Documenting learning of the key competencies: What are the issues? A discussion paper

Rosemary Hipkins, Sally Boyd, and Chris Joyce

∾∾∾

NEW ZEALAND COUNCIL FOR EDUCATIONAL RESEARCH

TE RŪNANGA O AOTEAROA MŌ TE RANGAHAU I TE MĀTAURANGA

WELLINGTON

2005

Acknowledgments

Table of Contents

Tables

Figures

1. Introduction

Through the Curriculum Project the Ministry of Education is engaged in an iterative co-construction process with the education community in New Zealand. One aspect of this work involves an intention to incorporate international understandings about the key competencies deemed necessary for lifelong learning into a reshaped curriculum framework.

Along with other informed commentators on the Curriculum Project we see this work as much more than a reframing or renaming of the essential skills in the current framework (Ministry of Education, 1993). The co-construction process draws on the theoretical underpinnings of the Defining and Selecting Key Competencies (DeSeCo) project co-ordinated by the OECD (as summarised in Rychen & Salganik, 2003). This is a project of international significance and discussion.

As in New Zealand's Curriculum Project, the development of the DeSeCo key competencies involved iterative processes of consultation. It was also underpinned by extensive theoretical research. The key competencies, as defined by the DeSeCo project, have been assessed for applicability to the New Zealand context and renamed to reflect this context. In the current iteration, five key competencies have been defined in a recent pamphlet for schools (Ministry of Education, 2005b) and in a tertiary education discussion document (Ministry of Education, 2005a). They are:

- thinking (creative, critical, and logical thinking, meta-cognition, self-awareness, and reflection);
- making meaning[1] (discovering meaning through the interpretation of cue and clues);
- relating to others (the knowledge, skills, and attitudes and values necessary for working and interacting with others);
- participating and contributing (contributing to communities); and
- managing self (making decisions for yourself including goal setting and planning).

This paper is premised on the assumption that assessment of a key competency is an assessment of a complex performance that integrates cognitive, meta-cognitive, affective and emotional aspects (de la Harpe & Radloff, 2000) such as skills, behaviours, attitudes and motivations, values, and understandings. Section Two, following this introduction, is a discussion of the nature of the key competencies, and further explains this assumption.

---

[1] This key competency has since been renamed "using language, symbols and texts".

Introduction to the case studies

As part of the Assessment Resources for Classroom Teachers contract, the New Zealand Council for Educational Research was asked to write this background paper to scope the challenges and issues that would need to be considered if the key competencies were to be assessed as a central part of the school curriculum.

This resulting report is a discussion document and should not be seen as a systematic review of the huge volume of assessment literature that is potentially relevant. While we did systematically search for literature about the assessment of key competencies, and the assessment of complex performances (*see* Section Two), only a small body of literature of direct relevance to our task was found. We then took a different starting point and began afresh by analysing case studies of situations in which key competencies (or some things nearly equivalent) have been assessed on a large scale. As we carried out this analysis the questions raised by the case studies pointed us towards assessment insights from other theorists. This process enabled the framing of the vast wider assessment literature in a more manageable way.

Selecting the case studies

We selected five case studies that allowed us to frame issues for the assessment of key competencies in both local and international contexts. The analysis of each case is reported in Section Four. These studies cover examples of key competency or complex performance assessment from early childhood to the tertiary sector. Most have been implemented at a national or, in the case of PISA, international level. All have generated a body of research literature that gave us a starting point for wider exploration of the relevant issues. The case studies are:

- Learning Stories (early childhood sector in New Zealand);
- New Basics (primary and secondary sector in Australia);
- PISA Problem-solving (secondary sector internationally);
- Vocational Education and Training (tertiary sector in Australia); and
- National Institute for Literacy (tertiary sector in USA).

Another important consideration in selecting our final group of cases was that between them they covered a range of *purposes* for large-scale assessment. Section Three outlines three overarching types of assessment purposes and relates these to three associated assessment paradigms. These are not seen as mutually exclusive, but there are some tensions that need to be resolved where there is an intention to employ several purposes simultaneously. The range of the selected case studies allowed us to sketch the different types of policy choices that have been made in the different contexts (countries and education sectors).

The structure of the report

As already noted, the paper starts with a discussion of the key competency model. Section Two also outlines potential areas of misunderstanding of this model, with associated assessment implications. It includes a brief discussion of the implications of locating key competencies within a socio-cultural theoretical framework.

Section Three outlines three assessment purposes and paradigms. It explores the implications for assessment in the New Zealand school context that arise from the analysis and also provides a framework for aspects of the case study analysis that follows. It concludes that assessment of key competencies will most appropriately be standards-based.

A description of each case study can be found in Section Four. These preliminary analyses are followed by a discussion of issues in common that emerged. A one-page summary of the various theory-practice interrelationships we found in the case studies is located at the beginning of Section Eight.

Sections Five-Seven cover some of the more practical considerations necessary when designing an assessment system. Section Five outlines types of assessment instruments and activities that could be appropriate for assessment of key competencies. Sections Six and Seven discuss questions and challenges for designing horizontal standards/progressions and vertical progressions. Both sections conclude that a co-construction process for designing any new system will address the potential misunderstandings of Section Two whilst drawing on the best of researchers' theoretical expertise and teachers' deep professional knowledge.

Finally, Section Eight is a synthesis of issues arising from the preceding sections and analyses. The key messages are:

- Assessment that supports lifelong learning and empowerment of learners should be the main focus for assessment of key competencies. This purpose is not easily compatible with assessment for other purposes such as accountability and this match would require careful consideration.
- To develop coherent systems that are widely understood, the current co-construction process should continue. Assessment processes should be developed alongside other ongoing curriculum work, and incorporate current theories about progression, learning and the key competencies themselves, as well as taking account of teachers' professional expertise.
- Professional development for teachers is a key factor in ensuring successful implementation.
- Demonstration of key competencies entails the assessment of a complex performance. Because transfer of learning is thought of as adaptation to a new situation, assessment should take place within a broad range of authentic situations. To enhance validity and reliability, each competency should be assessed more than once, and a variety of assessment methods should be employed.
- Standards-based assessment is more suitable for assessing and reporting on key competencies than standardised norm-referenced tests.

2. Nature of competencies

What is a key competency?

As they are defined in the New Zealand context, across early childhood, school and tertiary education sectors, key competencies align with the OECD's research-based DeSeCo model. Each competency "includes all the skills, knowledge, attitudes and values needed to do something" (Ministry of Education, 2005a, p.7). They are "more than discrete skills and attitudes: they integrate all aspects of learning" (Ministry of Education, 2005a, p.2).

The originating DeSeCo research similarly emphasises the holistic and universal nature of competencies that carry the prefix "key" and suggests that many other competencies, while important, are not "key" because they are situation specific (Rychen & Salganik, 2003). These OECD researchers assert that demonstration of a key competency requires a learner to show

> the ability to successfully meet complex demands in a particular context through the mobilization of psychosocial prerequisites (including both cognitive and non-cognitive aspects) (p. 43).

Rychen and Salganik emphasise that this is a functional approach to competence that places the

> complex demands and challenges that individuals encounter in the context of work and in everyday life at the forefront of the concept (p. 43).

Rychen and Salganik further explicitly state that it is inappropriate to apply the term "competence" to a single component such as cognitive functioning or motivation, because neither of these alone can make up a competence. This suggests a more complex approach to assessment that will require some rethinking of the assessment methods traditionally employed.

Differing views of key competencies: learning from the experiences of others

The experiences of others when implementing new key competency or generic skill frameworks suggest that it will be vital that all teachers, along with any other potential assessors of students' competency gains, are supported to develop a complex view of the key competencies, as outlined above.

In the VET sector in Australia there is no lack of agreement about the importance of key competencies, especially in regard to the need for young people to develop skills for the knowledge economy (Curtis and Denton, 2003; Comyn 2002; Blom and Clayton, 2003). But

although the TAFE[2] and school sector have been trialling and developing new forms of key competency assessment for a number of years, Curtis and Denton (2003) note that key competencies are not being "consistently promoted, presented or assessed in the VET sector in Australia" (p. 19). A variety of reasons is suggested for this including a lack of conceptual clarity, which has resulted in inconsistent understanding of the key competencies (Curtis and Denton, 2003; Kearns, 2001; Blom and Clayton, 2003).

Perhaps because of this lack of conceptual clarity, the term key competency has been used interchangeably with the term "essential skills" or "generic skills". Bryce (2003) suggests that it is difficult to distinguish "competencies" from "skills" as both terms are often loosely used to describe the same attributes. She observes that the definition of a skill has broadened in meaning beyond a discrete technical ability that can be isolated, practised, and perfected, to include habits and attitudes. Accordingly, she comments that definitions of the term competency can loosely be placed on a continuum from a behaviourist approach to an integrative approach. Similarly Blom and Clayton (2003) suggest that a lack of agreement about the constitution of competencies or skills leads to difficulties for recognising them in practice. They also consider this dilemma has been exacerbated by the interchangeability of the terms used to talk about generic skills, for example, key competencies, essential skills, core skills, or employability skills.

When discussing the pitfalls of past approaches Oates (2001) also alludes to definitional problems and a lack of understanding of key competency and employability skills models that preceded the current curriculum iteration in the UK. He argues that understanding the definition of competency is a significant problem because a "competency" does not exist in a simple way.

Carr (2004a) also suggests that it is difficult to succinctly summarise the definition of a competency in a way that acknowledges its complexity, and comments that it is important that the terminology used supports educators to adopt the model. She therefore suggests that to make the attitudinal dimensions more transparent, the key competencies could be called, for example, key skills and attitudes or key abilities and attitudes.

The combined experiences of these researchers and curriculum commentators suggests that any model of key competences that is adopted will need to be framed in a way that is coherent with the pedagogy and assessment purposes it promotes. This is challenging because it will also need to be presented in a manner that is easily understood and interpreted if it is to be appropriately used by the wider education community. If this balance can be achieved we may avoid the conceptual problems that have been suggested as a main reason for the variable uptake of previous key skills or competencies models. One way this clarity could be enhanced is through a careful consideration of the naming of the individual components of the key competencies. Other ways of avoiding these difficulties are through planned professional development as suggested by Oates (2001), or by co-construction of the assessment system as suggested in Sections Six and

---

[2]   TAFEs are equivalent to polytechnics.

Seven. However any implementation measures will need to take account of, and adequately address, aspects of the key competencies that remain contentious.

Ongoing debates about the nature of key competencies

Commenting on the expert papers written for the DeSeCo project, Kearns (2001) suggests that the variation observed between views demonstrates a lack of international consensus as to the identification and definition of generic skills. Discussions in more recent Ministry of Education background papers, and in the originating DeSeCo commentaries, highlight the differing views of various interest groups with regard to some specific aspects of the key competencies — in particular thinking, attitudes and belonging (Brewerton, 2004a; Rutherford, 2004). We next outline the nature of key debates as these relate to the definition and nature of the key competencies in the New Zealand school sector.

Varying views of the place of attitudes within the key competencies

The part played by attitudes and values within the key competency model, and the issue of whether they can be directly measured, or whether they can only be inferred, are two areas of contention. In an attempt to arrive at a description of a competency for the DeSeCo model, Rychen and Salganik (2003) state:

> A competence is manifest in actions, behaviors, or choices in particular situations or contexts. These actions, behaviours, or choices can be observed and measured, but the competence that underlies the performance, as well as the multiple attributes that contribute to it, can only be inferred… (Rychen & Salganik, 2003, p.48)

Carr (2004b) says this statement clearly shows that the DeSeCo model, and therefore the model of key competencies proposed for New Zealand, intentionally includes attitudes. Indeed attitudes, values, and motivations are integral components of the competency. However this has not always been seen as the case in earlier key competency initiatives.

In past initiatives, attitudes, values, and willingness to learn (motivation) have been excluded from some key competency models and included in others. The theoretical assumptions underpinning each model impact on the decision to exclude or include these attributes, as do difficulties of assessing these dimensions. For example Hill (29 April 2005, personal communication) comments that the term "competency" derives from, and therefore leads people towards an accountability framework, where attitudes might be seen as excluded. Her preference is for the use of a term such as "disposition", to fit within a socio-cultural discourse, and therefore lead teachers towards the use of more formative assessment practices. Within this type of framing, attitudes are relevant and should be taken into account.

Bryce (2003) suggests that definitions of competencies can be placed on a continuum from behaviourist approaches that assess "atomistic" skills which can be demonstrated (these

definitions tend to exclude attitudes and values), to integrative approaches which assume that what is being observed and assessed is an integration of knowledge, abilities, skills, understandings, and attitudes. (A theoretical framework for the competencies is further discussed below).

One example of a key competency model nearer the behaviourist end of the continuum is the model developed in 1992 in Australia by the Mayer Committee. This is the model underpinning the VET case study reported in Section Four. Values and attitudes were not included in the Mayer model because the development committee considered that these dimensions did not fit within their assessment criteria. Specifically, they felt that these aspects were not able to be "delivered, assessed and reported upon credibly" (Curtis and Denton 2003, p. 22). But although attitudes and values were overtly excluded, Bryce (2003) nevertheless notes that this conceptualisation of key competencies was underpinned by an assumption that what is being observed and assessed is an integration of knowledge, abilities, skills, and attitudes. Commentators suggest that this inconsistency has contributed to a lack of conceptual clarity for this model, and that this, as well as other problems with the model and within the VET sector, have contributed to the variable uptake of this framework.

Kearns (2001) observes that most countries are moving towards including attitudes and values in their models because the development of competencies cannot be separated from these dimensions. Likewise, Curtis and Denton (2003) observe that other frameworks in Australia include these dimensions, and Kearns (2001) notes that this is also the case internationally (and gives the example of a USA model of workplace basic skills that incorporates personal attributes, values, and ethics).

The general trend in the Australian VET literature is to suggest that an integrative definition of competency—one that incorporates attitudes and values—is more in line with current educational trends such as the growth in emphasis on lifelong learning (Bryce, 2003; Curtis and Denton, 2003; Kearns, 2001). Curtis and Denton (2003) and Kearns (2001) suggest that work done by the Mayer Committee needs to be revisited to better align the key competencies with current understandings in this area, including with the work being done by DeSeCo.

In a discussion paper for the Curriculum Project, Carr offers this advice:

> Given the vulnerability of any array of skills and attitudes to criticism from sectors with diverse (often implicit) views on curriculum, learning and teaching, it seems important that the array should sit inside a coherent theoretical framework. Also, given the current view of the Ministry that these Essential Skills centre on the individual learner, it is the view of this paper that the fundamental underlying principle will be to do with a coherent, consistent, and convincing viewpoint on *learning* (Carr, 2004a) p.3. Emphasis in original).

This is sensible advice and we return at the end of this section to the role of a coherent theoretical framework for relating the key competencies to learning.

Debates about the place of "thinking"

Kearns (2001) comments that all the Australian VET key competencies have meta-cognition running through them as a "meta-competency" or developmental progression, rather than being seen as a separate competency. Clayton, Blom, Meyers, and Bateman (2003) also view meta-cognition as a generic meta-competency. They suggest that, "…the ability to conceptualise and articulate possession of generic skills may be regarded as the most sophisticated of the generic skills — a meta-generic skill, perhaps" (p. 21). In their latest report Carr and Peters reach a similar conclusion about the role of critical thinking:

> ..the competencies call for a critical stance that enables people to make judgements about the information they receive, to "re-author" rather than just understand their feelings and emotions, and value, prioritise and resolve conflicts among abstractions….. Hence critical thinking is an aspect of all the competencies (Carr & Peters, 2005, draft cited with permission from authors).

This lends support to the argument that the theoretical underpinnings of the DeSeCo work, which places "thinking" as a cross-cutting dimension rather than as a discrete competency, should not be set aside (Rutherford, 2004). Currently the most recent New Zealand model places "thinking" as a separate competency in its own right (Ministry of Education, 2005b), but this is still obviously an area of contention.

Debates about the place of "belonging"

The quote from Carr and Peters potentially makes an interesting link between debates about thinking and debates about belonging. (The person who can critically "re-author" is demonstrating their ability to adjust their "being" to the demands of the context.) In her report of the initial consultation around key competencies, Rutherford (2004) notes that belonging has been "an issue of some debate" (p. 12) and these debates have included discussion of "identity, identities and relationships" (p. 12).

Carr (2004b) summarises the discussions at the Curriculum Reference Group on March 31 2004, where some participants suggested that "belonging" was an important condition for learning competencies rather a competency in its own right. She noted that support for a view of belonging as a competency was particularly strong amongst Mäori and Pasifika members of the reference group. Our own participation in the Curriculum project has led us to the impression that some of the debate centres around whether or not "belonging" is something that can be "done" and hence observed and assessed as a competence in its own right.

In subsequent work Carr and Peters (2005) have found that a sense of belonging was integral in some way to all nine action research projects carried out by early childhood and beginning school teachers in their research project. They also report that:

> The teachers developed indicators for this competence (as well as for others), trialled assessment procedures, and elaborated this category in terms of kaupapa Maori. This kind of

exercise provided very real evidence that this competence can make sense to teachers (Carr & Peters, 2005, draft cited with authors' permission).

They recommend that belonging should be included as a competency, at least for early childhood education and school. However OECD research suggests it is important to nurture this sense of belonging all the way through schooling so that students remain engaged with their learning (Willms, 2003). In a draft discussion paper Brewerton (2004b) takes a different view. She says that participation is a more important influence on learning than belonging, and so this is the competency that should be fostered for this purpose (p. 19). However early childhood educators tend to see participation and belonging as two faces of the same coin (Margaret Carr, personal communication).

Clearly there are unresolved tensions here that will need further discussion before or during the design of any proposed assessment system.

Debates about the place of "knowledge"

It is important that the relationship of traditional curriculum "content" to the key competencies be explored and clarified.

Brewerton (2004b) notes that previous surveys of business stakeholders in the wider community have found knowledge to be a lower priority as an outcome of schooling than attitudes and values and key competencies. (We think this contrasts with the emphasis accorded knowledge by secondary teachers). However she also cautions that the varying contexts in which different surveys have been undertaken may have influenced this result. A survey of university staff gave a much higher overall priority to "sound academic achievement" than have other similar surveys. Given the role and influence of tertiary educators on curriculum development and innovation (see for example White, 2003), this difference of priorities and perspectives is important to keep in mind.

Hargreaves (2004) presents a thought provoking perspective on links between the importance of knowledge, processes for curriculum review, and lifelong learning. He reiterates the importance of knowledge, whilst also addressing issues of the pace of change of knowledge and the associated necessity for curriculum content reduction:

> Dispositions and attitudes are often said to be more important than knowledge, which is now claimed by some to become quickly out of date, and the most important skills are taken to be transferable ones. There is some exaggeration here: much of the knowledge that is acquired in school, especially but not exclusively during the primary years, is the indispensable background knowledge of what it means to be an educated person, and its value does not diminish with age. Even when such knowledge is modified or replaced, the recognition that knowledge can sometimes change is itself educative. To assign more importance to skill acquisition may mean we have to be more selective in what established knowledge is transmitted through schooling: and this is merely a reiteration of an old

> problem that the school curriculum is inherently a selection of knowledge, and as such will always be a contentious matter (Hargreaves 2004, p.7)

Gilbert (2005) documents changes to views of knowledge that have significant implications for teaching, learning and assessment. She notes that rather than being valued for its own sake, knowledge is valued for its *performativity* — that is, its ability to be used in new and innovative ways to achieve new ends. In this view, new knowledge generation is no longer the preserve of experts who have served a lengthy 'knowledge apprenticeship'. Students of all ages need chances to be performative — to do things that create genuinely new knowledge. To help them do this they need to learn to take a 'meta' level view — to know more about where knowledge comes from, who decides what is worth knowing, and how tacit 'rules' of knowledge construction operate in different situations. The recent emphasis on the importance of "meta-cognition" in learning is one indicator of this shift. Another indicator is the increasing emphasis on learning about the "nature" of a subject — for example the "nature of science" or the "nature of history".

It seems to us that this more active view of knowledge is entirely compatible with its assessment as an *integral* component of key competencies that demand demonstration of an actual performance in a real context. However traditional knowledge-based assessment instruments are seldom structured in this way. There are already informal indications that secondary teachers, for example, may interpret "making meaning" (or as it now called, "using language, symbols and texts") as being able to be assessed by existing content-dominated NCEA achievements standards, and as being demonstrated when knowledge as a "thing" is recalled, written down, and perhaps explained or discussed.

This situation is complex and calls for very clear thinking and communication. As Brewerton points out, knowledge *is* an important part of all the key competencies:

> People need to be able to identify and 'perform' competencies in different contexts with an understanding of the demands of different contexts. This is where knowledge becomes an integral part of a key competency, and also highlights the importance of 'situational' learning (Brewerton, 2004b, p.41).

Brewerton goes on to identify contexts in which such use of knowledge might be profitably demonstrated, including family, social participation and work experience (p. 44). Interestingly, she also potentially links knowledge to issues of identity and belonging when she suggests that "Identity forms a condition for learning and can also provide content for learning for all students (not just those from non-Pakeha communities)" (p. 23). If identity is taken as an important aspect of belonging, then this comment seems somewhat at odds with her observation (reported above) that participation seems a more important influence on learning than belonging. However, this dilemma is neatly resolved if participation and belonging are seen as two faces of the same coin.

Issues of generalisability

Carr and Claxton (2002) note that dispositions reflect culturally determined values. For example, some cultures value cooperation over competition. Rychen and Salganik (2003) suggest that the differences may not be in regard to the types of generic competencies but rather in the weight given to them, or the way they are interpreted, between cultures.

Concerns about generalisability seem particularly strong in the vocational sector because specialised skills are required for some areas. Curtis and Denton (2003) note that failure to take account of the importance of specific generic skills in particular industries is one of the reasons for the variable uptake of the 1992 Mayer Committee key competency framework in Australia. However within the DeSeCo definition, as we have noted above, competencies that are not needed by everyone would not be regarded as "key" at all.

Issues of generalisability are probably not fully resolvable but we can learn from the Australian and international experiences here. Rather than trying for consensus, one approach could be to offer a more flexible framework which outlines core generic key competencies but also allows for the inclusion of other specific competencies that are valued by particular cultures or groups. Examples could include specific competencies important in kura kaupapa Mäori education, or in schools of special character, or in a particular industry in the case of the vocational training sector.

Issues of transfer or adaptability

Closely aligned with the question of the extent to which key competencies are generic, opinions vary on whether key competencies are transferable. Most commentators say that this issue needs to be addressed (Carr & Claxton, 2002; Clayton, Blom, Meyers, & Bateman, 2003; Curtis & Denton, 2003; Kearns, 2001; Kelly, 2001; McCurry & Bryce, 1997). Clayton, Blom, Meyers, and Bateman (2003) state that unless arguments about transferability can be resolved, definitional and assessment issues will not be resolved.

Kelly (2001) expresses concerns about the transferability of key competencies if they are assessed discretely within one defined area. She points to a lack of New Zealand research about the transfer of learning. However, we note some later research relevant to the tertiary sector, for example, McDonald (2001), and Doyle (2002). Likewise, Kearns (2001) notes that not enough is known about how competence is acquired and therefore how it can be taught and assessed.

McCurry and Bryce (1997) note that some commentators consider that competency is context or domain specific and hence they question the notion that there is such a thing as a "generic competency" which can transferred. Others take a more qualified view. For example Carr and Claxton (2002) suggest that dispositions are *both* transferable and situational. Carr (2004a) comments that the literature on transfer is complex, and suggests that transfer is related to a learner's meta-cognitive strategies, ability to practise competencies in a range of settings and, identity as a learner. A similar view is held by Delandshere and Petrosky (1998) who note that

"…as the structure of the construct or case becomes more complex, the conditions under which it generalizes become restricted and have to be qualified" (p. 20).

This argument suggests that the transferability of key competencies varies on a continuum that is individual to each learner. It depends on the nature, familiarity and complexity of the situation they face. Congruent with this view, Rychen and Salganik (2003) conceptualise the ability to transfer learning to new situations as "adaptation". Adaptation entails:

> actively and reflectively using the knowledge, skills or strategies developed in one social field, analyzing the new field, and translating and adapting the original knowledge, skills or strategies to the demands of the new situation (p. 48).

In this view, competencies can only be assessed when the assessment situation allows for adaptation to a new context to be demonstrated. Oates (2001) also views the concept of "adaptation" as more robust and helpful than the more traditional ideas of transfer.

Issues of transfer potentially also link to the debates on the place of attitudes, values, motivations, and belonging. Carr, (2004a) cites commentators who consider that identity as a learner is a key to transfer. Haskell (2001) describes a "spirit of transfer" that is a "psychological, emotional and motivational disposition to deep learning" (p. 117). He says transfer is influenced by traits such as persistence, locus of control, confidence, anxiety, and fear of failure. Making another type of link — this time to learning contexts and support — he also asserts that it is important that the issue of willingness to transfer is not seen as a concern for the individual alone but rather that a "culture of transfer" should be created in the classroom by setting up the conditions that foster this willingness.

Framing the key competencies as complex performances

So far this section has outlined definitional debates and found a lack of clarity surrounding past key competency models. We have noted that those who implement any intended New Zealand school intended model need to have a clear sense of what a key competency is (and indeed, is not). As we have worked on this background paper, we have found it helpful to frame the demonstration of the DeSeCo-defined competencies as demonstrations of *complex performances*. This is because they must integrate skills, behaviours, attitudes and motivations, values, and understandings. Brewerton (2004b) defines "performance" as "actively and intentionally doing things, including thinking" (p. 8) which is congruent with the view of knowledge discussed above. Furthermore the "performance" to be assessed also requires students to show adaptation to new contexts (Rychen & Salganik, 2003), thereby addressing the issue of transfer of learning. Thus we think that using terminology such as "complex performance" is one way of supporting educators to further develop their understandings of the model.

Additionally, the small literature we have found that relates to assessment of complex performances, (for example, Delandshere and Petrosky, 1998) has helpfully framed many aspects

of our analysis. Delandshere and Petrosky (1998) discuss differences between models that underpin the assessment of academic achievement and those that pertain to the assessment of complex performances. They show how traditional assessment paradigms have different underpinning assumptions from those they found helpful for assessing the work of master teachers — which they see as necessitating complex performances that integrate many different components in the moment of teaching. For example, they note that measurement theory is built on the assumption that the attributes (knowledge, attitude, or skill) being measured are in a steady state and that assessments sample from a domain of behaviours in a way that is indicative of likely performance in the whole domain. Any variations found in assessments on different occasions of measurement are then seen to be a "nuisance" and the result of error. In the complex performance paradigm these variations could very well be indicators of learning or maturation of the individual. The danger of applying an inappropriate assessment model is that the "part of the performance that uniquely responds to the specific context becomes translated as error" (p. 20). By contrast, tasks developed for the assessment of complex performances are based on assumptions that knowledge is individually and socially constructed and captures performance within contexts that vary "over time, across situations, and across individuals" (p. 19). Clearly, this description is very close to the idea of adaptation, as outlined above.

From this analysis Delandshere and Petrosky (1998) suggest that the "constructs underlying complex performances seem to be of a *different nature* to those that originally resulted from statistical analyses of test scores" (p. 19, emphasis added). Section Three explores the issue of assessment paradigms in more detail. For now, we turn to the question of a theoretical framework for learning key competencies, as recommended by Carr.

A theoretical framework for learning key competencies

As we have outlined above, the literature about the key competencies emphasises their holistic and contextual nature. Brewerton describes any intended curriculum initiative as:

> …taking an ecological or 'contextualist' approach to learning and living, where young people's learning is seen to be influenced by the various contexts of their lives (microsystems), the interactions between the contexts (mesosystems), and by the secondary and wider influences on those contexts (exo- and macro-systems). This perspective reflects the widely supported ecological approach of Brofenbrenner that underpins NZ early childhood education as expressed in Te Whäriki: Early Childhood Curriculum. … It also reflects the socio-cultural perspective on learning, which suggests all learning is mediated through cultural tools, primarily language (Brewerton, 2004b, p.7).

Socio-cultural theorists take the view that the individual is constructed in a web of relationships with other people, the natural world, and with the technological artifacts of our built world. Jay Lemke, an American linguist, began a recent article with the old saying "it takes a village to raise a child". He went on to draw a rich picture of learning as something that is socially situated, even when we perceive the child as a separate and distinct self at the heart of the process:

> As we learn, we gradually become our villages: we internalize the diversity of viewpoints that collectively makes sense of all that goes on in the community. At the same time, we develop values and identities; in small tasks and large projects, we discover the ways we like to work, the people we want to be, the accomplishments that make us proud. In all these activities we constantly need to make sense of the ideas and values of others, to integrate differing viewpoints and desires, different ways of talking and doing. As we participate in community life, we inevitably become in part the people that others need us to be, and many of us find our efforts unsupported or even strenuously opposed by others (Lemke, 2002, p.34)

If competency is seen not to reside in individuals alone there are implications for the theoretical underpinnings of any assessment system that will be compatible with the demonstration of competencies as complex performances. To illustrate with just one example, Carr (2004a) like Lemke, asserts that belonging must be seen as integral to the ability to demonstrate transfer of learning. Within the socio-cultural paradigm transfer is intimately bound up with identity. It is about anchoring our actions in our current ways of knowing and ways of being, while also changing these as necessitated by the challenge of adapting to new situations.

While all of Carr's work strongly supports the use of a socio-cultural theoretical paradigm as the key competency initiative is further developed, it is important to note that she does not see this choice as necessarily excluding the incorporation of work based in other paradigms. Rather she argues for a broad-spectrum framework to prevent "theoretical excesses" (Carr, 2004a, p.7) in any direction. She draws on a range of socio-cultural theorists, integrating Sfard's two metaphors for learning as acquisition and participation, Greeno's synthesis that subsumes behaviourist and cognitive theories into situative views of learning, and Rogoff's suggestion of fore-grounding and back-grounding the individual, the interpersonal and the cultural-institutional as units of analysis. This meta-synthesis leads her to an overarching theoretical perspective where:

> Learning is distributed across the resources of self, other people, cultural tools, and community. Learners need skills for accessing and developing these resources and for recognising their purpose over place and time (Carr, 2004a, p.8).

Carr goes on to link the competencies to the various components of this overarching framework thus:

- the self links to "managing self";
- other people links to "relating to others";
- cultural tools links to "making meaning from information" and "ways of thinking"; and
- community links to "participating and contributing" (p. 8).

This approach, which incorporates elements of a number of theoretical positions, seems to be a useful framework with which to proceed. As Section Three shows, it will be important to integrate different theoretical perspectives on assessment in a considered manner, especially where assessment is required to serve multiple purposes. This is the issue to which we turn next.

3. Policy-related purposes for assessment of key competencies

This section discusses the various reasons that might be proposed for the assessment of key competencies and relates these to differing sets of questions regarding the form the assessment system could take. The analysis points to three assessment paradigms, each related to one of three main purposes for assessment. Differences in underpinning assumptions about teaching and learning for each purpose/paradigm pair are discussed.

The analysis in Section Four shows that it is common for established assessment systems to mix purposes and systems features from across the three paradigms. Section Three shows that a mix of purposes is also intended for the introduction of key competencies into New Zealand's mandated school curriculum and scopes the choices that face policy makers as they seek to align assessment purposes with the desired and designed features of the assessment system.

Three assessment paradigms and related purposes

The assessment literature identifies three broad sets of purposes that assessment of key competencies might be expected to serve. Briefly, these are:

- assessing for accountability at school and wider policy levels, and for reporting learning progress to any stakeholders who have a need for this information (the traditional purpose for externally driven summative assessment);
- using assessment to support students' learning and as a driver for improving classroom teaching and learning practices (the focus of many formative assessment initiatives, for example); and
- using assessment to empower students to become lifelong learners (so-called critical or emancipatory purposes for assessment).

Each purpose is discussed more fully in the three sections that follow.

The three purposes align rather neatly with three assessment paradigms described by Aikenhead (1997). He set out to construct a framework that would make explicit the theoretical orientations that underlie debate about assessment issues. Following Habermas' analysis of sociological research more generally, Aikenhead described empirical-analytic, interpretive and critical-theoretic assessment paradigms. The next table shows the purposes/paradigms alignment as we see this.

Table 1 **Alignment between three assessment purposes and paradigms**

Paradigm	Most compatible purpose	Nature of assessment
Empirical-analytic	Accountability (school and wider policy levels) and reporting	Empirical methods based on psychometric principles, yield "robust" comparative data.
Interpretive	Improving teaching and learning	Evidence of achievement against specified standards, may combine descriptive and data—based components. Judgements made by others, not students.
Critical-theoretic	Fostering lifelong learning	Extends features of interpretive paradigm— collaborative methods fully involve students and empower them to continue learning.

In what follows, each of these paradigms is described, aligned with one of the key purposes of assessment, and related to the OECD key competencies research. The analysis is further informed by other relevant international assessment research and critique, particularly that relating to assessment of "complex performances".

Purpose One: Systems accountability and reporting

In many ways this is the most familiar and traditional of the three assessment purposes. There are well-established precedents for using large-scale assessments (i.e. those that are undertaken by students from many schools) to report on students' learning progress but also to gauge the success of different schools in helping students meet the intended outcomes of their learning at school.

International assessment trends suggest that any new type of assessment system will be expected to provide information for accountability purposes that extend well beyond a focus on the success of individual students, or even of specific schools. Such purposes include monitoring the success of government policies, for example in relation to "raising student achievement" (Broadfoot & Black, 2004). Illustrating this trend, the specification of key competencies for the tertiary sector is intended to support government funded agencies to "use the framework in policy and operation work to achieve consistency about desirable learning outcomes across the school and tertiary education sectors" (Ministry of Education, 2005a, p.4). This is seen as an important policy focus because "there are considerable social and economic benefits for all from a well-educated population" (p. 4). Although there has not, as yet, been open discussion of assessment of key competencies for accountability purposes in the school sector, this is also likely to be an expectation of any assessment system that could be introduced.

In the senior secondary school, the aggregated results of students' assessment for qualifications have traditionally been used as a proxy for rather broad accountability purposes. Analysis and reporting of "league tables" of schools' relative success in external examinations is such a measure. However, since the introduction of the National Certificates of Educational Achievement (NCEA) the construction of league tables has become much less straightforward,

and creates tensions where there are conflicts with the other assessment purposes such as supporting the development of lifelong learning {Hipkins, in press #83}. These other purposes suggest a more fine-grained analysis of success in meeting desired outcomes is required, and as the next paragraph shows, is now formally expected.

The NAGs set out the current government's education priorities for the school sector and the regulations that ensure that schools work in responsible ways (Ministry of Education, 2002). As already noted, NAG One encompasses "success for all" and is clearly aligned with aspects of assessment purposes one and two. NAG Two specifies that the manner in which each school addresses these priorities, and the success with which it has done so, must be made explicit. Desired outcomes must be planned for and then reported within a 3–5 year strategic plan, with an annually updatable section, and a section that reports on the previous year against this strategic plan. A process of school self-review is an essential component of meeting these governance requirements. This suggests that any system of assessment of key competencies might be expected to provide useful information to guide schools' planning and review frameworks, across all curriculum levels. Since such self-review is intended to inform the modification of curriculum and teaching to better meet identified student learning needs, accountability and improving learning purposes need to be in alignment, and to be understood to be so by school leaders and teachers. While this might seem obvious, the *National Survey of Secondary Schools 2003* showed that the newly introduced Planning and Review Framework is currently more likely seen to be an information gathering exercise for the benefit of the MOE (Hipkins & Hodgen, 2004).

When assessment provides exit-level credentials (for example the NCEA and other national certificates on the National Qualifications Framework) there is also an issue of accountability to those beyond the school system, including parents, prospective employers, and those who assign places in tertiary education courses. In this case the assessment system that is designed must provide "information about what learners can do that is credible to employers, educational institutions and policymakers, as well as to the learners themselves" (Stein, 2000, p.57). This is one of three guiding principles in the holistic adult learning framework developed by the National Institute for Literacy (NIFL) in the USA (one of the five case studies developed in Section Four). This principle is important for its potential to align aspects of all three purposes for assessment. However, as recent experience with the NCEA has shown, gaining credibility for new methods of assessment is likely to be challenging.

The empirical-analytic paradigm

Assessment for accountability purposes, as outlined above, aligns with what Aikenhead called the "empirical-analytic" paradigm of assessment. The key features he identifies for this paradigm are as follows.

Purpose of assessment

Empirical-analytic assessment is undertaken for certification purposes, and for national policy development and social control of schools. In science (which was the focus of Aikenhead's work) the assessment has a focus on fostering elite students, and on the products of learning.

View of learning/positioning of learner

This paradigm is underpinned by a view of knowledge as an accumulated product, and by behaviourist learning theories. The learner receives assessment judgements made by experts.

Nature of assessment

Assessment is standardised and summative. Assessment issues tend to be technical in nature and validity is defined by technical, rational psychometric principles.

Comment

There are clear links between the empirical-analytic paradigm and the accountability purpose identified above. Standardised, summative assessment has traditionally served a screening purpose, in which elite scholars were identified and enabled to undertake further study while other students were encouraged to end their learning to seek employment. While this screening function has been seen as problematic for many years, both in New Zealand (Renwick, 1986) and internationally (Apple, 2004) the issue has become acute with the rapid pace of change in the knowledge era, itself strongly associated with calls for lifelong learning {Gilbert, 2005 #87}. Thus there are also evident incompatibilities with the improved teaching and learning purpose and particularly with the lifelong learning purpose outlined below.

Within this paradigm the learner is not involved in either setting tasks or evaluating their own performance to create summative statements about, or evidence of, competencies attained. Assessment can be highly standardised and moderated, as in the case of PISA, where the purpose is to compare students' performances internationally. While the tasks are supposedly "authentic" with such a broad range of experiences there is obviously potential that they may not be relevant to all learners.

Alternatively the students may be judged against their own or others' responses to some sort of Likert scale, as in the Competent Children study (Wylie, Thompson, & Lythe, 1999). This study illustrates methods that may be associated with an intention to track competency over time and in a range of changing settings. Interestingly it has been suggested that students can be involved in judging their performance across a course of learning, via the use of appropriate Likert scales, if the encouragement of lifelong learning (see below) is the purpose for carrying out the assessment (de la Harpe & Radloff, 2000).

As the need to categorise students' responses for reporting purposes increases, it becomes correspondingly more difficult for the evidence to be collected within the context of a rich, complex and authentic task. Some commentators (Curtis & Denton, 2003; Kalantzis, Cope, & Harvey, 2003; Tittle, 1994) believe that standardised assessment tools are inadequate for assessing key competencies, with others going so far as to claim that normative evaluation may actually adversely affect the development of competencies (Carr & Claxton, 2002). One way around this dilemma is to assess opportunities to learn in the classroom setting, which directly places accountability for learning with the teacher (Carr & Wylie, 2004).

Kelly (2001) identifies a number of barriers to assessment of the key competencies in the New Zealand system that are related to the issues just raised. In particular she comments that such assessment demands a move from a quantitative assessment system to one that is more qualitative. This necessitates a change of expectations for the assessors and the users of data and has cost implications. Kelly notes that traditional norm referenced pen and paper assessments are easier to run and more cost effective, and parents and employers expect and value this type of "hard" data. By comparison the types of assessments outlined below would be costly, assessors would need sustained exposure to contextualised performance to be able to make valid judgements, and the measurement of key competencies could be prone to more bias.

Purpose Two: Improving teaching and learning

Assessment may be intended to serve the policy purpose of supporting the improvement of teaching and learning, as in the *New Basics* case study described in Section Four. Again, discussion of the key competencies in the New Zealand context suggests that this purpose should be taken into consideration when assessment is discussed. To illustrate—information designed to disseminate understanding and promote discussion of the proposed key competencies amongst those involved in the New Zealand school sector states that the key competencies "offer an alternative way of viewing curriculum and a clearer focus for teacher practice" (Ministry of Education, 2005b, p.2). There is a clear implication here that the introduction of key competencies into the mandated curriculum is intended to support improved teaching and learning.

The government priorities set out in National Administration Guideline One (NAG One) suggest that the intended teaching and learning improvements will promote success for all in a safe learning environment. There will be a focus on improving literacy and numeracy, better use of student achievement information, improving outcomes for students at risk, improving Mäori outcomes, providing career guidance, and on clearer reporting (Ministry of Education, 2002). This intention aligns with other MOE policy initiatives such as the production of various "Best Evidence Syntheses" (BES) (see for example Alton-Lee, 2003).

An intention to more clearly align curriculum, pedagogy and assessment is explicit in the *New Basics* assessment reforms underway in the Australian State of Queensland (see Section Four). The "rich tasks" devised for assessment of a cluster of four key competencies that they have

called the "new basics" were overtly intended to force a curriculum re-mapping and the widespread uptake of "productive pedagogies" in the classroom (Lingard & Mills, 2003). For example they assert that:

> Effective school reform demands that the three message systems [curriculum, pedagogy and assessment] be aligned and not work at cross-purposes. For instance, to achieve a focus on higher order thinking or on fostering strong citizenship attributes, assessment practices need to be focused in that direction, as well as pedagogies (Lingard & Mills, 2003, p.4).

Lingard and Mills go on to discuss the types of assessment that they see as working to better align curriculum and pedagogy with students' assessment experiences:

> ..those forms of assessment, which appear to work for all students, and especially disengaged students, are those which expect students to intellectually engage with the task, have some meaningful purpose beyond being for grading purposes, are supported through clear expectations, and encourage students to both value and make a difference in their world (Lingard & Mills, 2003, pp.11-12).

The last sentence, in particular, strongly aligns with the New Zealand key competency of participating and contributing and indicates that what they envisage involves curriculum intentions that extend well beyond teaching and assessing traditional knowledge and skills. Writing in a similar vein, Carr and Claxton (2002) suggest that it is important to assess dispositions because what is assessed is what is valued, by both teachers and students. They suggest assessment of dispositions will help young people develop these dispositions and provide indicators of their progress in doing so, thereby assessing the effectiveness of their learning programmes. Thus the assessment process itself will help demonstrate the efficacy of educating for dispositions, thereby potentially adding a different dimension to calls for the reinforcement of traditional approaches to learning, as expressed in calls to go "back to basics". A similar point is made by Bryce (2003) who suggests that one way of acknowledging the importance of key competencies is to assess them in a "high stakes manner" by making them an essential part of the school-leaving certification process.

The interpretive paradigm

The interpretive paradigm described by Aikenhead aligns well with the assessment purpose of improving teaching and learning. The features he identified for this paradigm are as follows.

Purpose of assessment

Assessment is undertaken to inform students' achievement of knowledge, skills and values, or to improve learning and teaching. In science there is a focus on "science for all" rather than as an elite achievement.

View of learning/positioning of learner

Assessment involves making formative judgements, with students involved in gathering the evidence needed to make these judgements. Knowledge is seen as individually constructed, and this aligns with constructivist learning theories.

Nature of assessment

Assessment is closer to the learning than in the empirical-analytic paradigm. Thus assessment issues tend to be pedagogical and validity "distils to trustworthiness" of one professional for another.

Comment

There are clear links between the interpretive paradigm and the "improvement of teaching and learning" purpose outlined above. The main purpose of the assessment is formative—to provide feedback to the learner. Assessment always takes place in authentic contexts, sometimes provided by the teacher, other times selected by the learner. In these cases the final judgement is made by somebody else, such as a teacher or parent, although the learner is involved in discussions about their learning and what they are trying to achieve. Indicators such as exemplars may give guidance (to learners or teachers, or both) on what to look for.

In a discussion of assessment for certification as an "expert teacher" Delandshere and Petrosky (1998) assert that if the overall purpose of certification is to improve teaching and learning, then only a considered discussion of the *evidence* will suffice. Thus, while aggregated data (as would be gathered within an empirical-analytic paradigm) allows certification, it does not provide a basis for a reflective discussion of what the assessment actually means, and hence provides no feedback that could allow the teacher to change and grow professionally. It is particularly apt in the context of assessment of key competencies because these, like expert teaching, are envisaged as complex performances that integrate many separate competencies in a holistic performance delivered in an authentic context (see Section Two).

Delandshere and Petrosky provide an interesting example of the tension between accountability and learning/teaching improvement at work. The situation they have critically analysed suggests that both types of assessment data (empirical and richly descriptive) may need to be generated if the tension is to be resolved. Others have reached similar conclusions — see for example Carr and Wylie (2004) and the various essays in Wilson (2004). Their critique suggests that assessments of key competencies cannot be "one off" events such as the formal examinations that have traditionally been employed in the empirical-analytic paradigm, nor should the judgements made be reduced to simple ratings if the assessment is to inform ongoing learning.

The types of outcomes sought within this paradigm suggest that it will be necessary to develop *standards* to specify the nature of the achievement to be demonstrated. As the NIFL developers put the design challenge:

You build a broad consensus on what results the system should achieve. You develop standards that express that consensus in a clear and measurable form. Then you use those standards as tools for focusing all parts of the system on achieving the desired results. Teachers use the standards to guide teaching and learning. Curriculum and assessment specialists develop new tools that are based on the standards. Professional development focuses on building the knowledge and skills that teachers and other program staff need to implement the standards. Policymakers target resources on building the capacity of programs to prepare students to achieve the standards (Stein, 2000, p.88).

Purpose Three: Supporting lifelong learning

The future-focused literature often emphasises the need for students to become "lifelong learners" (Codd et al., 2002; Gilbert, 2005). The recently released key competencies pamphlet also asserts that "the suggested framework of key competencies promotes a lifelong learning model" (Ministry of Education, 2005b, p.2). To achieve this outcome, school leavers are those who "have a positive sense of identity, take responsibility for themselves, can interpret and critique the world around them, can participate and contribute effectively in a range of contexts, and are equipped for ongoing learning" (Ministry of Education, 2005b, p.3). Again there is alignment with aspects of other MOE policy initiatives such as the reform of the school qualifications system, intended to help all students gain a sense they can be successful learners (Hipkins, Conner, & Neill, in press).

De la Harpe and Radloff (2000) define lifelong learners as those who can manage cognitive, meta-cognitive, affective, and emotional aspects of their learning. These include knowing what constitutes good learning in a range of situations, when and how to seek help, when and how to collaborate with peers, and how, when, where, and why they learn best. Lifelong learners have self-knowledge, self-confidence, persistence, positive feelings about themselves as learners, and a positive view of the value of learning. They are able to organise and manage their time and to plan, monitor, and adapt their own learning.

Weimer (2002) describes five key aspects of teaching practice that need to accompany any shift to students-centred practice that supports the development of the independence required for lifelong learning. These aspects of change are:

- a shift in the balance of power so that students are actively involved in decision-making about their learning and assessment;
- a related shift that allows students to take greater responsibility for their learning;
- a shift from content *coverage* to its *use* in the context of the planned learning; and
- a change in the role of the teacher from expert telling to guiding, coaching, modelling, collaborating and giving feedback (a similar description of changed teacher roles to support greater student autonomy has been provided in recent New Zealand research (Bartlett, 2005); and

- a move from summative assessment to forms of assessment that support students' self-awareness of their own learning processes and that include formative assessment practices, including self and peer assessment.

From an analysis of seven case studies in Australian secondary schools, Bryce and Withers (2003) identified five key elements of learning programmes that focus on lifelong learning. Their analysis of programme changes aligns strongly with Weimer's analysis of shifts in teaching practice. Bryce and Withers recommended that ownership of the need to learn and of its content should be given to individuals, that learning should be about learning how to think rather than what to think, and that teachers should be mentors and models of lifelong learning more than dispensers of knowledge. They see the purpose of assessment as assisting and encouraging further exploration, rather than relating students to some concept of a 'norm' and say learning should be viewed as an enjoyable and integral part of one's life (Bryce & Withers 2003, p. 2).

In summary, the lifelong learning literature strongly signals a shift to student-centred teaching practice. While all these shifts are compatible with the policy requirements of NAG One, they go well beyond any change implied within those guidelines and suggest that changes that support lifelong learning require much more radical shifts in teaching practice and, perhaps in school organisation, particularly at the secondary school level. This suggestion of a need for radical change is supported by other future-focused research (for example Gilbert, 2005) and by the OECD's originating key competencies research (Rychen & Salganik, 2003).

The critical-theoretic paradigm

There is a clear alignment between the lifelong learning purpose for assessment and Aikenhead's third assessment paradigm. The following is a summary of the key features he identifies.

Purpose of assessment

Assessment is undertaken to empower students as lifelong learners. In science there is a focus on "science for all" but with an "activist orientation".

View of learning/positioning of learner

Whereas students are involved in gathering the evidence needed to make judgements in the interpretive paradigm, here they are involved in what Aikenhead calls "formative evaluation". That is, they take an active part in collecting evidence and in the judgement of their own performance. Knowledge is seen as socially constituted, and this aligns with socio-cultural learning theories.

Nature of assessment

Assessment issues tend to be social, political, and cultural. Aikenhead does not address validity within this paradigm. (But see below for a critique of validity from another assessment theorist.)

Comment

Again it is evident that this paradigm aligns strongly with one of the three purposes described above — that of promoting lifelong learning. Learning that supports the development of the key competencies and assessment of students' acquisition of these is closely intertwined, so that assessment itself may become an important part of their learning.

A socio-cultural approach to learning empowers learners by situating them centrally in the process of learning, but also at the centre of the assessment. Their viewpoint is included (Ministry of Education, 2004) and they share the interpretation of the assessment information with the teacher (Tittle, 1994). This could involve students selecting the evidence that demonstrates the key competencies they can exhibit within the learning context. While this might take place within the classroom, it could also be within contexts outside formal learning situations (see for example the VET and Learning Story case studies in Section Four). In some cases it is not mandatory and students choose to complete this part of the assessment (VET case study). There is usually some sort of validation of students' own judgements. In the VET case study, the tutor does this, while in the early childhood scenario, the children may instigate the writing of the learning story but the actual writing is done by an adult. These sorts of assessments can be used for summative purposes (to demonstrate a competency for employers, for example), but can also be used formatively to identify areas for future learning.

The activist orientation of the critical-theoretic paradigm, as suggested by Aikenhead, aligns with the idea of "action competence" discussed in the DeSeCo research (Rychen & Salganik, 2003) and translated into the New Zealand key competencies model as "participating and contributing". Another perspective on the link between Aikenhead's framework and the DeSeCo work is provided by Gilomen (2003) who defines the dimensions of "well-functioning society" as being related to economic productivity, democratic processes, solidarity and social cohesion, human rights and peace, equity, equality and absence of discrimination, and ecological sustainability (p. 128–132). This powerful list of dimensions can arguably best be achieved if students are educated and assessed within the critical-theoretic paradigm that empowers them to take the necessary critical action.

This focus on action competence poses some interesting assessment issues. As Rychen and Salganik note the relationship between the individual and society is:

> dialectical and dynamic. Individuals do not operate in a social vacuum. Actions always take place in a social or socio-cultural environment, in a context that is structured into multiple social fields …. each consisting of a structured set of social positions dynamically organized around a given set of social interests and challenges. It is within these fields that demands

and the criteria for effective performance and action take form and manifest themselves, and individuals act to meet them (Rychen & Salganik, 2003, pp.45-46).

If assessment of key competencies is intended to encourage and support lifelong learning that includes action competence, then any assessment system that is designed will, at the very least, need to find ways to support the performance of individuals within the full complexity of social settings. Rychen and Salganik (2003) acknowledge that DeSeCo focuses on an individual rather than collective competence but say they sought to capture elements of this in the "relating to others" competence, which has also been included in the New Zealand key competencies.

Some commentators would extend this issue to assert that, since most new knowledge building activity now takes place in team settings, assessment of a solitary performance, long seen as a quality measure in the empirical-analytic paradigm, may no longer be appropriate {see for example \Gilbert, 2005 #87}.

Delandshere and Petrosky (1998) identify another tension between the empirical and the interpretive or critical paradigms. They point out that it is difficult for students to question expertly constructed test scores if they don't have the necessary expert knowledge. At least within the critical paradigm, ownership of the interpretation needs to belong to the student if they are to continue to learn as a result of assessment feedback.

Assessing attitudes and values

The usefulness of a framework that aligns paradigms/purposes can be further illustrated in the context of the assessment of attitudes and values. Section Two noted that these are integral to the idea of key competencies and should be assessed holistically within them. However we also noted that this aspect of past key competency initiatives has been problematic and will need to be handled carefully if new assessment systems are developed.

As the next table shows, the paradigm within which assessment is grounded influences the manner in which attitudinal dimensions are viewed and assessed. For example, traditionally in educational settings, adopting a behaviourist paradigm leads to the assessment of attitudes and values separate from context. This could be through the use of Likert-type scales as TIMSS[3] and PISA (see the case studies in Section Four) to assess attitudes towards curriculum areas such as mathematics and science.

---

[3]   http://timss.bc.edu/timss1999b/mathbench_report/t99bmath_chap_4_1.html

Table 2 **Approaches taken to the assessment of attitudes and values**

Critical-theoretic	Interpretive	Empirical-analytic
The premises of this framework suggest that attitudes and motivations should be discussed and explored with the learner, who is centrally involved in making inferences from the available evidence	Attitudes and motivations are inferred by an expert judge, although the learner may be involved in selecting the evidence	Attitudes and motivations assessed separately

Blom and Clayton (2003) note that often the assessment of key competencies is inferred rather than evaluated. Rychen and Salganik (2003) and Oates (2001) consider that the attributes (including attitudes and motivations) underpinning the manifestation of a competency can only be inferred, as do Delandshere and Petrosky (1998). However, as the table shows, the learner is differently positioned within the interpretive and critical-theoretic paradigms. Only in the latter are they empowered to move beyond the risk of misinterpretation — either their misreading of what is wanted by the assessor, or the assessor's misreading of the meaning of the evidence they present (Rychen & Salganik, 2003). Only in the critical-theoretic paradigm does assessment become a process of inquiry into the meaning of the learning that has taken place (Delandshere, 2002).

Possibilities for integrating the three paradigms

An analysis of assessment principles in another NZCER project (Doyle, unpublished manuscript) has found that assessment systems that are able to gather sufficiently rich data to support lifelong learning aims are also able to subsume the purposes of assessing to support better teaching and learning, and accountability purposes. We see similarities here with Carr's proposal for a broad spectrum but coherent theoretical framework for learning, as outlined in Section Two. It will be important to establish conceptual clarity around the match between assessment purposes and theoretical paradigms. The question is not which system should we privilege but how we design a system that is as compatible as possible with the mix of, and relative emphasis on, the purposes we wish it to serve.

Conclusions

Three main purposes for assessment can be identified. Each aligns closely to one of three paradigms that can be described for assessment practice.

- *Assessment for accountability and reporting* aligns closely with the familiar empirical-analytic paradigm. Assessment is standardised and psychometric tools may be used to establish validity and reliability. Students are not involved in the assessment process beyond demonstrating their learning in the manner required.
- *Assessment for teaching and learning* aligns with a less familiar interpretive paradigm. Assessment is more closely linked to classroom practice and the strength of the teacher's professional judgement is an important aspect of validity and reliability. Students are involved in collecting evidence of their learning but this evidence is interpreted by expert others.
- *Assessment for lifelong learning* aligns with an unfamiliar (for most teachers) critical-theoretic paradigm. Students are centrally involved in decision-making about all aspects of their learning and assessment, including the judgements made about their progress.

On balance, the least familiar critical-theoretic paradigm makes the most encompassing match to the purposes envisaged for introducing key competencies into school curricula internationally. There are a number of implications for policy-making that can be drawn from this analysis and conclusion:

- Assessment needs to be standards-based rather than normative in order to allow for the demonstration of learning as a complex performance.
- Demonstration of competencies as complex performances necessarily requires assessment inferences to be made. These are most likely to empower ongoing learning when students are directly involved in making the relevant judgements.
- Such approaches will necessitate considerable changes in traditional teaching and assessment practice, and hence there are significant implications for teacher professional development.
- Because traditional notions of validity and reliability (often expressed as "fairness") are deeply grounded in the familiar empirical-analytic paradigm, care will need to be taken with keeping the wider public abreast of any assessment developments, so that the changes are not misinterpreted.
- Any system that is thoughtfully designed so that the above conditions are met should also be able to serve accountability purposes.

4. The case studies

This section begins with a description and analysis of each of the five selected case studies. They are arranged in order of setting — from early childhood to tertiary education. A common framework has been used for the analysis of each case.

Each of these assessment initiatives has been supported by well-documented research. Aspects of this research relevant to the wider discussion in this background paper are briefly outlined in the introduction to each case.

In the case of the VET sector in Australia (Case Four) the overall initiative is outlined and then one specific case of implementation is used for the analysis. Because this initiative has been running for some time, a wider range of research information than in the other cases was available. The analysis is followed by a comparison between the TAFE case study and aspects of unit standards assessment for the National Certificate of Employment Skills (NCES) in New Zealand.

Key themes in common and issues that emerged are briefly outlined at the end of the section. A meta-summary of these aspects is included in Section Eight.

Learning Stories: Assessment in New Zealand early childhood education

The research project *Assessing Children's Experiences in Early Childhood Settings* (Carr, 1998a) was chosen as a case study for several reasons.

- Te Whäriki, New Zealand's Early Childhood curriculum, has been suggested as a starting place for considering key competencies. Its five learning strands, Belonging, Wellbeing, Exploration, Communication, and Contribution, integrate knowledge, skills and attitudes in a way that is "consistent with the recent understandings of competencies" (Brewerton, 2004a, p. 11).
- There is a clear match between the principles of learning on which this curriculum is based and the method of assessment. Both reflect a socio-cultural approach to learning, which Section Two identified as providing a theoretical framework compatible with the introduction of key competencies in the school sector.
- The initial research has been picked up by other practitioners, and Learning Stories have become an accepted way of assessing in Early Childhood, as demonstrated in the assessment exemplars recently published (Ministry of Education, 2004).

The first research project that developed the Learning Story framework for assessment was a Ministry contract directed by Margaret Carr, Assessing Children's Experiences in Early Childhood. The Final Report was published by the Ministry of Education in 1998. This research with teachers established the Learning Stories framework and trialled it in five contexts: a childcare centre, a kōhanga reo, a play centre, a kindergarten, and a home-based setting. The case studies were published in Part Two of the report.

Rather than "key competencies", this research focused on "learning dispositions" that incorporate both situated learning strategies and motivation. Carr (2001) defines learning dispositions as "situated learning strategies + motivation" (p. 5). Learning dispositions are described as having three parts: being ready (seeing the self as a participating learner), being willing (recognising places of learning), and being able (having the abilities and knowledge that contribute to being ready and willing). The nature of these dispositions, and their links to the curriculum and to observable actions is summarised in the next table, based on Carr's (1998b) support book for teachers (Reading 2, pp.14, 15).

Table 3 **The nature of learning dispositions in early childhood education**

Strands	Learning Dispositions	Actions and behaviours
Belonging	Courage and curiosity	Taking an interest
Wellbeing	Trust (and playfulness)	Being involved
Exploration	Perseverance	Persisting with difficulty, challenge, and uncertainty
Communication	Confidence	Expressing a point of view or feeling
Contribution	Responsibility	Taking responsibility

As the ministry sees them, links between these learning dispositions and the Ministry's suggested key competencies are as follows:

- Wellbeing and belonging = Managing self;
- Contribution = Relating to others and Participating and contributing;
- Communication = Making meaning; and
- Exploration = Thinking (Ministry of Education, 2005b).

The concept of "learning stories" was explored as a means of assessing and documenting pre-school children's learning and progress. Learning stories are structured narratives of children displaying some aspect of the learning disposition(s) during a learning event. An accumulation of learning stories over time provides a picture of the child's developing learning dispositions. The stories may be written by teachers, parents, or even instigated by the *learners themselves*.

Analysis of key features

Title	Children's Experiences in Early Childhood Settings
KC being assessed	Learning dispositions were derived from the curriculum strands of Te Whäriki. To assess the learning dispositions, actions that expressed them were described.
Sector	Early childhood
Description of the assessment	Learning stories describe critical incidents that: Include a description of the context; Often describe relationships with other people; Highlight the activity/ task; Include an interpretation from someone who knows the learner well; Focus on evidence of new or sustained interest, involvement, challenge, communication and responsibility. A collection of Learning Stories provides a cumulative series of qualitative "snapshots" of the learner. Assessment using Learning Stories involves four actions (the four Ds): Describing—what the learner is trying to achieve; Documenting—the evidence of learning; Discussing—with the learner, other teachers, parents; and Deciding—what to do next to improve the learning. Learning stories are interpretative, authentic, reciprocal methods of assessment. They focus on what the child is doing, not what they don't do (credit rather than deficit assessment).
Nature of the assessment	Formative. Collections of stories document learning over time
How are the assessment results reported and recorded?	Learning stories are kept in portfolios or folders. They may include photographs, photocopies of children's work, comments from the children. Portfolios are shared with children, parents, and other teachers. Early Childhood centres developed a format for writing up Learning Stories. Some of these included examples or cues; for example, beside "Being Involved" the cues for one recording sheet were — paying attention for a sustained period, feeling safe, trusting others, being playful with others and/or materials. Some also had sections for Short Term Review and What Next?, with prompt questions to guide responses (Carr, 2001).
Why was the assessment developed?	In broad terms, the research project was to provide support for the implementation of the early childhood curriculum. More specifically, Carr and Claxton (2002) say that it is important to assess learning dispositions: So the dispositions are not lost sight of; To evaluate the learning programme; Because what is assessed is what is valued; and To back up theory with evidence.
How was the assessment developed?	The framework was developed during an earlier project, the Project for Assessing Children's Experiences (P.A.C.E.), and then trialled in a range of New Zealand early childhood centres.
Theoretical underpinnings	Learning Stories originate from narrative or storied approaches in education — refer to Bruner (1986, 1990), Clandinin and Connelly (1990), Witherell and Noddings (1991), Genishi (1992), Middleton and May (1997) and Carr, May, and Podmore (2000). However, they are more structured than narratives. Te Whäriki is based on a socio-cultural approach to learning. Learning stories reflect the emphasis on the importance of relationships (Ministry of Education, 2004).

Title	Children's Experiences in Early Childhood Settings
Is the KC assessed discretely or integrated with other assessments?	Carr described a continuum of complexity of assessments. Skills and knowledge Skills and knowledge + intent = learning strategies Learning strategies + social partners and practices + tools = situated learning strategies Situated learning strategies + motivation = learning dispositions (Carr et al., 2000, p.5). Therefore, if learning dispositions are being assessed, implicit in this is that other things, for example skills and knowledge and application of these, are included in the assessment.
Are the KC assessed discretely or holistically with other KC?	The learning dispositions are assessed holistically. The Learning Story is a description of actions in an authentic context so more than one learning disposition may be described.
Is there evidence to show a connection between KC and achievement in other areas?	Mostly connections were made to "lifelong learning" rather than specific curriculum areas as presently taught in schools. The Competent Children study includes some competencies that are related to the key competencies. Correlations at age near-5 when the sample was in their final ECE centre were much higher between the key competency related areas (on average, more than r=0.40), than between these areas and literacy and numeracy (on average, less than r=0.20 (Wylie, Thompson, & Kerslake Hendricks, 1996).
How are motivations and attitudes dealt with?	Motivation and attitude are implicit in the domains of learning dispositions.
How is progression in the competency viewed and catered for?	Learning stories are collected over time and show progressions in the stories. There may be increasing competency, and also the competencies may be expressed in contexts of increasing complexity.
How are validity, reliability, and transferability dealt with?	Local validity is demonstrated through complex performances in complex environments and this is enhanced by teacher agreement. Reliability is enhanced by the same process, also involvement by the learner. A trail of evidence is provided of examples of the disposition, shared with the next teacher.
Moderation	Stories are collected by more than one staff member. Discussion is part of the process, with staff debating the interpretation of the story, and what should happen next. Families and the children themselves may also be included in this discussion.
Are the KC "taught"?	At the 'deciding' stage the next steps are identified, actions planned, and interventions made. Experiences are provided that promote the development of the targeted disposition.
What PD was given to teachers?	The research was carried out alongside professional development to support the implementation of Te Whāriki.
What are the perceived benefits of this approach?	Learning Stories: Offer respectful accounts of learners and their learning; Support future learning; and Offer ways of seeing, rather than looking to fit learners into categories. (Hall & Burke, 2003)
Main challenges identified by developers	Further longitudinal research is needed to establish a method to track across time and settings. There is a tension between local validity and beyond local validity. We need new ideas about validity and reliability. Carr {, 2001 #100}, citing Walsh, Tobin and Graue, 1993, suggested they should be replaced by judgements of 'plausibility' and 'trustability'.)

Title	Children's Experiences in Early Childhood Settings
	Assessing more complex learning outcomes may stretch our capacity for assessment to the limit. (Carr & Claxton, 2002)
Could this approach be generalised to other sectors?	Yes—portfolios for older learners can include their own written reflections, and they can select their own pieces of evidence.
What issues do we see with this approach?	Sadler (2002), in discussing dispositions, warned that the learner's goal is also of significance, as it is the goal itself that gives the learning meaning. He suggested that dispositions are highly situational — it is knowledge of success rather than the processes engaged in that provides reinforcement. With this in mind, it seems important to stress that a Learning Story should not only discuss the dispositions/key competencies demonstrated, but also provide a rich description of the context, including what the learner was trying to achieve. Teachers and schools would need to consider processes to manage the time needed to record Learning Stories.

The "New Basics" initiative

Assessment reform is one of three inter-related components of this project. The other two are curriculum and pedagogical change in the school sector (both primary and secondary) for schools in the state of Queensland, Australia. The purpose of the reforms is to transform schooling through the intentional alignment of change in all three areas. Assessment reform is seen as integral to achieving change in classroom practice.

The curriculum changes specify "new basics" which are seen as the multi-literacies necessary for the information age. These subsume but go well beyond the "old literacies" of reading, writing and numeracy. The four "new basics" and their relationships to the proposed New Zealand key competencies are as follows:

• Life pathways and social futures (aligns with aspects of *Relating to others* and *managing self*);
• Multi-literacies and communications media (most closely aligned with *Making meaning*);
• Active citizenship (aligns with *Participating and contributing*); and
• Environments and technologies (aligns with aspects of *Thinking, Making meaning*).

The new assessment instruments are "rich tasks" that assess these multi-literacies in the context of a small range of authentic tasks, to be progressively achieved in a three-year band. There are three such bands with completion of assessment and reporting points at years 3, 6 and 9 (New Zealand years 4, 7, and 10). These bands were deliberately designed to cross transition points so that teachers from different parts of the school system would need to co-ordinate curriculum planning. The tasks were designed to support and encourage the use of "productive pedagogies" in classroom practice. Extensive professional development has accompanied the implementation in a range of trial schools. A number of research projects has been carried out alongside the implementation, with the overall results reported in 2004 (Queensland Government, 2004).

Teachers provide formative feedback to students as they work on the tasks and then make summative judgements on their completion. Detailed rubrics are provided to assist them with their decision-making. However they must then produce a holistic judgement about whether the standard has been achieved, and to what level (pass, merit or distinction). It is stressed that the tasks are "complex" and "multi-faceted" and so "draw on multiple repertoires of practice developed over several years" (Queensland Government, 2004, p.55). Because of this complex nature, the judgement involves "trading off inconsistent performances on a type of holistic assessment against several desirable features" (*ibid*, p. 55) so that a single overall grade can be assigned for reporting. Teachers' decisions have been moderated for consistency, with the greatest number of changes made at the year 9 assessment point.

Although teachers have close involvement in the actual assessment, the initial design of the rich tasks was carried out by small teams of experts and teachers working together. The tasks are sufficiently generic that schools and teachers can choose contexts and task details to meet the learning needs of their own students.

Analysis of key features

Title	New Basics
KC being assessed	Four "new basics" that are in addition to "old basics" of literacy and numeracy (which continue to be seen as important). See above.
Sector	Primary and lower secondary school
Description of the assessment	"Rich tasks" are transdisciplinary and integrate cognitive, cultural, linguistic and social skills with each other and with key learning areas and other disciplinary fields as required. Tasks also link school learning to students' life worlds. They are sufficiently generic that details can be adapted to do this. They are "characterised by complexity" (Queensland Government, 2004, p.55). A specified number of tasks are required to be achieved from within three broad age bands (five tasks for years 1–3, seven tasks for years 4–6, eight tasks for 7–10). Tasks are intended to be spread across a band and not concentrated into the last 6 months. They should be part of the teaching and learning programme. Tasks are designed to be sufficiently different in what they assess that all must be used to cover all curriculum components.
Nature of the assessment	Essentially summative at the stage when reported, but tasks are designed to be ongoing within day-to-day learning environment until achieved. Holistic judgement — parts of the task are not assessed separately. Inconsistent performances are traded off across criteria to assign an overall grade. Because of the rich nature of tasks, evidence for any one task comes from "several products, in different mediums" (Queensland Government, 2004, p. 79).
How are the assessment results reported and recorded?	Results are placed on a scale A-C or A-E as specified by standards/criteria developed. "Students are required to keep records of evidence of progress and performance" (Queensland State Education, 2000, p.85). Teachers maintain ongoing report cards for formative purposes, until evidence is collected to show that the task has been achieved.
Why was the assessment developed?	To better match learning demands for twenty-first century realities (including challenges of identity, increase in number of possible learning pathways, necessity for multi-literacies). To align assessment as a powerful "message system" with "productive pedagogies" for classroom learning and with curriculum change (for example, content reduction, integration).

Title	New Basics
How was the assessment developed?	Initial design was informed by policy research — both wider literature (for example, about knowledge economy, lifelong learning) and Queensland's School Reform Longitudinal Study (QSRLS). Tasks were developed by 'expert panels" of teachers and educators. The intention (Queensland State Education, 2000) was to document developmental issues as case study schools trialed process. In 2004 there was extensive reporting of trial school students' achievements in comparison to students in non-trial schools. A key finding is that students have not been disadvantaged in achieving "old basics", while also adding new basics to their learning repertoire (Queensland Government, 2004).
Is the KC assessed discretely or integrated with other assessments?	Integrated. It is built from four clusters of "New Basics" competencies (see above) but tasks were designed to provide a "principled means for the selection of useful and valued knowledges and skills from Key Learning Areas (KLAs), traditional subject areas and interdisciplinary fields" (Queensland State Education, 2000, p.39).
Are the KC assessed discretely or holistically with other KC?	Holistically — see above. Also this explicit comment regarding "old" basics is relevant: "literacy and numeracy are not stand-alone activities or skills: they are always integrated with each other and with new repertoires of practice" (Queensland State Education, 2000, p.46). It was advocated that the design, with the necessary teacher professional development, helps foster a "consilience" which refers to "fluency across boundaries" as necessary for clarity of world view in new conditions" (Queensland State Education, 2000, p.92). A key finding was that students in trial schools perceived changes in teachers' pedagogy in areas of connectedness, supportive classroom environment and recognition of difference, but not in the area of increase in intellectual quality of their learning (Queensland Government, 2004).
Is there evidence to show a connection between KC and achievement in other areas?	38 schools completed the full four year trial. Many others were involved at some stages only (Queensland Government, 2004). Evidence shows there was no adverse impact on basic literacy and numeracy. Students in trial schools were compared to those in matched non-trial schools for PISA type tests. Multiple regression tests showed just 3–18% of variance in 'New Basics' achievement could be explained by performance in standardised literacy and numeracy tests — i.e. these rich tasks do assess something other than basic literacy and numeracy (Queensland Government, 2004).
How are motivations and attitudes dealt with?	Tasks are set in contexts of students' life worlds, with a focus on both old and new knowledges. NB: Students performed beyond many teachers' expectations.
How is progression in the competency viewed and catered for?	Tasks model progression across age levels. A Grading master provided for each rich task helps teachers establish progression in meeting more advanced standards within an age level. Grading masters are task specific but sufficiently generic to accommodate differences in context, how students tackle tasks etc. Each grading master has several "poles" that describe achievement at different levels (a bit like Achieve/Merit/Excellence differentiation but much more task specific). Teachers are supposed to evaluate each pole before making an overall holistic judgement. Research showed that possibly as few as 23% of sampled teacher judgements were actually made holistically. Others were decided using impression marking or some mechanical combination of individual poles. The researchers concluded that making holistic judgements is challenging for teachers but moderation helps (Queensland Government, 2004).

Title	New Basics
How are validity, reliability, and transferability dealt with?	Transparency of process. Intention to work backwards from tasks to assessment criteria is of interest, given that this has been recommended in other work on assessment of complex performances (Delandshere & Petrosky, 1998). Tests were conducted of reliability of teachers' judgement of richness of tasks and found this can be reliably consistent, especially after professional development in making judgements (Queensland Government, 2004). A separate research project assessed ability of tasks, as used by teachers, to deliver evidence that demonstrated richness sought (i.e. checking that they didn't become just another project as could happen if school reform subverted to business as usual). It was found the tasks as used did deliver the richness sought .
Moderation procedures	Developers reached a shared understanding of meaning of standards via moderation of selected portfolios of student tasks. Intention at outset was that teachers would be involved in discussions to reach shared understandings then verify own judgements through consultation with teachers from another school (Queensland State Education, 2000, p.89). When the intention to moderate was put into action some teachers reported receiving mixed messages from moderators but others found the interactions with other teachers very useful (Land, 2004, p.56) Research on the extent of change in grades post-moderation showed more changes at year 9 (circa 74%) than at years 3 and 6 (circa 35–45%), and a predominant effect of grades being moved down rather than up (see Queensland Government, 2004, p.80 for graphs of this finding). NB: The researchers comment, "The moderation strategy delivered high-level professional development for teachers, not only in assessment but also in discipline-specific knowledges" (Queensland Government, 2004, p.80).
Are the KC "taught"?	Yes — they are explicitly scaffolded. A stated intention is to align pedagogy, curriculum and assessment. It is recommended that teaching for and inclusion of tasks is across all three years of each age span, not just near the end.
What are the perceived benefits of this approach?	Teachers' views about assessment were sampled using a New Zealand developed instrument. (Developer was Dr. Gavin Brown, University of Auckland — see 2004 report, pg.53.) New Basics teachers were significantly more likely than New Zealand teachers to agree that assessment determines standards students have reached and that it improves learning. They were also less likely to agree that assessment is for school accountability. Schools are able to plan together across years 6–9 transition. Indigenous students are able to demonstrate their talents. It was implied that School Based Curriculum Design could be supported. "Teachers are finding solutions for student connectedness, although not initially exploiting this aspect of rich tasks. There are some local solutions that favour the aspirations of students and communities" (Queensland Government, 2004, p.60). Supportive teachers saw the initiative as getting emphasis back to the business of teaching (Queensland Government, 2004, p.78).
Main challenges identified by developers	As might be expected given the scope of the changes, there was a decrease in staff satisfaction because of increased workload. Staff turnover impacted on schools — for example understanding of joint curriculum planning was lost. There was some resistance from secondary teachers with a strong subject loyalty (fear of losing subject specificity and student numbers in their classes). It was a challenge to get teachers to see connectedness beyond the local community. (Technology needs to be available and reliable to go wider than this.) Assessment of group work was an issue. ESOL students make slower progress on tasks. It was necessary to map the whole three-year programme backwards from tasks (but this

Title	New Basics
	was seen as promoting collegiality and dialogue). It was also necessary to plan across schools at Years 6–9 transition point (but this was deliberate).
	There was a decrease in overall parental satisfaction with curriculum/learning in the trial secondary schools (but an increase in satisfaction in the trial primary schools).
	Another problem was parental misunderstanding or misreading of reports.
Could this approach be generalised to other sectors?	This approach already applies across most of the school sector in Queensland.
	Conditions that made uptake as intended happen in individual schools:
	Supportive school culture;
	Professional learning community of staff; and
	Strategic school leaders who encouraged risk taking and innovation.
What issues do we see with this approach?	Cost — the developers comment on this and say it could be reduced with an initial outlay for development of exemplars that could be delivered electronically.
	Scaling up — this was a limited trial. The 2004 report gives suggestions for expansion in the Queensland context.

PISA: Problem Solving

PISA assessments are OECD funded tests of competency in English literacy, mathematics literacy, science literacy, problem solving and belonging. Unlike the other case studies used for this background paper, they assess these competencies strictly for accountability purposes and are not intended to provide feedback to individual students. The various tests (each competency is assessed separately) are intended to provide the governments of all participating nations with comparative measures of the success of their education programmes and policies. As appropriate to this purpose, the assessment methods used fall totally within the empirical-analytic paradigm and adhere to strict psychometric parameters. While all the tests are developed in broadly similar ways, PISA Problem Solving has been chosen for this case study because it seems to incorporate elements of several of the New Zealand's key competencies. Brewerton (2004b, p.17) also notes an intention to extend sampling of *adult* competencies to include problem solving in the near future.

The PISA tests are of particular interest in the context of this paper because, notwithstanding their empirical-analytic orientation, they are standards-based. Set at age 15, three levels of problem solving are described. The "standard" for each of these levels is a rich description of a cluster of characteristics (*see* Section Six). These standards were developed through a rigorous process of theoretical research, combined with the practical expertise of a range of international experts in educational problem solving. The draft standards were then used to devise trial tasks and the range of responses to each task, from students in different countries, was analysed using standard psychometric tools. In this way the statistical validity of both the standards and a range of prototype tasks was established.

Tasks that can generate data seen as valid and reliable for empirical-analytic analysis are inevitably of a limited range of types, and all are pencil and paper tests that can be administered under examination conditions. In this respect, they assess something other than the "complex performance" demanded within other assessment paradigms. However efforts are made to set each task in a context that could be seen as "authentic" for a wide range of students, in that the problem to be solved might well be one they could face in real life. Tasks are marked according to rubrics validated at the trial stage and inter-marker moderation is carried out. The rubrics are devised to produce a range of test scores that translate students' actual achievements to one of the three levels of standard described.

Analysis of key features

Title	PISA: Problem solving
KC being assessed	Problem solving Tasks are designed to collect evidence of ability to understand, characterise, represent and solve the problem, then reflect on the solution and communicate it. Tasks are designed so that all or most of the above are integrated (OECD, 2004, p.50). Emphasis is on analytical reasoning and being systematic (OECD, 2004, p.51). Tasks incorporate elements of New Zealand's key competencies for thinking, making meaning and, depending on the nature of the problem, managing self.
Sector	Secondary school — age 15, towards the end of compulsory schooling.
Description of the assessment	Three types of problem solving capabilities are assessed: Making decisions under constraints; Evaluating and designing systems for a particular situation; and Troubleshooting a malfunctioning device or system, based on a set of symptoms (OECD, 2004, p.16). Tasks are set in contexts that "could conceivably occur in a student's life or, at least, be situations that the student can identify as being important to society" (OECD, 2004, p.26). The method of solving the problem should not be readily apparent so the student is required to exhibit a degree of flexibility in ways they access, manage, evaluate and reflect on information. Tasks draw on knowledge across several school subjects so require a degree of integration.
Nature of the assessment	Summative — paper and pencil tests, each two hours long. There is a mixture of open constructed responses, closed constructed responses, and multiple choice questions. There is some use of a unit format, where a series of questions of increasing complexity is clustered around the same stimulus material.
How are the assessment results reported and recorded?	Marking is done by trained markers, using a coding rubric. Marks from rubrics are converted to a four level proficiency scale (see progression).
Why was the assessment developed?	As a measure to compare the problem solving abilities of different students, in different nations, at the end of compulsory schooling.
How was the assessment developed?	International panels of experts worked co-operatively to develop and analyse the tasks.
Is the KC assessed discretely or integrated with other assessments?	Discretely — the specific focus of each task is on problem solving — but see notes on links to achievement in other areas below.
Are the KC assessed discretely or holistically with other KC?	Separate components of the key competencies are integrated with each other.

Title	PISA: Problem solving
Is there evidence to show a connection between KC and achievement in other areas?	The test is written, so basic reading literacy is needed to be able to demonstrate problem solving. Tasks use everyday language and also diagrams, maps etc. Any manipulation of figures is limited to "simple mathematical operations" (OECD, 2004, p.51). An exploratory factor analysis was carried out that showed problem solving ability is somewhat linked to mathematical reasoning (and specifically analytic reasoning) and to reading ability (OECD, 2004, p.53) but neither of these explain more than 10% of variation. The lowest correlation was between science and problem solving, but tasks did not require specific science knowledge. Generally a high level of correlation for all three shows students who do well in problem solving are also achieving well in other areas. The developers suggest students who do well in problem solving but not in mathematics could achieve better results in maths with better teaching (OECD, 2004, p.57).
How are motivations and attitudes dealt with?	Not explicitly measured in this test
How is progression in the competency viewed and catered for?	A four point scale was created: Level 3: Reflective problem solvers; Level 2: Reasoning, decision making problem solvers; Level 1: Basic problem solvers; and Below Level 1: Weak or emergent problem solvers. NB: Each level describes an overall cluster of attributes, not just one or two (see Section Six). Analysis of features of different types of problem solving, in terms of goals, processes involved and possible sources of complexity was used to determine the score attributed to each task (see OECD, 2004, p.29). Students could gain full or partial credits for some items. The overall score gained determined the level achieved.
How are validity, reliability, and transferability dealt with?	Validity — tasks are placed in real-life contexts. The nature of problem solving skills needed was carefully analysed and described. Reliability — all tasks were extensively trialled. Analysis of item performance after the tests indicated difficulty ratings had been accurately predicted when tasks were designed (OECD, 2004, p.101). Transferability — analysis of factors involved could help with analysis of other tasks.
Moderation procedures	Marker reliability studies were used to moderate between markers in different countries. More difficult tasks were marked by up to four markers, and sub-samples of these marked in at least two countries, then results compared. An inter-country reliability study was also completed on a sub-set of items.
Are the KC "taught"?	Not relevant — this is a summative measure of outcomes from many different school systems.
What PD was given to teachers?	Not relevant — teachers were not involved.
What are the perceived benefits of this approach?	The assessment is valid across a wide range of settings — different nations, different schools. Assessment can be managed by traditional examination conditions.
Main challenges identified by developers	None beyond those already addressed above.
Could this approach be	Possibly to tertiary sector. This approach is not compatible with the socio-cultural framework adopted for early childhood education.

Title	PISA: Problem solving
generalised to other sectors?	
What issues do we see with this approach?	Assessment is limited to the accountability paradigm (Section Three) so there is no direct impact on teaching and learning, or on support for lifelong learning, unless a new type of feedback mechanism is developed.

Vocational Assessments of key competencies in Australia

Defining the Australian key competencies

In Australia in 1992 the watershed report by the Mayer Committee (1992, cited in Curtis & Denton, 2003) set the scene for current work on generic skills and competencies in the Vocational Education and Training (VET) sector. The Mayer Committee's conception of key competencies centred on notions of generic employability skills, which involved the application of knowledge in an integrated way in work settings. The Mayer Committee requirements for key competencies were that they:

- were essential for employment preparation;
- were generic rather than industry specific;
- equipped individuals to participate in a range of social settings;
- involved the application of skills and knowledge;
- were able to be learned; and
- were amenable to credible assessment. (Curtis & Denton, 2003, p. 17).

Seven competencies were endorsed by the committee. These have some similarities to the essential skills outlined in *The New Zealand Curriculum Framework* (Ministry of Education, 1993). The competencies were:

- collecting, analysing and organising information;
- communicating ideas and information;
- planning and organising activities;
- working with others and in teams;
- using mathematical ideas and techniques;
- solving problems; and
- using technology (Curtis & Denton, 2003, p. 17).

Values and attitudes were excluded from the definition and assessment of the key competencies as the Mayer Committee considered that these dimensions did not fit within their assessment criteria in that they were not able to be "delivered, assessed and reported upon credibly" (Curtis & Denton, 2003, p. 22). But Bryce (2003) notes that the conceptualisation of key competencies by the Mayer Committee is underpinned by an assumption that what is being observed and assessed is an integration of knowledge, abilities, skills, understandings, and attitudes.

Along with the discussion and elaboration of the concept of key competencies, the Mayer Committee had several other mandates, specifically to operationalise the key competencies for the school and training sectors, specify levels of achievement, and make assessment and reporting recommendations. The committee specified that each competency should have three levels for learners to be assessed against. Curtis and Denton (2003) note that these three levels reflect an increasing degree of metacognitive control.

Case study: Validated self-assessment of problem solving at Torrens Valley TAFE

Denton (2003) charts the development of key competency assessment at Torrens Valley TAFE in Australia since 1990, and presents a model for the assessment of problem solving which is based on the Mayer Committee recommendations for assessment of the key competencies. This model is further described by Curtis (2003), and a validation study conducted by Curtis and Denton (2003) is also available. The model is discussed in other literature as one of the most comprehensive and advanced in the Vocational Education and Training (VET) sector in Australia (Clayton et al., 2003).

Although this model is developed from the Mayer key competencies framework, which locates its underpinnings within the interpretive or empirical-analytic paradigms, the model of assessment sits mostly within the critical-theoretic framework.

The key competency assessments were developed by staff at Torrens Valley TAFE to give more explicit recognition to the fact that students were developing enterprise and self-management competencies through the learning experiences offered on TAFE courses. The essence of the model is that key competencies are assessed by "validated self-assessment" which is grounded within a student-centred pedagogy. The assessment has a "prove and improve" teaching function and aims to support students so that they can develop understanding about the key competencies and learn how to demonstrate that they have the competencies necessary for their effective performance in the workplace. Students are given material about the key competencies and are able to select a key competency they wish to be assessed against. They select evidence from existing course work assessments of technical skills to perform a self-assessment of their ability against a set of criteria. Evidence can be presented in any form (written, verbal, demonstrated). This evidence is validated against the key competency criteria during a discussion between the tutor and the student. Students are given a certificate stating which key competency they had been assessed against and which one of three levels they have attained.

The case study below reports on the current state of play for this assessment system, which has been developed over the last 12 years.

Analysis of key features

Title	**Validated self-assessment of problem-solving as part of a Electronics and Information Technology Program at Torrens Valley TAFE.**
Key competency being assessed	A range of key competencies is included. Assessment of the problem-solving key competency is described in this case study as outlined in the validation study conducted by Curtis and Denton (2003) and also as described in Denton (2003) and Curtis (2003). This problem-solving key competency aligns with aspects of the New Zealand key competencies: *Managing self* and *Thinking.*
Sector	The TAFE (Polytechnic) sector in Australia.
Description of the assessment	Students are provided with a range of induction materials about the key competencies. Students select a key competency they wish to be assessed against. Students perform a self-assessment of their ability in a key competency using an Assessment Sheet provided by the TAFE. It is students' responsibility to then identify evidence to address the criteria for their chosen key competency performance level. They select evidence from existing course work assessments of "technical skills". Evidence can be presented in any form (written, verbal, demonstrated). The evidence is validated against the key competency criteria by their tutor.
Nature of the assessment	Formative — for the benefit of the learner — so they can "prove and improve". Summative — for the learner to demonstrate competence to employers and for the TAFE to report on.
How are the assessment results reported and recorded?	Students receive a Statement of Attainment (certificate), which lists the key competency and the performance levels achieved. This can be shown to employers. The students' evidence and portfolio of assessment is stored on the TAFE computer system.
Why was the assessment developed?	The key competency assessments aim to give more explicit recognition to the fact that students are developing key competencies through the learning experiences offered on TAFE courses. In particular, the assessments were designed to recognise the enterprise skills students were developing, such as thinking laterally, problem solving, self-direction and taking responsibility, working individually as well as part of a team, and communicating effectively in a range of forms. Making the assessment explicit shows students these skills are valued and enables the key competencies to be certified.
How was the assessment developed?	The key competency assessments were developed within an Employability Skills Framework using the Mayer Committee recommendations. Employers were consulted about their skill requirements. Theoretical approaches were explored to develop a coherent description of problem solving. Five major components of problem solving were identified (defining the problem, planning an approach, carrying out a plan, monitoring progress, and reflecting on the result). For each of the five, a set of indicators was developed. For each indicator a set of performance levels was described. (These provide the basis against which student evidence is compared.)
Theoretical underpinnings	A review of behaviourist, cognitive, and situative/sociohistoric approaches to problem solving was undertaken. Cognitive models were mostly drawn on to develop the five major components of problem solving.
Is the key competency assessed discretely or integrated with other assessments?	Integrated. Key competencies are "performed" as part of the course work, which is used to assess technical/vocational skills, but are assessed using separate criteria. Although technical assessments are mandatory, the key competency assessments are voluntary. The TAFE is considering also making the key competency assessments mandatory.

Title	Validated self-assessment of problem-solving as part of a Electronics and Information Technology Program at Torrens Valley TAFE.
Are the key competencies assessed discretely or holistically with other key competencies?	Each key competency is assessed discretely from other key competencies — but it is conceivable that students could select the same task to demonstrate evidence for more than one competency.
Is there evidence to show a connection between the key competency and achievement in other areas?	The validation study of the problem solving assessment that was completed by Curtis and Denton (2003) showed a significant association between attainment of the key competency and achievement on other course assessments.
How are motivations and attitudes dealt with?	As the assessment is voluntary motivation is implied for those who elect to do it! The validation study by Curtis and Denton (2003) did not show a clear association between attitude and attainment of the problem solving key competency or attitude and achievement on other course assessments.
How is progression in the competency viewed and catered for?	The problem solving key competency is framed and reported as a series of three performance levels which have associated criteria (indicators). Each level must be assessed and validated. Achievement of a higher level does not imply that the lower level can automatically be achieved. The levels reflect an increasing degree of meta-cognitive control. Level 1: Solves the problem by following an established procedure. Level 2: Solves the problem by selecting from several alternative established procedures. Level 3: Creates a new procedure or adapts an existing procedure to meet the demands of a task. The five level Structure of the Observed Learning Outcome (SOLO) taxonomy (Biggs and Collis 1982, cited in Curtis and Denton 2003) was used to describe performance levels for each indicator for the five components of problem solving. Not all five levels were used for each indicator. Only those which were deemed to be amenable to reliable judgement were selected.
How are validity, reliability, and transferability dealt with?	For students to gain recognition for a key competency performance level they must have demonstrated and been assessed for the key competency twice in different contexts (to show that the assessment is reliable and the learner can transfer the skill). More than one method of collecting evidence is able to be used (e.g., written, oral, observation). The tasks used for evidence are authentic and integrated into the learners' programme of study. More than one person makes the assessment. The student collects the evidence and this is validated by the tutor. A comprehensive validation study of the problem solving assessment was completed by Curtis and Denton (2003).
What are the moderation or validation procedures?	Standard instruments are used to assess levels of performance, but the judgement of the tutor is not moderated.
Are the key competencies	No: the TAFE uses an "assessment as learning" approach. Students are expected to learn about and develop the key competencies as part of the self-assessment and evidence

Title	Validated self-assessment of problem-solving as part of a Electronics and Information Technology Program at Torrens Valley TAFE.
"taught"?	collection process. The assessment instruments are viewed as "developmental tools".
What PD was given to teachers?	The TAFE has a key competencies co-ordinator who supports staff. Processes and resources are in place to ensure staff are inducted into their role as a validator.
What are the perceived benefits of this approach?	It is seen as practical and workable, as the key competency assessments are integrated with existing authentic tasks. It empowers students to have ownership over learning about the key competencies and developing their next learning steps. It provides students with the skills to be able to identify and describe their development of key competencies to employers (i.e., skills they need when they are trying to find employment).
Main challenges identified	Finding a practical and workable framework for assessment. The voluntary nature of the assessments.
Could this approach be generalised to other sectors?	Yes. The collection of a portfolio of evidence could also be undertaken by students at other levels of the school system and could potentially include information from sources other than school-based activities.
What issues do we see with this approach?	The framework has been developed within the VET sector. To support measurement of change over time further work to make connections to the secondary sector would need to be undertaken. The key competencies are "caught" within a self-directed framework, not "taught". Younger students would require an approach which involved more active teaching. Inter-tutor reliability is not evaluated.

Comparison of the VET key competencies assessment with one current New Zealand vocational model

A brief comparison with the assessment procedures for the *National Certificate in Employment Skills* (NCES), gives some indication that a framework with similarities to the VET model is already in existence in the New Zealand school and polytechnic sector. Only sections of the school or tertiary student population, usually low achievers or those targeted for transition classes, are currently being assessed for the NCES. The similarities and differences between the NCES assessment procedures and the VET model are discussed below.

One assessment approach taken by NZQA is the use of single unit standards to assess specific skills. Kelly (2001) notes that "the standard dealing with cross-cultural communication in the work place has the highest uptake of any of the over 15,000 registered Standards" (p. 15). This standard is part of the *National Certificate in Employment Skills* (NCES), which assesses core literacy and numeracy skills and attributes which are aligned to the key competencies (such as problem solving, cross-cultural communication, assertiveness, and team work).

Standards in the NCES are mostly assessed by collecting two main sources of evidence: information from observational checklists and from work sheets completed during students' course work. In this way the skills are explicitly taught. To deal with issues of transferability and

reliability, the observational evidence is usually required to be collected on several occasions or in three different contexts. For example for *Standard 1304: Communicate with people from other cultures*, candidates are required to communicate with people from three cultures. For *Standard 3503: Participating in a team or group to complete routine tasks*, candidates organise a witness to observe and rate their work in a group on two separate occasions. Evidence is collected by registered assessors and moderated by NZQA. Single standards are reported as a single level of competence using an achieve/not achieve system.

There are some similarities between the VET model and the NCES: to a certain extent the process of finding out about and developing the skills is managed by the trainee within their own timeframes; and trainees are located in apprenticeship on-the-job situations where they develop the skills.

Unlike the VET model there is only one level of performance and trainees do not have input into the final assessment decision. Also, unlike the VET model, this decision is moderated.

The National Institute for Literacy (NIFL) "Equipped for the Future" (EFF) Initiative

This initiative was undertaken to develop a standards-based assessment system that would both support and give credentials to the basic literacy learning of adult Americans in need of foundation learning. A feature of the project development was the extensive involvement of adult learners and literacy tutors from many different American states and types of learning programmes (both institutional and work based) over a period of some years. It was felt that shared ownership would be a key to the successful development and uptake of the system. Employers and others in the community who would use the assessment results were seen as important stakeholders, along with tutors and adult learners.

Before standards development began an extensive mapping exercise led to the construction of role maps that describe the **application** of adult literacy in the home and family, in the workplace, and as citizens and community members. The three "role maps" constructed from this analysis — one for each of the three types of settings—then formed the basis for the identification of sixteen "equipped for the future" (EFF) literacy standards. These standards are organised into four clusters (see table below) that have some correspondence with key competencies for New Zealand schools. For example "lifelong learning skills" looks similar to "managing self" and "interpersonal skills" is similar to "relating to others".

Once described, the EFF standards were used by tutors and adult students from many programmes, and information about the nature of evidence that could be collected for each was solicited and centrally collated. During this process four "key dimensions" of performance were kept in mind:

• structure of knowledge base;

- fluency of performance;
- independence of performance; and
- range of conditions for performance.

The empirical data collected was combined with relevant insights from the research literature to develop three-level judgement statements to support each standard. The 16 resultant EFF draft standards (as of April 04) are richly detailed because the intention is to support transparency of learning progress and evaluation — for both tutors and students. However they are also sufficiently generic that tutors may use tasks relevant to the contexts and needs of their students. A paraphrased example is included in Section Six. If this approach were to be translated to the New Zealand initiative, several aspects of each competency would be defined and elaborated separately, whilst still keeping an overall complexity of performance in mind.

The standards were designed for assessing literacy learning in the context of authentic tasks of personal relevance to the adult learners. A holistic judgement is made about the overall level reached by the student. Next learning steps are clear because of the descriptive nature of the progressions. There was a specific intention to make the first level sufficiently broad to acknowledge the achievements of any adult who has begun on a literacy learning journey. It was seen as a matter of principle that all learners should see places for themselves on the learning continuum and that they would be supported to continue this learning journey. Self-assessment is also seen as important.

This assessment initiative provides leeway for different programmes to decide how to document and report on student performance. In one early case study portfolios were used to collect evidence of performance levels, and a "career passport" that documented specific achievements and "marketable skills" was presented to students at graduation (Stein, 2000, p.84).

Analysis of key features

Title	National Institute for Literacy (NIFL) Equipped for the Future (EFF)
KC being assessed	There are 16 key literacy skills in four broad groupings, as these are applied in contexts of work, family life and community involvement: Communication skills (read with understanding; convey ideas in writing; speak so others can understand; listen actively; observe critically); Decision-making skills (solve problems and make decisions; plan; use mathematics to solve problems and communicate); Interpersonal skills (co-operate with others; guide others; advocate and influence; resolve conflict and negotiate); and Lifelong learning skills (take responsibility for learning; learn through research; reflect and evaluate; use information and communications technology).
Sector	Tertiary — learners are adults in many different kinds of foundation programmes across a wide range of American states.
Description of the assessment	Literacy skills are demonstrated in the execution of authentic tasks of personal relevance to the adult learners. They collate evidence of achievement to be judged against the detailed descriptors for each of 16 "Equipped for the Future" (EFF) standards.
Nature of the assessment	Both formative and summative. The adult learners are involved in collecting the evidence.
How are the assessment results reported and recorded?	This is left up to the programme designers, but it must be specific about what students know and can do.
Why was the assessment developed?	To address low levels of literacy in some sectors of the American population whilst simultaneously supporting the development of lifelong learning skills.
How was the assessment developed?	Through extensive sector consultation, collation of empirical data on levels of adult literacy achievements in the 16 identified areas, and use of relevant theoretical research on literacy development.
Theoretical underpinnings	Extensive — and included as background information with each standard. Overall development was strongly informed by research on lifelong learning, formative assessment, assessment of complex performances, and institutional learning (to achieve change in tutor practices).
Is the KC assessed discretely or integrated with other assessments?	Each key competency is assessed in an authentic task where it would be integrated with other skills as relevant. An underpinning principle of the initiative was that the KC assessment would be sufficiently generic to be compatible with any assessment of other competencies required in the context.
Are the KC assessed discretely or holistically with other KC?	Each is described in some detail as a discrete entity — but presumably some tasks, being complex, would support demonstration of more than one KC.
Is there evidence to show a connection between KC and achievement in other areas?	Not addressed in any of the reports that we have found.

Title	National Institute for Literacy (NIFL) Equipped for the Future (EFF)
How are motivations and attitudes dealt with?	Learners have maximum involvement in assessment and decision-making. Assessment tasks are designed to be personally meaningful to their lives and are discussed with the learner. At the first level, considerable support to achieve the task is provided.
How is progression in the competency viewed and catered for?	Horizontal progression is specified in rich detail, at three levels, for each standard (*see* the example in Section Six).
How are validity, reliability, and transferability dealt with?	Validity comes from the assessment of a complex performance. Reliability comes from transparency of information used to make judgements. Both aspects were enhanced at the outset by the extensive collaborative involvement of the sector in the initial design and use of the standards.
Moderation procedures	Unsure — have not seen any reports of this.
Are the KC "taught"?	Yes — achievement of the 16 standards is integral to each programme.
What PD was given to teachers?	Many tutors were maximally involved in the system design and implementation as co-researchers with the assessment experts.
What are the perceived benefits of this approach?	Authenticity, credibility, manageability, positive impact on teaching and learning, development of lifelong learning skills.
Main challenges identified by developers	Unsure — have not seen any reports of this.
Could this approach be generalised to other sectors?	Yes. DeSeCo have shown interest in adopting aspects of this work (Texas Centre for the Advancement of Literacy Learning, 2003).
What issues do we see with this approach?	Development of standards would be expensive and take considerable time. Literacy learning is a relatively bounded context — ways to link to the key competencies and curriculum learning areas would need to be debated.

Discussion of the case studies

Most commentators note that it is important to realise that the attributes which contribute to a unified *performance* are interwoven and not discrete (Carr & Claxton, 2002; Curtis & Denton, 2003; Kearns, 2001). At any one time a learner may be drawing on two or three of the competencies to deal with the complexities of a situation. This interwoven nature has implications for how the key competency model is presented, as well as how the competencies are taught and assessed. These implications are discussed now, with reference to the idea of competency as a complex performance, and to the solutions arrived at in the various case studies.

Should key competencies be assessed with each other or discretely?

The case studies offer different perspectives on this question.

- The New Basics and Learning Stories case studies provide models for the holistic assessment of competencies, as these are demonstrated in authentic, complex performances.
- The VET and NIFL case studies offer models within which the key competencies are performed in an integrated manner but are assessed discretely.
- The PISA case study models an approach where the competency is both performed and assessed discretely.

All except the PISA case study demonstrate an assessment approach that potentially aligns well to the socio-cultural theoretical framework for the key competencies outlined in Section Two. In all four cases students perform assessment tasks in real settings where they must interact with other people and with cultural artefacts. This suggests that the question of whether to assess competencies discretely or separately could be decided in a range of ways without compromising the intention of their introduction into the national curriculum.

Commenting on the VET model, Curtis and Denton (2003) suggest that although key competencies are not used in isolation, their application is context dependent, and holistic assessment may not address the subtleties of each key competency.

Should key competencies be assessed with academic or technical skills?

Commentators appear in general to agree that key competencies are best developed through authentic tasks that address curriculum goals. However, whether academic or technical skills should be assessed *at the same time* as key competency goals is an area of debate. Carr and Claxton (2002) suggest that capabilities (academic goals) and dispositions should be kept distinct from each other. In contrast Sadler (2002) argues that dispositions are situational and relate to particular goals or motivations and therefore suggest that assessments need to be holistic and incorporate two key dimensions: assessment of the dispositions and also the "product" or the technical/academic skills being developed. Carr and Wylie (2004) recommend that school entry level assessment of key competencies be placed within "expanded" literacy and numeracy contexts, to address manageability issues for new entrant teachers.

Reynolds and Mackay (1997, cited in Clayton et al., 2003) describe four ways key competency assessments can combine academic knowledge or technical skills with competencies:

1. Inferred (evidence for key competencies is inferred from academic performance).
2. Parallel (key competencies are taught and assessed separately from academic goals).
3. Separate tasks (key competencies assessment tasks are separate from academic assessment tasks and are only for one key competency).
4. Integrated (inference for key competencies is drawn across groups of academic subjects).

Again all the case studies except PISA model ways in which assessment of key competencies can be integrated with assessment of other aspects of the curriculum. In the case of the early

childhood case study the intention to develop the dispositions that are inferred from children's performance *is* the main curriculum focus. Similarly the NIFL curriculum is largely shaped by the literacy demands revealed in the first instance by the role map analysis and then enshrined in the 16 EFF standards. It is food for thought that in both cases the nature of the intended curriculum integration between competencies and other aspects of learning has been carefully designed and made overt.

Curtis and Denton (2003) report that the integrated approach is the most common in the VET sector in Australia. They say it reinforces the importance of key competencies as well as their connection with academic or technical competencies. As we have seen above, key competencies are performed holistically as they are embedded in existing vocational "technical" assessments but are explicitly assessed separately. However, Curtis and Denton note that assessments of key competencies need to be clearly distinguishable from other assessments of vocational or academic skills.

Other New Zealand research has pointed to an example of the first model on the above list. Kelly (2001) notes that in New Zealand some industries have subsumed key competencies within technical standards. The result is that these key competencies must be inferred as having been acquired as the technical competencies are acquired. She notes however that this approach does not lend itself to being able to report on achievement of the key competencies.

What is the relationship between key competencies and core academic or technical skills?

The discussion of whether key competencies can be inferred from academic performance raises further questions about the nature of relationships between competencies and other academic or technical learning.

Most commentators suggest that any assessment framework needs to encompass or consider both academic and key competency goals (Brewerton, 2004a; Curtis, 2003; Kress, Norris, Schoenholz, Elias, & Seigle, 2004; Queensland State Education, 2000). The New Basics project was deliberately designed so that the integrated holistic nature of the key competencies would drive an imperative for curriculum integration. In part this was related to an intention to align "message systems" (*see* below) so that students are taught, and hence have the opportunity to learn, what will be assessed. But there is more to it than that. Reviewing the connection between curriculum standards and social and emotional learning, Kress, Norris, Schoenholz, Elias, and Seigle (2004) report that the guidelines for developing curriculum materials suggest that:

> …academic and SEL goals [be] unified by a comprehensive, theory-based framework that is developmentally appropriate (p.139, Elias et al., 1997, cited in Kress et al., 2004).

As already noted, both the Learning Story and NIFL/EFF case studies are examples of this approach, albeit quite different from each other. While the developmental basis is less clear in the

literature we have found, the New Basics initiative also centred on a carefully designed framework of rich tasks.

Some difficulties caused by a lack of such alignment are discussed in the literature. For example, Janinski (1996 cited in Kearns, 2001) notes that a "diverse understanding" of the Mayer key competencies in the VET sector in Australia could be indicative of a lack of understanding of the relationship between key competencies and vocational competencies.

In New Zealand, taking the earlier iteration of competencies as essential skills, we could predict that many teachers might perceive a problematic relationship between competencies and other aspects of the curriculum. How, for example, does numeracy interact with the Mathematics curriculum? Or literacy with any curriculum other than English? "Relating to others" has considerable overlap with, for example, the Health and Physical Education strand "Relationships with other people" as shown in the matrix of progress indicators in the exemplars.[4] "Thinking" has overlap with the characteristics of learning in technology as expressed through the decision-making and discernment matrix of progress indicators in the technology exemplars.[5] There are many other examples we could have included. We also see a danger that, in the absence of a clearly understood theoretical framework, "making meaning" could be interpreted as simply understanding content in relation to a particular curriculum subject.

One approach to creating a clear sense of coherence between the key competencies and other curriculum objectives could be to map the fit between existing curriculum strands and the key competencies. This exercise would need to address questions such as: "What is already being taught and assessed?" and "What is not taught or assessed?" It would also need to ask whether the current curriculum focus was compatible with the intentions of the key competency initiative or whether modifications would be necessary. The debate about what constitutes knowledge is referred to in section two.

Kress, Norris, Schoenholz, Elias, and Seigle (2004) provide one model of such a mapping exercise to show the overlap between curriculum standards and opportunities for social and emotional learning. Another local example of a mapping of essential skills against Learning Area projects is shown for the inquiry-based integrated programme at Kuranui College described by Bartlett (2005).

Teaching key competencies: aligning message systems

Teaching, enactment of the curriculum, and assessment practices are intertwined, and any proposed key competency model needs to take this interconnectedness into account. The New Basics case study (Queensland State Education, 2000) is a well-documented example of how this might be approached. It is based on a premise that:

---

[4] http://www.tki.org.nz/r/assessment/exemplars/hpe/matrices/matrix_rwop_e.php

[5] http://www.tki.org.nz/r/assessment/exemplars/tech/matrices/index_e.php

> Improved student outcomes require a systematic, principled and practical coordination of the message systems of curriculum, pedagogy, and assessment (p. 26).

The implication of this statement is that the overall key competency model and the assessment model need to be grounded within a wider framework to ensure the alignment of pedagogy and curriculum. Oates (2001) similarly comments on the importance of aligning pedagogy with key competency assessment practices.

The developers of the Learning Stories, New Basics, and VET assessment systems, as well as other commentators (Kearns, 2001), note that key competencies are better acquired through the use of student-centred, project-based pedagogies that are meaningful to learners. This has implications for teaching practice and teaching resources, as well as for any professional development that surrounds the introduction and assessment of key competencies.

Although all the developers of the case study assessment systems agree on the pedagogical approaches which best support the development of key competencies, they adopt varied positions in regard to whether key competencies are explicitly "taught" or "caught". As already noted, the Learning Stories and NIFL/EFF case studies present models for key competencies as the core curriculum, with the main teaching emphasis on the development of these competencies. In New Basics the competencies are explicitly taught alongside other academic goals. (One aspect of the research accompanying this initiative inquired into the students' progress in the "old basic" of literacy and found it unaffected by the change in focus (Queensland Government, 2004). By contrast the VET case study offers a model where vocational or "technical" skills are the main emphasis of the curriculum and key competencies are "caught" alongside these skills. However, there is debate about this approach, with some commentators querying whether incidental acquisition is in the learners' best interests (Janinski, 1996 cited in Kearns, 2001).

What is the relationship of key competencies to existing initiatives that assess the national curriculum?

In New Zealand the development of national assessment models has traditionally been attempted within individual learning areas of the overall curriculum. AsTTle, NEMP, ARBs, and Exemplars are all structured this way. Even so, each initiative appears to have taken a somewhat different approach to the establishment of a stated relationship to the existing curriculum.

Exemplars and ARBs were developed discretely within curriculum areas, and appeared to lack an overarching theoretical framework or attention to theoretical progressions, at least in the early stages of their respective developments. Rather, the collective professional experience of the researchers and teachers who developed each ARB item or exemplar provided the link between curriculum and assessment.

The initial NCEA achievement standards were similarly developed from existing examination prescriptions, which were in most cases based on existing curriculum document, or subject syllabi. Again teachers' professional experience, rather than a clear theoretical framework,

appeared to guide the development. More recently some cross-curricular achievement standards for Environmental Education have been drafted. However we have no knowledge of the theoretical basis for their development beyond an alignment with the Environmental Education Guidelines (Ministry of Education, 1999a).

AsTTle is different. That project began by building a clear theoretical framework to link aspects of students' intellectual and knowledge development to the existing curriculum (*see* Section Seven).

We note that any traditional curriculum-specific model of assessment development will lack congruence with the integrated view of key competencies proposed by DeSeCo. Furthermore, the potential place of key competencies within the NCEA is unclear. It is our understanding that recent investigations by NZQA have shown that current NCEA assessments are weighted towards knowledge rather than wider competencies. We have already noted above the danger that that assessment of students' understanding of knowledge will be interpreted as evidence of acquisition of the "making meaning" competency. Yet Section Two also notes that it will be very important to reassure teachers that knowledge does matter and indeed is integral to each competency. For these reasons we think the alignment between the assessment of key competencies and the NCEA will require further exploration, and in all likelihood the development of some new standards, that may be quite different to the present standards (see Section Six). This is an important consideration that will influence the uptake of the key competencies model in secondary schools.

Should assessment of key competencies incorporate acknowledgement of learning in other settings?

Finally in this section we consider the wider question of learning that takes place outside of school settings. Both tertiary case studies (VET and NIFL/EFF) make explicit acknowledgement of the fact that learners develop key competencies in many settings. These case studies provide models of how this learning can be incorporated into the assessment process. The NIFL maps literacy use within three settings: the home and family, in the workplace, and as citizens and community members. Learners are able to collect evidence from these settings to show how they have developed key forms of literacy. A similar model is used in the VET case study. Learners are able to use their course workplace experiences as evidence for their key competency assessment, but learning acquired in other settings such as at home, is not incorporated into the process.

In New Zealand secondary schools the assessment processes for vocational courses also incorporate the learning that occurs in other settings. Students are able to gain credit for workplace learning through the Gateway (an evaluation of this programme is documented in Tertiary Education Commission, 2003) and STAR (an evaluation of this funding is documented in Vaughan & Kenneally, 2003) programmes or from work experience placements which are part of their National Certificates, such as the NCES (New Zealand Qualifications Authority, 1998).

Kelly (2001) comments that current assessment models in the compulsory school sector do not report on the learning which occurs outside school. The lack of recognition of students' out of school learning is also noted in the literature concerning students' transition from secondary school to adulthood. This literature discusses how young people develop their identity, or learn to be somebody who is not just a "student" or a "worker", through engaging in adult activities, such as paid work or community work, whilst still in education (Bye, 2001; Higgins, 2002). Congruent with this, Brewerton (2004b) advocates the use of such work-related experiences as rich contexts in which to both develop and assess key competencies, noting that their salience will be motivating to students.

Models that show how home and community learning might be incorporated into the assessment and reporting of key competencies are needed for early childhood, primary, and secondary sectors. The recent case studies that modeled sharing of information about early childhood learning with new entrant school teachers (Carr & Peters, 2005) could be one good place to begin to build such a model. Another example might involve the incorporation of discussion about students' development of key competencies into three-way student-parent-teacher conferences at any level of schooling. This would have the added benefit of supporting teachers to create continuity across home and school settings in the teaching and learning of key competencies. The importance of creating continuity is discussed in relation to other curriculum areas and in particular in regard to literacy learning by students of diverse cultural backgrounds (McNaughton, 2002).

5. Selecting assessment types and managing the assessment

Suggested types of key competency assessments

A number of different authors have commented on the nature of assessment suited to measuring complex performances. In general their reflections are premised on an assumption that the purpose of the assessment has two functions: to support the development of attributes or competencies, and also to provide summative evidence. They also assume that the assessment of a complex performance requires different forms of assessment than the standardised pencil and paper tests of the past. A summary of this thinking is presented below.

Assessment of complex performances

Delandshere and Petrosky (1998) suggest that if the overall purpose for setting up the assessment is to improve teaching and learning, then only a considered discussion of the evidence can achieve this purpose. Like Rychen and Salganik, they note that attributions of competence are "fundamentally inferences" (p. 49) and say that in order to assess a complex performance it is first necessary to determine the nature of the evidence that can be used to infer the competence exists. As we saw in Section Four, exactly this type of exercise was carried out as part of the research process that accompanied the development of assessment of dispositions for the early childhood learning stories.

Rychen and Salganik (2003) say that inference is strengthened when the relevant behaviours are observed multiple times, in a range of settings, and the performance takes account of cognitive, motivational, ethical and emotional aspects. Accordingly, it becomes necessary to integrate evidence from a range of sources. But there is a likelihood that the judgements made and integrated will remain "implicit and invisible" (Delandshere & Petrosky, 1998, p.15) if they are subsumed into one overall rating. Accordingly Delandshere and Petrosky suggest using "interpretive summaries". They envisage these to be case studies where assessment is a "process of inquiry into one's knowing" (p. 21).

In summary, assessment of a complex performance needs to integrate evidence from a range of sources and requires the performance of a number of complex tasks set in contexts authentic to the person being assessed.

Assessment of dispositions

Carr and Claxton (2002) use a four-point set of criteria to critique the possible methods of assessing dispositions. Their review of these methods is pertinent to the assessment of key competencies given that the assessment of a disposition is an evaluation of a complex performance. The criteria they used are that the assessments:

i.   Are manageable for busy teachers to implement;
ii. Attend to construct validity;
iii. Are locally flexible but longitudinally reliable; and
iv. Support a curriculum in which learning dispositions are integral (Carr & Claxton, 2002, p. 28).

Carr and Claxton (2002) distinguish between three groups of assessment methods. The points they make about each are summarised next, with examples of the methods.

- **Dispositions are inferred through observations:** Observational tasks vary as to their authenticity, reliability, and suitability for formative purposes. Tasks low on authenticity and suitability for formative purposes, but high on reliability include standard experimental tasks or "dynamic" tasks that require learners to attempt things that are too hard for them. Tasks high on authenticity and suitability for formative purposes, but low on reliability include "Learning stories" that are a portfolio of structured observations of "critical incidents" in everyday settings (*see* Section Four).

- **Dispositions are judged by outsider questions and interviews:** These methods are reliable at the expense of validity. Students' performance is judged by others, for example, and could be in the form of Likert-type rating scales as used in the Competent Children/Learners project (Wylie et al., 2004). These scales can track development over time as they have the potential to be produced from pre-school to adults, are quick for teachers to complete, but are not necessarily detailed enough to provide a rich picture of learning which can be used for formative purposes or to support changes in practice.

- **Dispositions are self-reported or self-assessed:** The advantages of self-report methods are that the learner is interpreting their own actions and may be able to specify their development of dispositions in ways that may not be obvious to the teacher. These methods are more suitable for formative purposes. Carr and Claxton (2002) describe five types of self-report or assessment:

    1) questionnaires containing Likert scales;
    2) interviews connected with observations;
    3) self-created learning stories/journals;
    4) in-depth interviews; and
    5) portfolios (which can contain information from other assessment methods).

Carr and Caxton conclude that no one method met all their criteria. Less standardised forms of assessment are appropriate to assess complex aptitudes and attitudes, but are less reliable. They therefore suggest that a combination of approaches is the most logical way forward. The solution

they propose is a combination of a learning disposition grid which is assessed by collaborative teams of teachers (to enhance reliability) and a Learning Disposition Portfolio (LDP). The LDP includes narratives about activities that are completed by learners as they work on valid local tasks. These activities are observed and written up by teacher or learner or both depending on the age of learner. The LDP can also include samples of work. This LDP can be kept from a young age right through school (giving a trail of evidence that enhances reliability). Further explanation of a similar approach is provided in the Learning Stories early childhood case study in Section Four.

Assessment of multi-literacies

Like Carr and Claxton (2002), Kalantzis, Cope, and Harvey (2003) note that traditional assessment measures such as standardised tests are not appropriate for assessing multi-literacies, which they define as the skills required for the new economy, such as, problem solving and working cooperatively. As alternatives they suggest:

* project assessment (of in-depth tasks that require planning and problem solving);
* performance assessment (based on planning and completing a task);
* group assessment (of whole group or members of the team); and
* portfolio assessment (to document the body of work undertaken).

The "rich tasks" of the New Basics case study (*see* Section Four) exemplify the manner in which one or more of these alternatives could be incorporated into assessment tasks suitable for the school sector.

Assessing key competencies in the VET sector

In Australia the Technical and Further Education (TAFE) sector has been trialling and developing new forms of key competency assessment for a number of years. In their review of the Australia Vocational Education and Training (VET) literature Curtis and Denton (2003) identify four approaches used to assess key competencies, and discuss the merits and pitfalls of each. Their analysis is summarised below, with other related observations added.

1. **Holistic judgements by teachers:** This approach is more suited to the school sector as teachers have more opportunity to observe students and know their attributes, and less suited to the VET sector where tutors are less likely to have opportunities to observe students. McCurry and Bryce (1997) report on the successful trial of this approach in the school during which groups of at least three teachers globally assessed students' performance against a three level assessment framework based on the Mayer key competencies.

2. **Portfolios created by students:** This form of assessment can be effective in supporting students to develop an awareness of their skills but often is not presented in a form that is readily "digestible or comparable". Curtis and Denton (2003) cite other research which has

queried the reliability and validity of portfolios. Orland-Barak (2005) found that teachers as learners were more likely to produce *descriptive* reflections to accompany their portfolios. She questions whether learners will demonstrate critical thinking that questions either their own progress, or the nature of the learning tasks set, unless ways to do so are clearly modelled and seen to be safe.

3. **Assessment based on work-experience:** Workplace assessment has similar reliability and validity concerns to those mentioned for portfolio assessment, but enables learners to demonstrate key competencies in authentic situations. A drawback of this approach is that the ability of a learner to display key competencies is dependent on the context provided by the workplace.

4. **Assessment using standardised instruments:** This form of assessment is efficient, provides reports that are easily interpreted, has high reliability and low validity, and has been criticised for its lack of authenticity.

Reaching the same overall conclusion as Carr and Claxton (2002), Curtis and Denton (2003) consider that the relative strengths of these four approaches to assessment, as outlined above, lend themselves to being viewed as complementing rather than competing with each other.

Clayton, Blom, Meyers, and Bateman (2003) also analyse the common forms of key competency assessment in the Australian secondary school and VET sectors. These are: portfolio assessment; assessment by several teachers; self-assessment; and collection of examples of work. They conclude that their analysis suggests an unexamined premise: that it is possible for students to produce a portfolio (hard copy) of their key competency attainment. This approach is currently being used in some New Zealand schools for learning area and essential skill assessment, for example, Year 9 and 10 learners in the inquiry-based integrated programme at Kuranui College were required to collect and keep evidence for a portfolio which was intended to be stored online (Bartlett, 2005).

Issues of reliability and validity

In their papers which describe various forms of assessment of complex performances Carr and Claxton (2002), Curtis and Denton (2003), Clayton, Blom, Meyers, and Bateman (2003), and Denton (2003) all discuss issues of reliability, validity, and manageability. Some of the key areas of agreement are presented below.

Having **more than one person assessing** the key competencies is one commonly suggested way of supporting reliability. In an early childhood setting Carr and Claxton (2002) suggest that this could be achieved by a team of teachers agreeing on the level of a disposition. This approach has also been used in the school sector in Australia (Curtis & Denton, 2003; McCurry & Bryce, 1997). For older students an approach which incorporates the input of the learner is discussed by Denton (2003). This involves validation of students' self-selected evidence by discussion with

their tutors. Clayton, Blom, Meyers, and Bateman (2003) also note that it is important that assessment evidence is validated.

Rychen and Salganik (2003), those in the VET sector (Curtis & Denton, 2003; Kearns, 2001), and those in the early childhood sector (Carr & Claxton, 2002) all agree that key competencies are **grounded in real-life situations**. To a certain extent this requirement could be seen to be driving the type of assessment and the tasks used for assessment — they too need to be located in real-life contexts. Indeed Rychen and Salganik (2003) note that assessment of complex performance requires complex tasks and most authors viewed real-life authentic tasks as a more valid form of assessment than generic tasks (Carr & Claxton, 2002; Curtis & Denton, 2003; Delandshere & Petrosky, 1998; Denton, 2003). However designers of the PISA assessments have attempted to design tasks grounded in real-life contexts that are sufficiently generic for international use (*see* the PISA problem-solving case study in Section Four) and the New Basics tasks are somewhat generic so that teachers can fit them into their intended curriculum.

Most authors agree that learners need to be given a **range of opportunities** to demonstrate competence. They say that more than one task needs to be used to increase reliability and to ensure that the competency being assessed is not situation specific (Carr & Claxton, 2002; Clayton et al., 2003; Curtis & Denton, 2003; Delandshere & Petrosky, 1998; Denton, 2003). This is of particular importance if it is accepted that the attributes (including attitudes and motivations) underpinning the manifestation of a competency can only be inferred (Delandshere & Petrosky, 1998; Oates, 2001; Rychen & Salganik, 2003). For example Rychen and Salganik (2003) suggest that inference is strengthened when:

- the relevant behaviours are observed multiple times; and
- the relevant behaviours are observed in a range of settings.

Curtis and Denton (2003) also suggest that students need to have multiple opportunities to practise and demonstrate skills. This is because the application of key competencies is context dependent. Denton (2003) notes that for students to gain recognition for a key competency performance level they must have demonstrated and been assessed for the key competency twice in different contexts. This is common practice for some national assessments in New Zealand. For many of the unit standards which contribute to the *National Certificate in Employment Skills* (NCES) students must demonstrate competence in three different situations (New Zealand Qualifications Authority, 1998).

To enhance reliability, more than **one method of collecting information** should be used. For example, observation could be combined with self-report or other-report (Carr & Claxton, 2002). Portfolio approaches enable more than one form of evidence to be collected, and also can provide evidence over time and contexts. The VET case study as described by Denton (2003) shows how different forms of evidence are able to be utilised (for example, written or oral evidence, or observations). Looking at a New Zealand example—assessment of unit standards in the NCES— we see that assessments tend to rely on two assessment methods: student completion of worksheets/books and observations (New Zealand Qualifications Authority, 1998).

Clayton, Blom, Meyers, and Bateman (2003) note that assessment tools need to be trialed and validated. Carr and Claxton (2002) suggest that to track across contexts, assessment tools need to be reliable.

Writing about high stakes assessment Keightley and Keighley-James (2001) note that validity is more difficult to deal with than reliability because:

> ..validity studies rely heavily on judgements made about human behaviour, human abilities and human decision-making processes which makes it a difficult concept to report on and defend. There also seems to be an inherent mistrust of the role of human judgement in assessment mainly as it involves acts of interpretation that are subject to disagreement which can seem to undermine the impartiality of assessment (Keightley & Keighley-James, 2001, p.4).

If these are the challenges seen for public examinations, the validity challenges outlined are likely to be magnified in situations where judgements are necessarily made at the local/school level in order to satisfy the requirement to use authentic tasks. In view of this, the process of "secondary inquiry" being developed by the Senior Secondary Assessment Board of South Australia (SSABSA) may be of particular interest to policy-makers. The inquiry process is retrospective and involves "validation through challenge" (Keightley & Keighley-James, 2001, p.6). Educators who are skilled in critical analysis and argument, but who bring an outsider's perspective to the curriculum area in question, work through an analysis process with the assessors. Together they identify areas that might be "vulnerable to alternative explanations" (p. 6). The intention is to strengthen assessment expertise in the teaching community through an open process of dialogue and professional growth. This model appears to be an interesting example of what Delandshere (2002) would call "assessment as inquiry".

Issues of manageability

The manageability of assessments influences their uptake and therefore their reliability. Carr and Claxton (2002) and Denton (2003) comment on the need for assessments to be manageable for the teacher or tutor, and Kelly (2001) notes that assessments need to be cost effective. Oates (2001) also discusses the tensions between validity, reliability, and utility. One way of ensuring manageability is through the use of tasks that are being carried out for a learning or work-related purpose. For example, the VET case study shows how tasks that assess vocational skills are also used to provide evidence for key competency assessment.

Clayton, Blom, Meyers, and Bateman (2003) observe that in the UK the requirement to separately assess generic skills contributed to them being taught individually (rather than in an integrated manner) and created an unworkable assessment burden for teachers. As a consequence this system was phased out.

Some teachers expressed concerns about the workload associated with the implementation of rich tasks assessment in the New Basics case study (Queensland Government, 2004). However research of the NCEA implementation suggests this could be as much about the associated move to standards-based assessment (Alison, 2005; Hipkins, Vaughan, Beals, & Ferral, 2004), or the need for considerable curriculum revision and pedagogical change, as about the assessment process *per se*. Now that familiarisation with standards-based assessment has occurred in New Zealand, at least for secondary teachers, the instigation of a similar system for key competencies is likely to be more easily interpreted and acted on. Nevertheless there is a need for close attention to be paid to the resourcing and professional development that will need to accompany any new assessment initiative.

In this context, it is interesting to note that most of the case studies describe a slow and thorough development process. For example, the New Basics trial took four years, and involved among other support mechanisms, a range of teacher development sessions, resource materials, critical friends assigned to support schools, and chatrooms and forums for students, staff and community members. Selected cohorts of students were not involved until the second year (Queensland Government, 2004). The SSABSA "secondary inquiry" process briefly sketched above could provide an interesting model for strengthening assessment expertise at the same time as proactively addressing the validity challenges that the NCEA implementation experience suggests can be expected.

Summary

There is considerable consensus in the literature concerning the characteristics of assessment that will be suitable for the assessment of key competencies.

In Section Three we noted that assessment for lifelong learning is likely to provide a richness of information that can also be used for teaching and learning purposes, and can be reduced for accountability purposes, provided validity and reliability requirements are satisfied. A synthesis of the Section Three analysis with the assessment requirements outlined in this section suggests that:

- Assessment is standards-based so that assessment involves making a judgement about a performance;
- This performance should be as authentic as possible and located in real-life contexts;
- Several assessment events and/or contexts should be used and different types of evidence should be collected when assessing overall competency in any one aspect;
- More than one person should be involved in making the judgement, and preferably the learner should be included in this process when they are old enough to do so meaningfully;
- The learner and the assessor should both be clear about the types of evidence that will be used to infer the successful performance; and
- Reporting on the assessment should provide clear feedback based on the evidence collected, so that the next learning steps are clear.

This section shows that there is a range of ways in which these requirements could potentially be satisfied, and indeed suggests that it will be preferable to use a variety of types of assessment tasks, as appropriate to each new context.

We also conclude that support and professional development will be very important aspects of implementation of any assessment change. Carefully designed teacher support could address issues of manageability whilst also proactively establishing confidence in the validity of a system that will need to place a heavy emphasis on teacher professional judgement.

6. Describing achievement at a specified curriculum level

This section of the report considers issues related to the specification of what learners will have achieved — the benchmarks against which they could be assessed at any one curriculum level. There are two parts to this process. The first part of this section reports on processes for specifying standards to be assessed. The second part looks at how others have determined levels of achievement within these types of standards (that is—horizontal progression). Following that, Section Seven of the report discusses vertical progression across age/curriculum levels.

This section assumes that any proposed assessment system will be standards-based. The analysis in Section Three shows that a standards-based system is needed for assessment purposes and paradigms that are most compatible with the assessment of key competencies as complex performances. The case studies outlined in Section Four all use some form of standards-based assessment. More pragmatically, the recent NQF and NCEA assessment reforms are standards based, and harmonious integration of any new assessment requirements would be most easily achieved if the new assessment instruments were compatible with those already in use. It would be unwise to squander the gains made via the huge effort that has already been put into educating New Zealand secondary school teachers in standards-based assessment, especially given the closer alignment of standards-based assessment paradigms with future-focused purposes for assessment.

Designing standards for key competencies

Historically, assessment for accountability purposes has focused on students' progress in gaining new knowledge and, sometimes, skills. The nature of key competencies demands evidence of new types of learning outcomes that we have no history of assessing in this way. Thus an important consideration for policymakers is how we might approach the development of the necessary standards, given the relatively "blank slate" with which we must start. The analysis that follows draws on experiences of those who have faced this same challenge, with varying degrees of success. In particular, it draws on the case study of the National Institute for Literacy (NIFL) process for designing Equipped for the Future (EFF) standards to assess adults' knowledge and skills for daily life in the 21st century (*see* Section Four). With its intention to assess adult learning in the actual contexts of family, community life and work, this initiative has close parallels with the DeSeCo competencies that underpin the New Zealand curriculum initiative.

The NIFL initiative has designed and continues to refine 16 "Equipped for the Future" (EFF) standards. The design process drew widely on both tertiary sector expertise and on adult learners' own views of the knowledge and skills that were seen as important enablers of purposeful actions in their daily lives — at home, at work, and in the community. Obviously it would be impossible to focus on everything that adults are required to do. From an analysis of the future-focused literature the developers chose to work with four key purposes for adult learning:

- gain access to information and ideas;
- communicate with the confidence that their message makes sense and can be understood by others;
- make decisions that are based on solid information and reached through thoughtful analysis, consideration of options, and careful judgement; and
- keep on learning so they won't be left behind (Stein p. 17).

The subsequent steps taken for the design of the standards were:

1. Determining the parameters of a typical "performance"

Beginning in 1994 the designers gathered "snapshots" of adult performances in many different settings.

2. Creating validated "role maps"

In 1996–97 the developers created "role maps" that defined broad areas of responsibility, in relation to the four identified purposes, in each of the three contexts. This analysis then led to the identification of "key activities" in each setting and these were validated in an extensive consultation process. While the developers found obvious overlaps (for example it is necessary to use communication skills in all three types of settings) the actual skill use differed with the setting. This level of detail was captured in "role indicators" that specify aspects of the performance such as "How? How well? With what?"

3. Turning role maps into standards

During 1998–99, through an "iterative field and expert review process" (p. 15), the material was refined into 16 standards in four clusters that match the four purposes listed above. These were published in 2000 and began to be used by adult educators across many states of the USA.

4. Creating levels of performance within each standard

Meanwhile the project continued to the next stage — the ongoing design of a "performance continuum" for each standard, specifying at least three levels at which the standard may be met. Again the design process has been iterative, drawing on teacher and learner expertise about how adults actually perform in each type of task, and relating this type of evidence to the theoretical perspectives provided by research. Draft versions of the standards as at April 04 are available at http://eff.cls.utk.edu/assessment/standards.htm.

Throughout the various stages of the NIFL adult literacy research, the design of the 16 EFF standards was guided by six specific quality standards:

- EFF standards must accurately reflect what adults need to know and be able to do;
- EFF standards must be reflective of broad consensus;
- EFF standards must be specific enough to guide instruction and assessment;
- EFF standards must be able to be measured;
- EFF standards must define multiple levels of performance for students to strive for; and
- EFF standards must be written clearly enough for all stakeholders to understand (Stein, 2000, pp.18-19).

Commentary

The design process followed by the NIFL could be of interest to New Zealand policymakers. It has obviously been resource intensive — rough dates have been provided in the description above to show the time scale, for instance. However the involvement of many learners in role mapping, and of both educators and research experts in validating and trialing the standards as they evolved was seen as necessary to ensure public confidence in, and understanding of, the standards. The numerous state-specific Internet references to these standards support the claim that this initiative has been widely adopted (Texas Centre for the Advancement of Literacy Learning, 2003).

Similarly intensive consultation processes were followed by staff at Torrens Valley TAFE as they developed key competency assessments for VET students and as the New Basics rich tasks were devised (see Section Four case studies). Even the PISA assessment case study, which falls more into the empirical-analytic paradigm, drew on a wide range of practical expertise at the stage of designing the parameters for and details of tasks that were in turn analysed psychometrically to create the standards that were ultimately used.

It is worth noting here the recent controversy surrounding the NCEA achievement standards. These were developed more or less exclusively by those with secondary school curriculum expertise in the relevant area, and levels of progression were determined at the same time as each standard was described. In comparison to the EFF standards, NCEA standards lack detail in their description and in their progressions. They tend to be generic in their specification of increasing complexity, rather than task and context specific. Because of this, the actual standard is defined in practice by the tasks used for assessment and this has led to controversy about a lack of consistency (Alison, 2005). Wider involvement at the design level might be a prudent way of avoiding similar controversy about the actual content of any standards devised to assess key competencies, whilst simultaneously supporting the development of richer indicators of learning progress.

Describing achievement levels (horizontal progression)

As outlined above, there is a difference between specification of a standard and specification of performance levels within a standard. A further distinction needs to be made between *horizontal* indicators — differing levels of performance amongst a similar group of learners — and *vertical* indicators of progression over time (these are discussed in Section Seven). The analysis of other projects suggests that richly described indicators for both types of progression are needed if assessment is to effectively inform teaching and learning, and to promote the development of lifelong learning skills.

Describing horizontal levels of performance

The table on the next page is a greatly abbreviated summary of the information that describes a performance continuum for one of the 16 NIFL EFF stands. Other detail not included here gives information similar to that in NCEA range statements, as well as indicators of how easy or difficult learners could find the task and the consequent level of support needed. Examples of contexts in which learners might carry out the task, and a brief summary of supporting background research, with a bibliography, are also included. While this is draft material that has doubtless yet to be professionally published, the richly descriptive material extends to 13 pages of text. This stands in sharp contrast to NCEA achievement standards. The indicators of progression for a typical NCEA research standard are included below to allow a comparison to be made.

Table 4 **AS Geography 1.5: Carry out and present directed geographic research**

This achievement standard involves carrying out and presenting directed geographic research, describing the relevance of geographic ideas to the research, and evaluating the research process.

Achievement	Achievement with Merit	Achievement with Excellence
Carry out and present directed geographic research.	Accurately carry out and present directed geographic research.	Accurately carry out and present directed geographic research.
Describe the relevance of a geographic idea to the research.	Describe, in detail, the relevance of a geographic idea to the research.	Describe, in detail, the relevance of geographic ideas to the research.
	Evaluate the research process.	Evaluate, in depth, the research process.

Table 5 **Paraphrased summary of descriptions for the performance continuum of the draft EFF standard "Learn Through Research"**
http://eff.cls.utk.edu/assessment/standards.htm.

	Paraphrased summary of descriptive indicators of performance		
Aspect of performance addressed	**Level One**	**Level Two**	**Level Three**
Ability to pose question or problem to guide research	Pose simple questions or predictions that can be adequately researched using a few familiar resources	Pose a question or prediction that can be adequately researched using a range of resources	Pose complex, sometimes novel, precise questions for purposeful development of a coherent inquiry
Sophistication of information gathering strategies	Use multiple simple strategies to draw information from a range of sources	Use multiple strategies to gather a store of information from a range of familiar and less familiar sources	Use a range of sophisticated strategies to gather information and adjust as necessary to enhance comprehension
Cognitive processes involved in organising, analysing and evaluating	Evaluate usefulness of the information gathered and integrate it with prior knowledge	Restate, summarise, compare and contrast, and evaluate information, integrating as relevant with prior knowledge	Analyse and synthesise information from multiple sources, and use a range of evaluation strategies such as cause/effect analysis
Meta-cognitive aspects of research process	Use simple strategies to monitor effectiveness of inquiry	Use a range of strategies, including expert feedback and prediction based on pattern recognition to monitor and adjust progress	Use strategies such as interim summary, verification via multiple lines of inquiry, and adjust approach based on feedback
Interpretation and reporting of findings	Make brief but accurate oral report of findings	Communicate through accurate oral or written presentations	Communicate through extensive oral or written reports, complex graphics

Rich descriptors are also a feature of the PISA Problem Solving case study. What follows is a paraphrased version of the many aspects that go to make up each of the four horizontal levels described (for a full version see OECD, 2004).

Level 3: Reflective communicative problem solvers

- Think about underlying relationships in a problem and relate these to solution;
- Are systematic and can construct own representations of the problem;
- Communicate solutions using accurate written statements and other representations;
- Monitor multiple variables, account for temporal and other restraints;
- Cope with multiple inter-related conditions; and
- Organise and monitor own thinking.

Level 2: Reasoning, decision-making problem solvers

- Use reasoning and analytic processes to solve problems;
- Can use inductive and deductive reasoning, and can link cause and effect;
- Systematically compare all possible variations in well described situations;
- Combine and synthesise and represent information from a variety of sources; and
- Draw inferences based on two or more sources of evidence.

Level 1: Basic problem solvers
- Solve problems with a single data source and discrete well defined information;
- Locate and retrieve information related to the major features of a problem;
- Can carry out basic transformations of information (e.g. from table to graph); and
- Check some well-defined conditions for a problem.

Below level 1: Weak or emergent problem solvers
- Deal with straightforward problems and carefully structured tasks;
- Responses require few or no inferences; and
- Show difficulties in making decisions and analysing or evaluating systems.

While the processes followed were very different, the PISA standards, like the EFF standards, and the VET case study standards, were defined by the combination of deep theoretical expertise in the relevant area with empirical data that demonstrated the range of learners' actual abilities. Obviously, such processes require a range of people with different types of expertise to work together.

Theoretical inputs to horizontal progression: The SOLO taxonomy

The assessment systems described above drew on theoretical ideas of relevance to the area of learning being demonstrated. Progressions may draw on *developmental* theories as well. A search of the literature quickly reveals that the SOLO taxonomy has been widely used for both horizontal

and vertical progressions in a wide range of contexts and discipline settings. Horizontal progressions are discussed here and vertical progressions in the following section.

The SOLO taxonomy (Structure of the Observed Learning Outcome) is a widely used empirically derived general taxonomy of levels of learning outcome (Biggs & Collis, 1982). The taxonomy describes five levels[6] to which a learning outcome may be demonstrated. These levels are "based on a person's ability to handle, with increasing sophistication, relevant clues" (Pegg & Inglis, 2005, p.1) and so can be distinguished by their increasing complexity:

Prestructural: no evidence of anything learned, student may repeat the question or convey some irrelevancy.

Unistructural: the student shows at least one element of knowledge but does not apply or transfer this easily.

Multi-structural: several relevant elements are present but these aspects are not interrelated and inconsistency results.

Relational: the relevant elements are integrated into a generalised structure; there is evidence of induction.

Extended-abstract: the structure of elements is related to other relevant domains of knowledge; answers are not bounded by the question allowing the student to suggest alternative outcomes.

Use of SOLO in asTTle

Here in New Zealand the SOLO taxonomy has been integrated with an analysis of the "big ideas of mathematics" and other relevant features such as considerations of text and diagram demands to construct "item signatures" for individual questions in the asTTle item banks (Thomas, Tagg, Holton, & Brown, 2002). [7] Responses at the prestructural level require students to have actually attempted an answer but the taxonomy was used to classify item potential at the *development* stage and so a four-level progression beginning from the unistructural was devised (Thomas et al., 2002). These item classifications determine the four levels of horizontal progression from "surface" to "deep" learning that underpin all the asTTle assessment tools (Hattie & Brown, 2004).

Hattie and Brown (2004) describe a range of advantages for using the SOLO taxonomy to underpin horizontal progressions. Some of these can be clearly related to considerations of key competencies as complex holistic performances. For example they assert that the use of SOLO "separates the concept of 'difficulty' from 'complexity'"(p. 17), and they point out that questions

---

6   The taxonomy also describes modes of thinking but as these are primarily used for determining vertical progressions they are discussed in Section Seven.

7   All asTTle technical reports can be retrieved from www.tki.org.nz

can become more difficult without necessarily becoming more complex. SOLO does not separate context from concepts, or student performance from teacher performance (p. 28). Describing an earlier meta-analysis of study skills (pp. 21–22) Hattie and Brown link the two "deep" levels of SOLO (relational and extended abstract) to features such as self-regulation and transfer to new learning situations. Such features are linked to the development of skills and dispositions for meta-cognition and for lifelong learning, and in turn, the support of lifelong learning as an educational goal is seen as an important purpose for curriculum, teaching, and assessment reform (see Section Three). On a more practical note, they say that the use of SOLO provides an "item writing methodology" and "can be used for meaningful reporting to teachers and students" (p. 27).

Use of SOLO in the VET sector in Australia

Curtis and Denton (2003) note that staff at Torrens Valley TAFE used the SOLO taxonomy to describe performance levels for key competencies, using the taxonomy to reflect an increasing degree of meta-cognitive control. Again, not all five levels have been used for each indicator, with only those that were deemed to be amenable to reliable judgement being selected. In the following example of a problem-solving competency, this plays out as increasing autonomy and flexibility in addressing a work-related problem.

Level 1: Solves the problem by following an established procedure.

Level 2: Solves the problem by selecting from several alternative established procedures.

Level 3: Creates a new procedure or adapts an existing procedure to meet the demands of a task.

Curtis and Denton (2003) note that the three levels of the seminal Mayer key competencies reflect an increasing degree of **meta-cognitive** complexity.

Practical issues when using SOLO for determining assessment progressions

While it is possible to assess the unistructural level of SOLO using quantitative questions, the requirement to relate elements of an answer into a coherent whole suggests that questions that require something more than a brief response are necessary for determining the three higher levels (Teaching and Educational Development Institute). Hattie and Brown (2004) recommend writing individual questions that invite responses at one of the four levels, and then ensuring that the questions collectively cover the whole range. While useful to bear in mind, these types of recommendations presuppose a formal written assessment in the empirical-analytic tradition, which Section Three has posed as problematic for the assessment of complex performances.

In fact, one critique of SOLO is that the taxonomy was constructed using written or verbal evidence for analysis, and may need modification before it can be used in situations where other types of outcomes are sought. (Although we note that the VET sector in Australia appear to have

successfully used the taxonomy to judge other types of evidence.) Jackson (1995) says that the use of the taxonomy in a design learning context requires that interviews, reports or other written evidence be produced specifically for analysis. This critique suggests that, whatever its attractions, SOLO would need to used with care when determining progressions for key competencies.

Standards-based assessment and the reporting of key competencies

As outlined above, well-developed standards can be richly informative of students' learning. However that potential will only be fully realised if reporting processes allow the richness to be shared in ways that are meaningful to all the assessment stakeholders. This means that the performance of any individuals needs to be accessible to them and to their parents and employers (as relevant), in a form that is easy to read (Kelly, 2001), and provides meaningful information over time (Curtis & Denton, 2003). Kelly (2001) notes that reporting needs to allow for recognition of the competencies and also the context in which they are acquired.

Clayton, Blom, Meyers, and Bateman (2003) provide a useful model for discussing the various ways key competencies could be reported in relation to an existing curriculum. The alternatives they identify include:

- cross-curricula reporting of key competencies as these are developed across a range of subjects;
- reporting development of a key competency within one subject;
- grouping similar subjects such as science and technology when reporting; and
- developing a portfolio of actual tasks for reporting, without necessarily making reference to discrete subject areas.

There is scope for considerable creativity in designing reporting formats. The early childhood case study illustrates the relevance, power and clarity of learning stories, as appropriate to the developmental level of the learners and the needs of a wide range of parents to understand the assessment process (see Section Four). By their very nature, stories include a context, as recommended by Kelly (2001). One institution involved in the adult literacy NIFL case study used "career passports" as a reporting format. These relate the competency to a prospective context — that of work.

7. Describing vertical progression in achievement

New Zealand's school curriculum specifies eight vertical levels of achievement. These assume measurable progression in learning over time, specified within "achievement objectives". There is a well-established curriculum tradition, here and elsewhere, of specifying and assessing progression in knowledge development, although the basis for this is not always clearly explained. If the key competencies are to be assessed, the issue of what we might expect progression to look like across the years of schooling will need to be addressed. But the question is complex. The Competent Children/Learners longitudinal study has shown that children's cognitive development is not linear and orderly. Yet specification of progression (which have tended to assume linear trajectories) has critical consequences for learners:

> As we move towards both a greater use of student assessment to support student learning, and a greater use of assessment data to set individual and school targets as levers towards improving performance levels, the questions of what learning progress we can expect in given lengths of time, and what we are basing this expectation on, are increasingly important. Not prejudging students' likely progress, and approaching their teaching with optimism, is crucial to the chances of students flourishing (Wylie, Hodgen, & Ferral, 2005, p.113-114).

This quote from Wylie et al. is a timely reminder to take care when making assumptions about learners and progression. McGivney (2002) points out that progression is a word with many meanings, although "at its simplest it connotes the idea of improvement in a sense of advance from a less favourable to a more favourable state or situation" (p. 11). She says that learning outcomes, to which ideas of progression are typically applied, may extend well beyond advances in specific knowledge and understandings of formal learning to encompass changes and actions in daily life that occur as a *consequence* of learning. When viewed this way progression for adult learners can be variously seen as:

- personal progression (for example the development of greater confidence, autonomy, self-esteem, tolerance and understanding; wider aspirations; changes of attitude);
- social progression (improved social interaction and higher levels of civic and social participation such as volunteering and membership of local organisations);
- economic progression (acquisition of skills for employment, gaining a job or advancement within a job);
- collective progression (when groups of learners move together to achieve collectively defined goals or to create new local networks (McGivney, 2002, p.17).

McGivney's list is a useful reminder that there exists a range of aspects that can show progression as a consequence of learning. Some of these relate to what she calls "soft" outcomes such as an increase in self-confidence, that are personally and subjectively determined, and are a matter of degree rather than absolute (p. 21). However the distinction is not necessarily straightforward and there is evidence that hard and soft learning outcomes are "symbiotically linked":

> For many learners, the new understanding and acquisition of skills acquired in a supportive learning environment creates a heightened sense of self-worth which in turn imparts the increased confidence they need in order to progress to higher levels of education, to apply for jobs, or to undertake more demanding skills (McGivney, 2002, p.22).

Given the intention to develop competencies holistically, and to look for demonstration of them in personally meaningful contexts, these insights from Wylie et al., and of McGivney, are clearly very relevant to the deliberations of this section of the report. Here we outline merits and pitfalls of various approaches that could be employed when addressing vertical progression.

First, we outline recommendations from British assessment experts who have considered pitfalls to be avoided when specifying assessment benchmarks for progression in a school curriculum. Next, issues for constructing developmentally based sequences for progression of competencies are outlined and the use of the SOLO taxonomy for determining the nature of vertical progressions is discussed. Finally, other theoretical and pragmatic methods that could be used for determining vertical progressions are outlined.

Ways to specify levels of attainment

Writing about recent UK school curriculum decisions, Wiliam (2001) cautions that policy-makers must be very clear about the purposes for which they want levels of attainment specified, or these are likely to be misused. The UK curriculum specifies "attainment targets" that appear to serve a similar function to New Zealand's achievement objectives — especially when these are actually read more as learning *outcomes* than as learning *intentions*. Attainment targets specify what learners should know and be able to do at certain curriculum transition points but Wiliam asserts that:

> ..it has become clear that policy-makers have never really understood the nature of levels of attainment in the national curriculum. The result is that current policy has driven teachers and schools in ways that are antithetical to high-quality teaching and learning (p. 1).

The research programme led by Black and Wiliam (see for example Black, Harrison, Lee, Marshall, & Wiliam, 2002; Black & Wiliam, 1998), and widely recognised around the world, places its emphasis on assessment for learning, and recommends designing assessment systems with this as a priority. In this document and elsewhere they argue that "formative and diagnostic assessments can be aggregated to serve summative and evaluative purposes" (Wiliam, 2001, p.2). This takes very careful design, however, and one important issue to be addressed is how to specify and interpret levels of attainment in the curriculum.

According to Wiliam, a key progression issue to be resolved is how to make learning outcomes/attainment targets that are sufficiently challenging for a wide range of students without either overwhelming and demotivating the least able or "undermotivating" the most able. This is likely to happen when decisions are based on simple definitions of what children should normally be expected to know, understand and do at age-specific points of their learning:

> The difficulty with such simple benchmarks is that if they are sufficiently demanding so they provide real challenges for the most able, then they are so far beyond the reach of most students that the students are likely to give up. Conversely, if they are set so as to be motivating for the lower attainers, then they provide no motivation for the high attainers, who will quickly see they can obtain a 'satisfactory' score with little effort (Wiliam, 2001, p.3).

Wiliam also advances another type of argument against age-specific clustering of attainment targets or standards, especially when these are reported as overall grades (as in the UK National Testing programme and in the New Basics assessments). When assessed in this way students may perceive that they have made no progress if they gain, say, a "B" grade at age 7 and again at age 10, even though they will have made steady progress in real terms. If their grade drops as the work gets harder or more complex, they may perceive that they are now "failing" or "falling behind" even though, again in real terms, they will have actually made some progress. Wiliam cites the research of Carol Dweck (1999) who is also widely cited in discussions of the development of self-regulated learning skills. Dweck found that students who view learning ability as a fixed entity are more likely to be discouraged when they strike challenges in their learning than those who think they can surmount challenges with more effort. Wiliam notes that the second of these dispositions — relating ability to effort — is more likely to be fostered by an assessment system that reports actual achievement, whereas broad age-related grades can act to reinforce the "fixed entity" view.

Wiliam suggests that an element of horizontal progression (relative levels of achievement of the same target or outcome, as in the NCEA) can address some motivation issues for less able students. However this will not address the challenges of motivating the most able if they are disposed to choose the least taxing achievement option. We have recently seen exactly this dilemma play out in some teachers' concerns about the NCEA (Alison, 2005; Hipkins et al., 2004). Therefore, as an overall policy response, Wiliam and his assessment team recommend an age-independent "ladder" of related learning outcomes as the best solution. Students at any one age-level may be at different rungs of the ladder but all will be able to register actual progress on the same broad pathway. The boundaries of New Zealand's eight curriculum levels are deliberately "fuzzy" to discourage age-specific associations and some progressions in the New Zealand curriculum already adopt the pattern Wiliam recommends. But many other suites of achievement aims and their associated achievement objectives show distinct disjunctions at different levels.

While this policy decision is straightforward in principle, Wiliam goes on to note that:

79

> With a progressive model, the essential task is to define the nature of progression — in other words, when someone gets better at a subject, what is it that gets better? Once we know this, we can say what it is that a student should do next, given they have reached a particular point in their learning. The answer to this question may be difficult, but to say there is no answer is to deny the possibility of progress in the subject at all (Wiliam, 2001, p.5-6).

A subject is not the same as a competency, but obviously they are related. The question is valid, if more complex. Do we expect students to get better at demonstrating competencies? If so, what will this progress look like? And how, if at all, will it be related to progress in their learning of various subjects? This question raises the challenge that competencies are demonstrated in multiple contexts, but students may very well make faster learning progress in one subject or competency than in another.

Faced with these complexities, Wiliam's assertion that decisions about progressions are value-based is perhaps encouraging. He says that once we have decided what it is that actually gets better, specification of levels of attainment is arbitrary and we can "site them where we like" (p. 7). However it does seem to us to be important to explore and learn from the various conceptual frameworks, assumptions, priorities, and values on which progressions could be constructed. This is the issue to which we now turn.

Is it possible to construct a developmental framework for progression in the key competencies?

Many researchers comment on the need to locate the key competencies within a clear theoretical (Carr, 2004b) or developmental framework (Kearns, 2001). However, writing in the context of the Australian VET sector, Kearns notes that not enough is known about how competence is acquired and that while some models of progression incorporate theories of human development, others, including the widely used Mayer model do not. For this reason he says that "considerable development work" would need to be undertaken to construct a framework that identifies and defines a developmental progression in the components of the various key competencies and in the sub-skills for each. We think this recommendation begs a question of whether sub-components can be meaningfully identified and assessed without compromising the intention to develop competencies as complex holistic demonstrations of competence.

Weimer (2002) notes that developmental frameworks can trap users into viewing development as linear. She says that a linear model cannot take into account situations where the context gets more complex but the learner's level of competency does not. Wylie et al. also point out a range of potential pitfalls when assuming an unproblematic linear progression:

> We know that students can suddenly 'get' things that surprise their teachers; that gaining of knowledge and understanding can occur at different rates at different times, depending on a range of factors, sometimes leaping ahead rapidly, sometimes showing a flatness because of consolidation, sometimes showing a dip that may simply indicate attention and growth

elsewhere (e.g. an area of new discovery or interest), or a querying of previously held assumptions that make it hard to respond in the former way that allowed them to succeed (Wylie et al., 2005, p.113).

This quote accommodates the multiple complexities that can influence progression in a way that is a timely reminder that the key competencies are envisaged as complex, contextual, and so on. Addressing the same issue, Carr and Claxton (2002) suggest that assessments need to be flexible to take into consideration the different contexts of learners' lives but also need to be longitudinally reliable. While this seems eminently sensible in principle, the recommendation *per se* does not address the underlying design question.

Postmodern research raises a quite different set of philosophical questions. From this perspective, it is important to be clear about what we take as the "givens" of young people's development as learners. Howley, Spatig, and Howley (1999) present an overview of various types of postmodern critiques that have been made about developmentalism. They note that existing ideas of development mainly specify cognitive development, as this has been defined within a positivist science discipline, with all the critiques that inhere to that. Depending on the ideological perspective of the critic, existing taxonomies that specify a developmental progression are variously seen as grounded in white, male, middle class, Western European, libertarian, capitalist cultural perspectives. They are seen to be essentialist and determinist and to almost always focus on a unitary self. The latter does not sit well with the situated DeSeCo model in which individuals act in many social fields and need to demonstrate that they can adapt appropriately to each one, while still acting with autonomy.

While clearly signalling pitfalls, none of this brings us any closer to a basis for actually deciding what will "get better" and what we might want to see progress. Any approach chosen to describing vertical progressions in learning will need to be mindful of the above pitfalls and also take up the challenge of keeping the complexity of each competency in mind. Accordingly, we turn now to several different approaches for addressing this challenge, beginning with the developmentally-based SOLO taxonomy.

The use of the SOLO taxonomy

The five *levels* of SOLO described in Section Six identify differences in ways learners can respond to a task by making connections of differing degrees of complexity. Pegg and Inglis (2005) explain how a second dimension is added to the SOLO taxonomy to specify the nature of vertical progressions. This dimension adds five *modes* to the SOLO taxonomy as shown in the diagram overleaf.
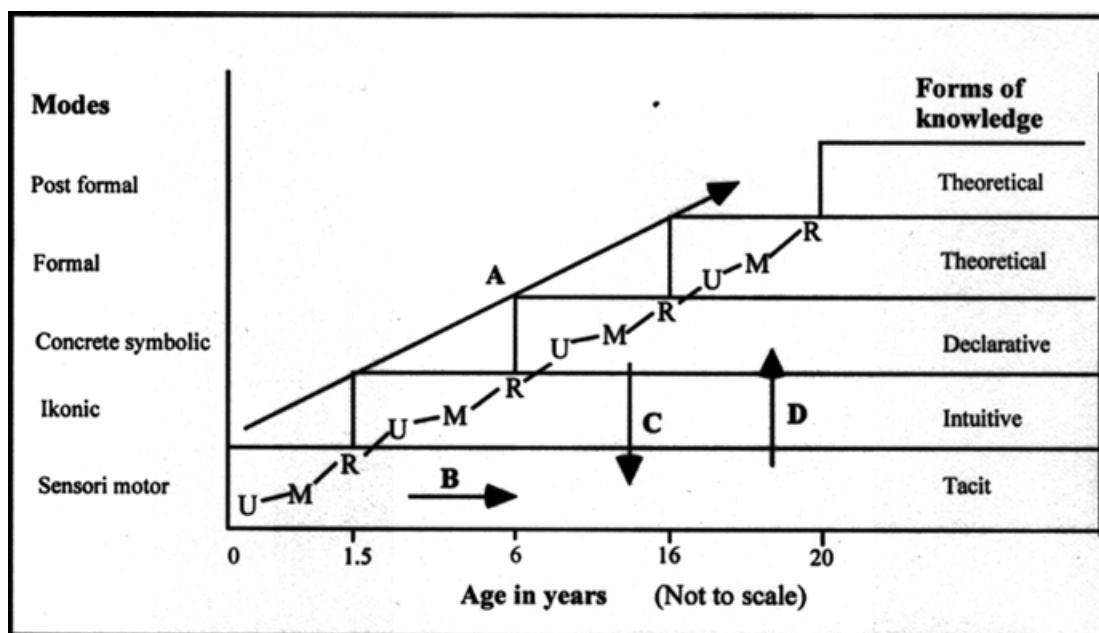
Clearly based on Piagetian principles, five modes of development are identified. These are sensori-motor, ikonic, concrete symbolic, formal, and post-formal. This aspect of the taxonomy differs from Piaget's work in that the descriptors are applied to the students' response to the specific task, and the context in which it is carried out, not to the student as an individual. Thus the taxonomy, as applied, is task specific and of limited use for making wider inferences about achievement (Panizzon, 2003).

In this version of SOLO, students can demonstrate the various levels of achievement within any one of the five modes of progression. Thus a student operating in the concrete symbolic mode may begin by making unistructural responses but after having achieved the "deep" learning levels of multi-structural or relational responses, they may move up to the formal mode, where again they might initially slip back to making unistructural responses, and so on.

Pegg and Inglis assert that "the strength of this model is the linking of the cyclical nature of learning and the hierarchical nature of cognitive development. Each level of functioning within a cycle has its own integrity, its own idiosyncratic selection and use of data" (Pegg & Inglis, 2005, p.5). They say the use of this taxonomy allows for the development of frameworks for assessment of learning that have "criteria specific enough to provide support for teachers in planning their work and in allowing students to interpret their performance" (p. 6). Support for this assertion comes from a recent two-year study of changes secondary science teachers in rural New South Wales made to their teaching practice as a result of ongoing support to use the multi-dimensional SOLO taxonomy for assessment of learning (Panizzon, 2005). These changes included better structuring of questions and other teaching strategies to allow students to display a wider range of levels of learning, more focused lesson planning, and paying more attention to levels of student cognition.

The Berkeley Evaluation and Assessment Research (BEAR) project has also used the multi-dimensional version of SOLO in the construction of assessment tasks that led to the production of computer generated Performance Maps (Kennedy, 2005; Wilson & Draney, 2004). The developers of this project took as a "first principle" that their assessment system would be based on a "developmental perspective of student learning" because

> A developmental perspective helps researchers move away from "one shot" testing situations — and away from cross-sectional approaches to defining student performance — toward an approach that focuses on the process of learning and on an individual's progress through that process (Wilson & Draney, 2004, p.134).

This American team, like the Australian team (Pegg & Inglis, 2005), advocate for using information based on well-researched learning progressions to improve teaching and learning via the assessment process. As in the New Basics case study, they focus on the assessment of outcomes in a way that allows a match to be made between classroom learning and the assessment tasks — a match that they describe as "aligning" the curriculum, teaching and learning, and assessment. For each curriculum area they have worked with, detailed generic scoring rubrics have been produced. These are supported with exemplars to guide teachers' decision-making about the level of development the student has reached and to enhance their professional learning. The students' assessment results are then fed into a computer programme developed through Rasch modeling, to produce both individual and whole-class "performance maps". It has been noted that the BEAR analysis process, while promising, is very time consuming and would require "'many more years of effort to map detailed classroom assessments to national frameworks" (Suter, 2004, p.174).

Other methods for determining progression

If the benefits of developmental perspectives are to be used but the pitfalls avoided, what other methods of determining vertical progressions might be considered? This final part of Section Seven draws together other theoretical perspectives and teacher expertise to suggest a pragmatic co-construction process for moving forward with this complex question.

Grounding progressions in socio-cultural theory

Since the DeSeCo competencies are strongly oriented to socio-cultural perspectives (Section Two) is it important to consider how socio-cultural theory might inform the process of constructing progressions for their assessment.

Margaret's Carr's policy papers for the curriculum project provide strong theoretical links between socio-cultural theories of development and the notion of progression as the "transformation of participation" (Carr, 2004c, p.2). She draws on Broffenbrenner's ideas of power shifts from the teacher to the developing learner (an idea that is compatible with

pedagogical change to a lifelong learning framework, as outlined in Section Three). His ideas of increasing complexity of the context and wider domain of operation within which the student can demonstrate their growing competence provide other possible theoretical aspects to explore when developing vertical progressions.

In another discussion paper Carr introduces the work of Beach, who links the idea of progression with the pragmatists' ontological perspective on development as "becoming" (see for example Packer & Goicoechea, 2000). From this perspective development is about taking one's learning and new knowledge onwards into new contexts and ways of knowing and being in the world (Carr, 2004b, p.6).

Progression as the development of "expert" knowledge

Theories of the structure of expert knowledge could be useful for determining progression in key competencies because they link developmental and socio-cultural perspectives. This literature emphasises the move from a novice stage, where there is a focus on learning facts, rules and procedures, through an "associative stage" where understandings strengthen and performance becomes more fluent. Finally in the "autonomous stage" knowledge is used with speed and accuracy, and procedures are applied more appropriately and rapidly and in some cases more creatively. These ideas underpinned the criteria for "fluency" of performance used in conjunction with the EFF standards described in Section Six.

The development of expertise requires both experience and effort. Ericsson (1996) argues for "deliberative practice" which:

- Is at an appropriate level of difficulty;
- Is accompanied by informative feedback;
- Provides many opportunities for repetition; and
- Provides opportunities to correct errors.

These conditions are clearly compatible with the idea of supporting students to become lifelong learners (*see* section Three) and with the socio-cultural notion of scaffolding learning.

Brewerton (2004b) notes that that expertise is context dependent because experts must show they can use and integrate their competencies in complex contexts (p. 47). Discussing the implications of this for the curriculum and for assessment she concludes:

> This understanding of expertise suggests that learners need to have opportunities to use, practise and adapt their skills in a range of increasingly complex contexts. This supports the need for key competencies to be integrated into all curriculum areas and levels, as well as the need for a common understanding of skills so that learning from other contexts can be built on (p. 47).

This recommendation introduces the interesting dimension of complexity of the context as an aspect of progression. This is discussed further in the example below.

Drawing on teachers' professional judgements

In their work on the assessment of complex performances Delandshere and Petrosky (1998) suggest working backwards from the desired outcomes to determine the characteristics of what should be assessed and how. The key question according to Delandshere is "what does it mean to know?" (Delandshere, 2002). She asserts that assessment should be a process of inquiry into the consequences (the ultimate meaning) of learning. This is coherent with pragmatist philosophy in which *consequences* are a measure of "truth" (Biesta & Burbules, 2003; Johnson & Onwuegbuzie, 2004) and also with the DeSeCo focus on ways students use their competencies to act in new situations. If such a process was followed, the question of progression could be addressed from the perspective of teachers' existing professional knowledge about desirable and achievable outcomes for students at different stages of their schooling.

In our direct experience, this is the process that already underpins curriculum development. For example the Health and Physical Education in the New Zealand Curriculum describes progressions for three factors of direct relevance to the key competency *relating to others*:

> come to understand the nature of relationships;
>
> increase their understanding of personal identity and develop sensitivity to, and respect for, other people;
>
> use interpersonal skills effectively to enhance relationships (Ministry of Education, 1999b).

These progressions were described by an iterative process in which expert teachers worked together and then consulted with wider reference groups (Tasker, 2004). Since such progressions are already in use, it seems sensible to draw on them when developing assessment approaches to the key competencies.

The use of existing curriculum progessions and teacher judgement could help avoid the pitfalls of an excessive emphasis on ladder-like developmentalism, making spaces for progressions more akin to the climbing frames or bouncy castles envisaged by those who advocate for a more complex view of progression (for example McGivney, 2002). Of course developmentalist assumptions are likely to inform teachers' professional judgements at least to some degree, so the distinction being drawn is not straightforward.

The New Basics case study illustrates how a "professional judgement" design process might play out. Interestingly, the principles used are directly linked to pragmatist principles through the use of Dewey's work. He argued that curriculum and instruction should be based on integrated community-based tasks and activities that involve learners in pragmatic social action of value in the real world (Education Queensland, 2000, p.51). New Basics used an iterative process in which a range of teachers, teacher educators and educational researchers were involved in designing rich tasks that they believed could assess the identified competencies at three curriculum transition points (at the end of years 3, 6, and 9). Once the draft tasks had been designed they were mapped to the key competencies (the so-called New Basics) and to the range of existing curriculum documents across the school levels. This ensured that the tasks envisaged would draw on

knowledge and skills that were part of the various curricula leading up to each assessment point, allowing assessment to be integrated with teaching and learning. Thus the progression judgements involved were of suitable tasks to allow students to demonstrate complex performances, not of determining features of the learners per se.

Using insights from a range of research projects

Our final example of a possible process for determining progressions was chosen because it was already familiar, but also because the work was relatively small-scale and hence shows what can be done without a burdensome financial input. The competency explored is the ability to carry out scientifically designed research — a performance of "thinking" as the New Zealand key competencies are currently defined, although not a "key competency" *per se.* This example shows how the *generic* idea of key competency might translate into progressions linked to a specific curriculum learning area.

A recent NEMP probe study drew on a wide range of research, drawn eclectically from both developmental psychology and science education literature related to children's science investigations. The research concluded that it was possible to identify five clusters describing a sequential development of meta-level characteristics of children's science thinking. Rather than drawing on separate aspects of development, the research pointed to possibilities for clustering four types of attributes. In this context they were: views of the nature of knowledge; meta-cognitive awareness; meta-level views of the purposes for which activities such as 'investigations' are carried out, by children and by scientists; and mental models of causality. All these factors appeared to develop in ways that are in broad alignment within each of five developmental clusters (Hipkins & Kenneally, 2003).

At the same time as these clusters were being theoretically derived, 200 NEMP tapes of children carrying out simple science investigations were analysed for any evidence of progression between years 4 and 8. Although the tasks constrained the range of types of evidence that could be observed, there was strikingly little change seen, except in the level of complexity of the contexts used. Furthermore, the tasks appeared to challenge children to carry out much the same sequence of activities (except for formal reporting) that would be required of a student at year 11 — level one of the NCEA. While the *theory* suggested that much more could be expected of learners as they progressed through school, the *empirical evidence* pointed towards progression seen as increasing complexity of contexts in which tasks were set, rather than competencies. As noted above, increasing complexity of the context is one possible means for determining progression within socio-cultural and "expert" theoretical frameworks. However the NEMP probe study outlined here suggests that contextual challenge should be considered alongside other factors, so that the possibilities for students to learn and grow are not inadvertently proscribed.

The findings of this project provide a caution to use of the pragmatist approach described above. While teacher professional judgement is a rich source of expertise and should be made better use of, it has its limitations. Teachers may not be aware of assumptions that limit expectations that, at

least according to the research, could be expanded. Perhaps the most important question to ask is how to take account of both theory and practical judgement. While examples of such an approach have been described for horizontal progression (for example the EFF standards, the VET standards, and the PISA levels for problem solving at age 15) we have not found clearly described examples where this has been done for vertical progressions.

Conclusions

Vertical progressions are complex and need to be developed and used with caution if they are to support students' sense of themselves as learners whose progress is acknowledged. As for horizontal progressions, qualitative descriptors of clusters of aspects to be sought at any one level may need to be defined.

While existing curricula may provide a useful guide in some aspects of key competency progressions, it will be important not to create simple linear or age-specific progressions that could inadvertently limit learning expectations of students — either by themselves or their teachers.

The SOLO taxonomy is the most widely used developmental theory we have found for determining vertical progressions. Work that draws on this is very detailed and clearly involves expensive and time-consuming research processes. It is possible to determine progressions, at least for subject-specific aspects of competency, that draw on other types of theoretical perspectives and research.

It would appear most manageable to use a combination of teacher professional judgement and carefully debated theoretical perspectives if vertical progressions are to be designed. However the role of the context in setting the complexity level of assessment tasks should not be overlooked.

8. Where to next?

This review took as a starting point the assumption that an assessment of a key competency is an assessment of a complex performance that integrates skills, behaviours, attitudes and motivations, values, and knowledge.

As outlined in Section Two, such performances take place in real-life contexts. A competent person will be able to show they can adapt their performance appropriately as details of the context change. Section Four shows how this means that more than one assessment event is likely to be necessary. Furthermore, the "success" of the performance will depend to some extent on the nature of the interactions that are made possible in that context.

As just outlined, the key competencies are clearly intended to be situated within a socio-cultural theoretical framework. Competency does not reside in individuals alone. This has implications for the theories of learning and pedagogy that are adopted, and hence for the theoretical underpinnings of any assessment system that will be compatible with the demonstration of competencies as complex performances.

Figure 1 presents a summary of the three theoretical perspectives on assessment outlined in Section Three and locates the case studies within selected aspects of this framework. This diagram illustrates the range of choices possible for policy makers when aligning assessment purposes with the features of an assessment system.

Figure 1 **The fit between the three theoretical perspectives on assessment and the case studies**

Framework	Critical-theoretic	Interpretive	Empirical-analytic
	S V L	N	P
Purpose of assessment	To support lifelong learning and learner empowerment S V L	To consider the knowledge, skills, and values learners are acquiring and improve teaching and learning N	To provide certification and systems accountability P
Assessment focus	Developing and interpreting performance through a focus on context, process, and product S V L	Understanding achievement through a focus on process and product N	Measuring behaviour through a focus on product P
Nature of assessment	Co-constructed formative (Learners are involved in selecting, discussing, and judging the evidence) V S L	Formative (Learners are involved in discussing the process but not the final judgement) S N	Summative (Learners are not involved/experts validate) V N L P
The place of attitudes and values	Attitudes and motivations are discussed with learner L	Attitudes and motivations are inferred S N V P	Attitudes and motivations are assessed separately P
Relationship of key competencies to the curriculum	Key competencies are the main emphasis of the curriculum S N	Key competencies are integrated with other aspects of the curriculum V L N	Key competencies are discrete from other aspects of curriculum P
Relationship to other curriculum assessments	Key competencies are assessed holistically with other aspect of the curriculum S N	Key competencies are assessed discretely but using the same tasks as other assessments V L	Key competencies are assessed discretely from other aspects of the curriculum P
Relationship to other key competency assessments	Key competencies are assessed holistically S N	Key competencies are assessed in sub-clusters	Key competencies are assessed discretely V P L
Nature of assessment tasks	Tasks are authentic and locally-designed/selected S V L	Tasks are adapted locally N	Tasks are generic with contextual links/expert designed P

Key

Learning Stories (Sector = early childhood)	S
New Basics (Sector = primary and secondary)	N
PISA Problem-solving (Sector = secondary)	P
Vocational Education and Training (Sector = tertiary)	V
National Institute for Literacy (Sector = tertiary)	L

This analysis demonstrates the manner in which the dominant purpose envisaged for assessment influences the choice of a paradigm to drive the development of the assessment system. Some flexibility between paradigms is possible. The table shows that each case study is not always placed within the same paradigm for all the aspects indicated, or may have elements that locate it in two paradigms. This is especially the case for assessment systems that are designed to provide information for both formative and summative purposes, as illustrated by the VET and NIFL case studies (*see* Section Four).

Important policy considerations

From this analysis a number of key issues and themes have emerged. These point to specific areas where it will be important that there is a coherence between theory, policy and practice as any system proposed to assess key competencies is developed. The key messages emerging from these issues and themes are summarised next.

Key messages about assessment purposes

On the whole there is considerable agreement among commentators about the purpose of assessing key competencies. Most of the case studies point towards a system that places lifelong learning and learner empowerment at the heart of the assessment. However, most of the assessments also have a second purpose, which is to evaluate progress for reporting and/or accountability.

Achieving a coherent alignment of these potentially conflicting purposes will be challenging. The view that emerges in the literature is that an assessment system that provides rich information for teaching and learning should take design precedence. This is necessary because assessment information for reporting purposes will almost inevitably involve a reduction in this richness (for a comprehensive summary of this argument see LeMahieu & Reilly, 2004) and carries costs to the learner (Carr & Wylie, 2004).

Key messages about the theoretical underpinning of the model

Many authors comment on the importance of incorporating a range of theoretical understandings, as well as the varied perspectives of stakeholders, into the key competency assessment model. Oates (2001) notes that, where implementation of similar initiatives have been successful in England, the initiative has been supported by a carefully constructed model which reflects a sound theoretical base and which has clear connections with pedagogy.

The theoretical understandings that have already been incorporated into the key competencies model are those developed from the DeSeCo framework. From this review of the literature, we suggest that other theoretical understandings are also important to incorporate. These include but

should not be limited to current understandings about horizontal and vertical progressions. One model which is particularly prevalent in the case study literature, and which has been used in New Zealand by the asTTLe team is the Structure of the Observed Learning Outcome (SOLO) taxonomy (Biggs & Collis, 1982). However we note that asTTLe also drew on a rich theoretical framework for understanding the development of knowledge in the relevant curriculum area. It will be important to consider such theoretical perspectives in every curriculum learning area as the key competencies are developed.

Other debates about the framing of the key competencies model suggest it will also be necessary to:

- ensure that the language used to frame the key competencies reflects the complexities of the model, and leads educators towards the teaching and assessment paradigm being emphasised; and
- offer a flexible framework which outlines core generic key competencies but also allows for the inclusion of other specific competencies that are valued by particular cultures or groups. This approach could avoid concerns expressed internationally about the existence of "generic skills".

Key messages about process

In the case studies we selected, and in many other assessment systems (for a meta-review see McMillan, 2000) the initial construction of the assessment initiative is underpinned by a set of assessment principles which guide the development process. These principles typically incorporate understandings about the purposes of the assessment and the process to be followed in developing the assessment system. For example, one NIFL principle is:

> The levels defined in the EFF Assessment Framework must be the products of a national consensus-building process that assures portability of certificates and credentials (Stein, 2000, p.58)

Many of the case studies show how teachers and other stakeholders can have a role in the development of the assessment system. Given this, it appears vital that the Ministry of Education continues with the co-constructed processes currently underway as part of the curriculum stocktake. A model for such co-construction of an assessment system is provided by the NIFL case study (*see* Sections Four and Six).

Placing educators at the centre of this process has a number of benefits:

1. A co-constructed process allows for theoretical expert knowledge and teacher practice to be combined. This supports the incorporation of current theoretical understandings underpinning each key competency while also affirming and supporting teachers' expertise.
2. A co-constructed process allows time to shape developmental progressions that need to be based on a combination of theoretical expert knowledge and teacher practice.

3. This model supports the key stakeholders (teachers and educators, learners, parents, and employers) to deepen their understanding of the theoretical underpinnings of the key competencies and in particular to understand how they differ from the existing essential skills. This should increase their ability to support learners to develop key competencies, and their ability to assess this development effectively. It should also enhance community understanding of any reports generated by the assessment process (Stein, 2000).

4. The readings from this literature review have led us to become increasingly aware of the importance of alignment of curriculum, pedagogy, and assessment. Depending on how the co-construction is managed it could also support this alignment to ensure that there is consistency between curriculum areas.

When discussing the avoidance of the pitfall of current key skills/competencies initiatives, Oates (2001) notes that successful implementation has been supported by adequate levels of staff development to ensure that parties concerned (teachers, trainers, and learners) understand the underpinning principles and theoretical perspectives. A framework for this process would need to be developed. Given the cross-curricula nature of key competencies, and current understandings of good practice professional development, one model could be to locate this professional development within the framework of school-wide learning communities, especially in secondary schools where teachers work in separate curriculum areas.

Key messages about assessment type

There is considerable agreement among commentators about the types of assessment best suited to the evaluation of complex performances such as key competencies.

- Most agree that **new forms of assessment** are needed to assess complex performances. These forms of assessment move us away from standardised testing towards forms which promote co-constructed formative assessment.

- Most suggest that **more than one form of assessment** is needed to adequately deal with issues of reliability and validity.

- Most suggest that **more than one task and more than one form of evidence** are needed to adequately deal with issues of reliability.

- **Portfolio-type self-assessments or observations** of key competencies grounded within **authentic learning situations** are suggested as methods that are better suited to supporting learners to develop key competencies and for validly assessing key competencies.

- Most suggest that the learner should be **involved in discussion about progress** or in making **decisions about selecting the evidence**. Most suggest that more than one person should be involved in decisions about judging or validating the evidence. Some suggest that the learner should be involved in decisions about **judging the evidence**.

- Commentators agree that key competencies are **performed holistically** — that is more than one key competency is drawn on in any given situation, but have less agreement about whether key competencies should **be assessed holistically or discretely**. Models for both forms of assessment are provided in the case studies.

93

- Students develop key competencies in many settings. Therefore exploring ways to capture the learning that occurs outside of school, through the involvement of parents, and members of the students' wider community, could be incorporated into the model. The tertiary case studies provide models that illustrate how key competencies developed in settings outside tertiary institutions can be incorporated. New models would need to be developed for the early childhood, primary, and secondary sector. This could be achieved through the inclusion of student-parent-teacher conferences as part of the discussion of students' portfolios. Such a process could also support teachers to provide continuity across home and community settings.
- There is a tension between the provision of rich tasks and manageability. For this reason assessments need to be **manageable** and connected with existing practices, that is, aligned with curriculum and pedagogy. A mapping of the curriculum is suggested to provide an overview of how key competency assessment could fit within elements of existing assessments. NCEA assessment instruments may need to be revised in the light of this mapping.
- Most commentators propose a key competencies **assessment and reporting system that is standards-based**. This aligns with current directions in assessment in New Zealand.

Overall summary

The understandings developed in this review strongly suggest that supporting lifelong learning and learner empowerment should be a key purpose for assessing key competencies. Accordingly, this purpose would need to drive any assessment system that is developed. This requirement is not necessarily incompatible with assessment for other purposes such as accountability, if care is taken with the design of the assessment system. A range of assessment options and tools exist and these should be further investigated as no one overall "best" solution emerged from the analysis.

A coherent and defensible system would need to incorporate current theoretical understandings about progression, about socio-cultural theories of learning, and about the components of each key competency, and to appropriately integrate all these with teachers' existing professional expertise. We suggest that the Ministry continues with the current co-construction process to develop the assessment system, and extends the process to include cross-curricula mapping and assessment development.

Teacher professional development, community consultation and a system that provides coherence between assessment, curriculum, and pedagogy will all support better understanding of the key competency model and increase the likelihood of appropriate and well-supported implementation.

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.

Alison, J. (2005). *Teachers talk about NCEA*. Wellington: Post Primary Teachers Association.

Alton-Lee, A. (2003). *Quality teaching for diverse students in schooling: Best evidence synthesis*. Wellington: Ministry of Education.

Apple, M. (2004). *Ideology and Curriculum: Third Edition*. New York: RoutledgeFalmer.

Bartlett, J. (2005). *Developing independent learners. The planning and implementation of an inquiry-based curriculum integration programme at Kuranui College*. Wellington: New Zealand Council for Educational Research.

Biesta, G., & Burbules, N. (2003). *Pragmatism and Educational Research*. Lanham, Boulder, New York, Toronto, Oxford: Rowman and Littlefield.

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

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box. Assessment for learning in the classroom.* London: nferNelson.

Black, P., & Wiliam, D. (1998). *Inside the black box. Raising standards through classroom assessment.* London: nferNelson.

Blom, K., & Clayton, B. (2003). *We can't teach them that! Reinstating the place of generic skills in VET.* Paper presented at the The Changing Face of VET, Sixth Australian VET Research Association Conference, 9-11 April, Australian Technology Park, Sydney. http://www.avetra.org.au/abstracts_and_papers_2003/refereed/Blom.pdf

Brewerton, M. (2004a). *Reframing the Essential Skills: Implications of the OECD Defining and Selecting Key Competencies Project. A background paper for the Ministry of Education*. Wellington: Ministry of Education (unpublished).

Brewerton, M. (2004b). *Thoughts on the characteristics of a 'Successful School Leaver'.* Wellington: Ministry of Education.

Broadfoot, P., & Black, P. (2004). Redefining assessment?The first ten years of *Assessment in Education*. *Assessment in Education, 11* (1), 7-27.

Bryce, J. (2003). *Constructing intra and interpersonal competencies in a context of lifelong learning.* Unpublished PhD, Royal Melbourne Insitute of Technology University, Melbourne.

Bryce, J., & Withers, G. (2003). *Engaging secondary school students in lifelong learning*. Melbourne: Australian Council for Educational Research.

Bye, J. (2001). Youth-in-transition - managing the risks. In F. Beven,C. Kanes & D. Roebuck (Eds.), *Knowledge Demands of the New Economy. Proceeds of the 9th Annual International Conference on Post-compulsory Education and Training* (Vol. 1, pp. 122-128). Bowen Hills, Queensland: Australian Academic Press Pty Ltd.

Carr, M. (1998a). *Assessing children's experiences in early Childhood. Final Report to the Ministry of Education*. Wellington: Research Division, Ministry of Education.

Carr, M. (1998b). *Assessing children's learning in Early Childhood settings. A professional development programme for discussion and reflection. Three videos and a support booklet.* Wellington: NZCER.

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

Carr, M., & Claxton, G. (2002). Tracking the development of learning dispositions. *Assessment in Education, 9* (1), 9-37.

Carr, M., May, H., & Podmore, V. (2000). *Learning and teaching stories: Action research on evaluation in early childhood. Final report to the Ministry of Education*. Wellington: New Zealand Council for Educational Research.

Carr, M., & Peters, S. (2005). *Te Whariki and links to the New Zealand Curriculum Research Projects 2005: Final Report*. Hamilton: University of Waikato.

Carr, M., & Wylie, C. (2004). *Assessment of key competencies in school entry assessment*. Waikato and Wellington: University of Waikato and New Zealand Council for Educational Research.

Clayton, B., Blom, K., Meyers, D., & Bateman, A. (2003). *Assessing and certifying generic skills. What is happening in vocational education and training*. Station Arcade, South Australia: NCVER. http://www.ncver.edu.au/publications/1351.html

Codd, J., Brown, M., Clark, J., McPherson, J., O'Neill, H., O'Neill, J., Waitere-Ang, H., & Zepke, N. (2002). *Review of future-focused research on teaching and learning*. Wellington: Ministry of Education.

Curtis, D. (2003). *Assessing employability skills: The case of problem-solving.* Paper presented at the Sixth Australian VET Research Association Conference: The Changing Face of VET, Australian Technology Park, Sydney. http://www.avetra.org.au/abstracts_and_papers_2003/Curtis.pdf

Curtis, D., & Denton, R. (2003). *The authentic performance-based assessment of problem-solving*: NCVER. http://www.ncver.edu.au/research/proj/nr2007.pdf?PHPSESSID=6567bb66ad7c28284fa7b59c133ced24

de la Harpe, B., & Radloff, A. (2000). Informed teachers and learners: the importance of assessing the characteristics needed for lifelong learning. *Studies in Continuing Education, 22* (2), 169-182.

Delandshere, G. (2002). Assessment as inquiry. *Teachers College Record, 104* (7), 1461-1484.

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

Denton, R. (2003). *Down-to-earth key competencies assessment - a very practical perspective.* Paper presented at the "The Changing Face of VET, Sixth Australian VET Research Association Conference, 9-11 April, Australian Technology Park, Sydney. http://www.avetra.org.au/abstracts_and_papers_2003/Denton.pdf

Doyle, M. (2002). *Learning to transfer: The distance learner and transfer of learning*. Wellington: Victoria University of Wellington (unpublished doctoral dissertation).

Dweck, C. (1999). *Self-Theories: Their role in motivation, personality and development.* Philadelphia: Psychology Press.

Education Queensland. (2000). *New Basics project technical paper*. Brisbane: Queensland Government.

Ericsson, K. (1996). *The road to excellence*. Mahwah, N.J.: Erlbaum.

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

Gilomen, H. (2003). Desired outcomes: A successful life and a well-functioning society. In D. Rychen & L. Salganik (Eds.), *Key competencies for a successful life and a well-functioning society.* Cambridge, MA. and Gottingen,: Hogrefe and Huber.

Hall, K., & Burke, W. (2003). *Making formative assessment work*. Berkshire: Open University Press.

Hargreaves , D. (2004). *Learning for life: The foundations of lifelong learning*. Bristol: The Policy Press.

Haskell, R. (2001). *Transfer of learning: Cognition, instruction, and reasoning*. San Diego, CA: Academic Press.

Hattie, J., & Brown, G. (2004). *Cognitive processes in asTTle: The SOLO taxonomy. asTTle Technical Report 43.* Auckland: University of Auckland/Ministry of Education.

Higgins, J. (2002). Young people and transition policies in New Zealand. *Social Policy Journal of New Zealand, 18*, 44-61.

Hipkins, R., Conner, L., & Neill, A. (in press). *Shifting Balances 2: The impact of NCEA implementation on the teaching of geography and home economics*. Wellington: New Zealand Council for Educational Research.

Hipkins, R., & Hodgen, E. (2004). *National Survey of Secondary Schools 2003*. Wellington: New Zealand Council for Educational Research.

Hipkins, R., & Kenneally, N. (2003). *Using NEMP to inform the teaching of scientific skills: A report to the NEMP Board.* Wellington: New Zealand Council for Educational Research.

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.

Howley, A., Spatig, L., & Howley, C. (1999). Developmentalism deconstructed. In J. Kincheloe,S. Steinberg & L. Villaverde (Eds.), *Rethinking intelligence: Confronting psychological assumptions about teaching and learning*. New York: Routledge.

Jackson, B. (1995). Assessment practices in art and design: a contribution to student learning? In G. Gibbs (Ed.), *Improving student learning - through assessment and evaluation*. Oxford: Oxford Centre for Staff Development. www.city.londonmet.ac.uk/deliberations/ocsd-pubs/islass-jackson.html. Retrieved 20 April, 2005.

Johnson, R., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33* (7), 14-26.

Kalantzis, M., Cope, B., & Harvey, A. (2003). Assessing multiliteracies and the new basics. *Assessment in Education, 10* (1), 15-26.

Kearns, P. (2001). *Generic skills for the new economy: Review of research*. Station Arcade, South Australia: NCVER. http://www.ncver.edu.au/research/proj/nr0024.pdf

Keightley, J., & Keighley-James, D. (2001). *Negotiating multiple interests in high-stakes assessment: getting inside construct validity*. Research Report No. 1. Wayville: Senior Secondary Assessment Board of South Australia (SSABSA) www.ssabsa.sa.edu.au/research/highstks.pdf. Accessed 10 April 05.

Kelly, F. (2001). *Defining and selecting key competencies: A New Zealand perspective*: Swiss Federal Statistics Office.

Kennedy, C. (2005). *The BEAR assessment system: A brief summary for the classroom context*. Berkeley: University of California. http://bearcenter.berkeley.edu, retrieved 21 April 05.

Kress, J., Norris, J., Schoenholz, D., Elias, M., & Seigle, P. (2004). Bringing together educational standards and social and emotional learning: Making the case for educators. *American Journal of Education, 111* (1), 68-69.

Land, R. (2004). *Teachers talking about New Basics: Teachers' stories on trialing the New Basics framework, 2000-2003*. Brisbane: Education Queensland.

LeMahieu, P., & Reilly, E. (2004). Systems of coherence and resonance: Assessment for education and assessment of education. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability.* (pp. 189-202). Chicago: University of Chicago Press.

Lemke, J. (2002). Becoming the village: education across lives. In G. Wells & G. Claxton (Eds.), *Learning for Life in the 21st Century*. Oxford: Blackwell.

Lingard, B., & Mills, M. (2003). Teachers and school reform: Working with productive pedagogies and productive assessment. *Melbourne Studies in Education, 44* (2), 1-18.

McCurry, D., & Bryce, J. (1997). *The school based key competencies levels assessment project*. Melbourne: Australian Council for Educational Research.

McDonald, B. (2001). *Transfer of training in a cultural context: A Cook Islands study*. Wellington: Victoria University (unpublished doctoral dissertation).

McGivney, V. (2002). *A question of value: Achievement and progression in adult learning: A discussion paper.* Leicester: National Institute of Adult Continuing Education (NIACE).

McMillan, J. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research and Evaluation, 7* (8), http://PAREonline.net/getvn.asp?v=7&n=8. Retrieved 10 April 05.

McNaughton, S. (2002). *Meeting of minds*. Wellington: Learning Media Ltd.

Ministry of Education. (1993). *The New Zealand Curriculum Framework*. Wellington: Learning Media.

Ministry of Education. (1999a). *Guidelines for environmental education in New Zealand schools*. Wellington: Learning Media.

Ministry of Education. (1999b). *Health and Physical Education in the New Zealand Curriculum*. Wellington: Learning Media.

Ministry of Education. (2002). *Planning for better student outcomes*. Wellington: Ministry of Education.

Ministry of Education. (2004). *Assessment for Learning: Early Childhood Exemplars 2. Sociocultural Assessment.* Wellington: Learning Media.

Ministry of Education. (2005a). *Key Competencies in tertiary education*. Wellington: Ministry of Education.

Ministry of Education. (2005b). *Key Competencies: The New Zealand Curriculum/Marautanga Project*. Wellington: Learning Media.

New Zealand Qualifications Authority. (1998). *National Certificate in Employment Skills: Assessment guide*. Wellington: New Zealand Qualifications Authority.

Oates, T. (2001). *Key skills/competencies - avoiding the pitfalls of current initiatives*. London: Qualification and Curriculum Authority (unpublished paper).

OECD. (2004). *Problem solving for tomorrow's world: First measures of cross-curricular competencies from PISA 2003*. Paris: Organisation for Economic Co-operation and Development.

Orland-Barak, L. (2005). Portfolios as evidence of reflective practice: what remains "untold". *Educational Research, 47* (1), 25-44.

Packer, M., & Goicoechea, J. (2000). Sociocultural and constructivist theories of learning: ontology, not just epistemology. *Educational Psychologist, 35* (4), 227-241.

Panizzon, D. (2003). Using a cognitive structural model to provide new insights into students' understandings of diffusion. *International Journal of Science Education, 25* (12), 1427-1450.

Panizzon, D. (2005). *Assessment and teaching practices to enhance student learning: Experiences of secondary Australian science teachers.* Paper presented at the American Educational Research Association (AERA) conference,, Montreal, 11-15 April.

Pegg, J., & Inglis, M. (2005). *Setting the context: Current challenges concerning assessment: An Australian perspective.* Paper presented at the Paper presented at American Educational Research Association Conference, Montreal, 11-15 April.

Queensland Government. (2004). *The New Basics Research Report*. Brisbane: Department of Education and the Arts, State of Queensland.

Queensland State Education. (2000). *Draft New Basics technical paper, Version 3 April 2000*. Brisbane: New Basics Unit, Queensland Government.

Renwick, W. (1986). *Moving targets: Six essays on educational policy*. Wellington: New Zealand Council for Educational Research.

Rutherford, J. (2004). *Key competencies in the New Zealand curriculum: A snapshot of consultation, December 2004*. Wellington: Ministry of Education (unpublished report).

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

Sadler, R. (2002). Learning dispositions: Can we really assess them? *Assessment in Education, 9* (1), 45-51.

Stein, S. (2000). *Equipped for the Future content standards: What adults need to know and be able to do in the 21st century*. Washington DC: National Institute for Literacy.

Suter, L. (2004). Tools for two masters: Classroom assessment and school system assessment. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability.* Chicago: University of Chicago Press.

Tasker, G. (2004). Health Education: Contributing to a just society through curriculum change. In A. O'Neill,J. Clark & R. Openshaw (Eds.), *Reshaping culture, knowledge and learning: Policy and content in the New Zealand curriculum framework*. Palmerston North: Dunmore Press.

Teaching and Educational Development Institute. *Bigg's structure of the observed learning outcome (SOLO) taxonomy*. Brisbane: University of Queensland. http://www.tedi.uq.edu.au/downloads/Biggs_Solo.pdf. Retrieved April 20 2005.

Tertiary Education Commission. (2003). *Evaluation of the Gateway pilot: Final report.* Wellington: Tertiary Education Commission.

Texas Centre for the Advancement of Literacy Learning. (2003). Equipped for the future - an update. *Literacy Links, 8* (1), 1-5. http://www-tcall.tamu.edu/newsletr/dec03/dec03d.htm, retrieved 09 April 05.

Thomas, G., Tagg, A., Holton, D., & Brown, G. (2002). *Numeracy item signature study: A theoretically derived basis: Technical report 25*. Auckland: School of Education, University of Auckland.

Tittle, C. (1994). Toward an educational psychology of assessment for teaching and learning: Theories, contexts and validation arguments. *Educational Psychologist, 29* (3), 149-162.

Vaughan, K., & Kenneally, N. (2003). *A Constellation of Prospects: A Review of the Secondary-Tertiary Alignment Resource (STAR).* Wellington: Ministry of Education.

Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. San Francisco: Jossey-Bass.

White, R. (2003). Changing the script for science teaching. In R. Cross (Ed.), *A vision for science education: Responding to the work of Peter Fensham.* London and New York: RoutledgeFalmer.

Wiliam, D. (2001). *Level best? Levels of attainment in national curriculum assessment.* London: Association of Teachers and Lecturers.

Willms, J. (2003). *Student engagement at school: A sense of belonging and participation. Results from PISA 2000*. Paris: Organisation for Economic Cooperation and Development.

Wilson, M. (Ed.). (2004). *Towards coherence between classroom assessment and accountability*. Chicago: University of Chicago Press.

Wilson, M., & Draney, K. (2004). Some links between large-scale and classroom assessments: The case of the BEAR assessment system. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability.* Chicago: University of Chicago Press.

Wylie, C., Hodgen, E., & Ferral, F. (2005). *Completely different or a bigger version? Experiences and effects of the transition to secondary school. Draft report to Ministry of Education.* Wellington: New Zealand Council for Educational Research.

Wylie, C., Thompson, J., Hodgen, E., Ferral, H., Lythe, C., & Fijn, T. (2004). *Competent children at 12*. Wellington: New Zealand Council for Educational Research.

Wylie, C., Thompson, J., & Kerslake Hendricks, A. (1996). *Competent children at 5: families, early education, and schools*. Wellington: New Zealand Council for Educational Research.

Wylie, C., Thompson, J., & Lythe, C. (1999). *Competent children at 8: Families, early education, and schools*. Wellington: New Zealand Council for Educational Research.