Assessing adult learning: literacy and numeracy competencies

NZCER Conference Proceedings

August 2010

Wellington 2010

NZCER PRESS

New Zealand Council for Educational Research
PO Box 3237
Wellington
New Zealand

ISBN 978-1-877293-82-5

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

Contents

1. Welcome and introduction

Jessica Hutchings and Robyn Baker

Jessica Hutchings

Welcome to NZCER's conference on assessing adult learning. My name is Dr Jessica Hutchings and I am the Manager of the Māori Research Unit Te Wāhanga at NZCER. We hope that today's conference and korero and discussion will combine opportunities for you to all peel back layers to be able to deepen and broaden our individual and our collective understanding about the many diverse ways that we think about assessment. It is my personal hope that this conference reminds us that assessment is culturally located and that this serves as an important reminder for us to check whose ideologies and assumptions underpin assessment frameworks. Who is the framer of the frame? This is a critical question when thinking about Māori and assessment and the need for kaupapa Māori or culturally responsive wellbeing assessment frameworks. I am looking forward today to listening to the speakers, to learning more about assessment and to being challenged in my thinking. I am now going to pass over to Robyn Baker, the director of NZCER.

Robyn Baker

My name is Robyn Baker and I am the Director of NZCER and it is my pleasure, too, to welcome you to this conference. NZCER has been supporting education in New Zealand for more than 75 years. In doing this, what we are aiming to do is to contribute to an education that supports all learners to be well equipped for the world ahead of them and to also support a society in which everybody participates and learns through their lives. Now of course many educational organisations share this kind of vision and the unique niche of NZCER is that we're a purpose-built research and development organisation. We undertake research, we publish books and resources, we provide services to the sector and we provide independent advice to whoever is interested in education. Some of you have worked closely with us for a very long time and others won't be so familiar with the organisation. In the 75 years we have largely worked in a school sector, more recently in early childhood but now we're well and truly in the tertiary space in adult learning and in literacy and numeracy in particular. We hold conferences like this on a regular basis, seeking to provide a forum where we examine a current educational issue in a way that provides new insights and extends our thinking. We're very pleased to be doing it this year in this exploration in the context of assessment of adult learning, particularly because we know there haven't been many opportunities for a national debate in this sector and in this particular area of focus.

Over the years assessment has been a big focus of the work at NZCER. We've undertaken research into ways assessment practices can support and of course fund our student learning. One example of the work that we've done recently is the assessment of foundation learning, a project that Rose Hipkins will talk about in her presentation. Another example is some work we've just finished for the Industry Training Federation that focuses on assessment in workplaces and led to the production of a booklet that some of you may have seen that gives advice—practical advice to assessors and tutors about how to support effective assessment and moderation practices in the workplace.

Over the years we've also sought to produce tests and assessments that are based on an agreed framework which might be a curriculum or learning progressions of some sort. And it provides teachers and learners

with valid and reliable information with which to support learning. While we built our theoretical and practical knowledge of assessment working in the school sector we've always been keen to make a contribution in the tertiary sector as well. The development of the assessment tool for adult literacy and numeracy gave us a chance to do just that. We had a successful tender with our partners, the IT company Fronde and the Australian Council for Educational Research, some of whose members of this team are here today and we'll be hearing from them. It was an exciting project for us and for New Zealand because it gave us the ability to bring together our collective expertise and to learn as we sought to design and develop a tool that was purpose built to support the teachers of adult learners in literacy and numeracy and the students themselves. It was also a rich opportunity for NZCER itself as we sought to align our deep knowledge of assessment with the rapidly changing scope that IT offers so that we could produce a tool where you will get fine-grain information rapidly back to learners and teachers. This kind of activity would not have been possible even a few years ago. As you're aware, however, the assessment of learning is complex and can be very contentious. There is no one assessment that will provide a valid and reliable picture of a learner's expertise—as a writer, for example, or in reading comprehension or in numeracy. And it is this complexity that we are keen to explore today. We want a day where there's a focus on the evidence, the research, on the tools of assessment, on the practice and on the actual experience of being assessed.

Today we are going to take a multidimensional look at assessment and consider questions such as: What might the purpose of assessment be? How can assessment support learning? Can an assessment serve multiple purposes and be diagnostic, formative and summative? And how might assessment be viewed through a Māori world view? As we consider these questions and many more it will be very clear that there are many things to weigh up, many choices to make and new things to learn. In planning this day we sought speakers whom we were confident would give rich insights into the complex area of assessment and who would inform and challenge us and would present a range of views. There is unfortunately no one answer, no silver bullet and no easy one solution. It will be up to you to make your own sense of the presentations and to think about it within your own professional practice and take it back to your work, whatever that might be. And today we have a very diverse audience, people from all over New Zealand whom I welcome to Wellington and I thank you for coming. There are practicalities in implementing a day like this; we couldn't get everybody to select and share their deep knowledge. However, the collective knowledge in this room about adult learning, literacy, numeracy, assessment, adult learning and research is very deep and the way we have set up this conference with you sitting at tables is designed to at least provide the opportunity for you to have discussions to share the knowledge you have and to discuss the ideas raised.

2. Panel session

Introduction—Jenny Whatman

Jenny Whatman is a Senior Researcher at NZCER.

We chose to open this conference with a panel of practitioner experts: expert learners, experts at being assessed and at being assessors. The stories from this group of experts will be the reference points for the other speakers during the day and I hope for you, the participants.

As Robyn said, networking is one of the things I specialise in and I've spent most of the year hunting down panellists and I did locate some likely suspects at various forums. Some of the people at this table had to be cajoled and others had to cajole their organisations into letting them participate. But it's really great to have these four people here representing very different kinds of assessment literacy and numeracy experience and knowledge. The stories that they'll tell are a little different–some are a journey and some are a snapshot in time.

Charlie is going to start and by way of introducing him I want to tell my own story, which is a very small story about assessment. Earlier this year I did a certificate in teaching English as a second language. It was offered through a New Zealand organisation but part of a global organisation so the certificate is international and apparently I will be able to teach anywhere in the world with it. The first part of the certificate that I did was a face-to-face intensive over several evenings and weekends with about 20 classmates. It sounds like a really gruelling schedule—I was working during the day and then going to the classes at night but the classmates were great and the teacher was fantastic. Our final assessment was a paired teaching session—we had to present to the rest of the class. It was very carefully scaffolded by the teacher from day one. He gave us a very clear and manageable task to do with clear assessment criteria, we had time to practise our learning and our teaching during the days of the course and we got lots of individual help from the tutor. When we'd successfully taught this short lesson we got very useful constructive feedback from the other class members and from the tutor. Feedback was oral and in writing, and we in turn took part in each other's lessons, providing feedback. That's a fairly familiar story to those of you who are involved as tutors and educators.

However, the second part of the story is not quite so successful. The next two modules we did were completed online. We had a year to complete them but we were warned that unless we did them straight away we'd probably never pass the certificate as most people never do, they never complete it. So because this was an international qualification all of the online teaching materials were relevant to another country and not to New Zealand, they were very boringly and poorly presented and there was a little bit of scaffolding in to what the assessment might look like but not really. I chose to do a course called *Teaching adults at home* because I'd long thought that I would like to work in a voluntary capacity in that area. It turned out to be a very bad choice. I had to learn a whole lot of information about teachers' salaries and various sectors in Canada and answer some very poorly constructed multiple-choice questions about Canada and teaching in Canada. I managed to get 51 percent which was as you imagine a scrape through pass and I felt utterly humiliated by that because I've been a teacher for 30 years and felt I ought to know

quite a lot about teaching adults. The second module I was much more wise. I actually did the boring assignments online, anticipated the kinds of multiple-choice questions I'd get and needless to say flew through that part of the assessment. I think the moral of my story and by way of introducing Charlie is I really regret that I spent all my pent-up emotions shouting at the computer instead of writing to the organisation and saying how disappointed I was having had such a brilliant start to the certificate.

Charlie Hayward

Charlie Hayward is the Literacy and Numeracy Co-ordinator for the Boating Industry Training Organisation. He is completing a Masters in Adult Literacy and Numeracy Education at Auckland University of Technology.

We were asked to talk about our journey into literacy and numeracy and how our experiences along the way have shaped our ideas around assessment. My starting point was in my first job in the Royal New Zealand Navy (RNZN) which is very much a training-oriented organisation. My last posting was to Cambodia as part of the United Nations mission there. I was a team leader of an international riverbase tasked with patrolling a section of the Mekong River and its tributaries. After safely conducting the elections and with tourism on the rise, UN-sponsored ESOL teachers started arriving in Phnom Penh as tourism was seen as the key to the economic development of the country. I met a lovely teacher from Australia and before long, and without quite realising how, I had become a volunteer English teacher in my time off.

I have nothing to offer on assessment from this period as I had no idea what I was doing in the classroom. What I did know, however, was that I had caught the English teaching bug and that this would definitely be a part of my future.

Fast forward a few years and here I am in Japan with one of the high school classes that I taught. I also taught one to one and small groups and the beginning of every lesson was an assessment of sorts. Carefully worded questions would give us information on a student's fluency and whether or not they could use the future tense/past tense/conditional or whatever other grammatical structure we had used to frame the question. What we learnt would influence the content of the lesson we delivered.

The Japanese Ministry of Education has stated that the purpose of English lessons in schools is to give students basic communication skills in everyday life situations. Parents and institutions, however, want to be able to measure the students' progress and often tests, like the Test of English for International Communication (TOEIC), are used to measure this. As a result, English lessons are often held in Japanese by Japanese teachers and students spend their time picking out the subject/verb agreement errors or spotting prepositions that haven't been used correctly and after nine years of "English lessons" are very good at passing tests, but can't actually speak in English. That is why progressive schools are now hiring "gaijin sensei" like me to try to get back to the original objectives of school English programmes.

I think this illustrates an important point that if teachers are made to use a specific assessment method, they may well "teach to the test" and not to the goals of the programme or the needs of the students.

Here is another point on assessment in Japan. The programmes were often fixed and so the quality of each programme was often judged by the quality of the teaching. In one school I worked at there were video cameras in the ceiling and the managers could switch these on at any time to record the lessons. These recordings were then analysed to provide evaluations of the teachers. Another school had microphones in the ceiling and the managers had a switchboard by their desks so they could listen in on lessons. This, too, was a primary means of providing evaluation on teacher performance. Not all teachers were happy with this arrangement!

So these two parts of my life, the Navy and ESOL teaching, came together when the Boating ITO was looking for a literacy and numeracy co-ordinator. I was aware, though, that I needed some upskilling in the area of literacy and numeracy. In my first six months on the job I completed my National Certificate in Adult Literacy Education (Vocational Tutor) by delivering "deliberate acts of literacy and numeracy teaching" to one of our boatbuilding night school classes, and now, as you know, I am doing the MadLitNumEd, or Masters in Adult Literacy and Numeracy Education.

It gives me the opportunity to step away from the coal face of teaching and really think about different approaches to teaching and learning. Ideally, of course, I would have learnt about learning theories first, and then constructed my teaching approach and style based on the ideas and theories that resonated with me. In my case, however, I started ESOL teaching while working in Japan and now, 10 years later and with the advantage of hindsight, I am trying to see where my teaching style fits with the theories I am learning about and having some interesting reflections and a few wake-up calls too.

Now I am over halfway through the Master's course, I can recognise some of the key influential players in this area. They are kind of my heroes at the moment which I hope doesn't sound too sad. I'm thinking about people like Paulo Freire, from Brazil, the father of the critical approach to literacy which recognises a strong political aspect. He was influenced by Marxist and anti-colonialist thinkers and believed that literacy could lead to empowerment for oppressed groups. He taught 300 sugarcane workers to read in 45 days, and they taught some friends, who taught some friends … All this was at a time when to vote in the Presidential elections, literacy was a requirement. You can see how Freire's work was politically motivated, increasing the voting/political power of the underclasses. His most famous book is called *Pedagogy of the Oppressed*. (By the way, I don't know how the election officials measured the literacy of voters, perhaps they had their own "Brazilian National Assessment Tool"!)

But I want to talk about another figure today. His name is Brian Street and he classified all the approaches to literacy and numeracy teaching into two models. These are the "autonomous model" and the "ideological model".

The autonomous model grew out of the "evolutionary framework" of the early 1900s. This model placed human beings along a continuum moving towards "civilisation", a state measured by the adoption of written texts. There were two groups, "civilised" and "not yet civilised" and people talked of "The Great Divide" between these two groups.

In the autonomous model I guess there are two groups too. There are "those who can cope with the demands of living and working to their full potential in a knowledge society" and those who can't. The distinction is measured by "literacy skills". These are a set of value-free, decontextualised technical skills around reading and writing. The same is said to be true with numeracy. Mathematics is a beautiful, finite body of knowledge with lovely rules that hold true in all times and in all places and which can be broken down into manageable chunks and taught. Practitioners who use the autonomous approach tend to focus their lessons around texts. The texts are seen as value-free artefacts that nicely follow all the rules associated with reading and writing. These skills are identifiable and quantifiable and so we can measure people to see which of the skills they have and which of the skills are missing. For this reason it is labelled "a deficit model"—focusing on what is missing—and the job of literacy teachers is to identify the deficit and then to fix the problem by teaching the missing skills.

In this model we hear terms like "deficit" or "lack of" and learners have literacy "issues" or "problems". From here it is only a small step to think that people with low levels of literacy are somehow deficient

themselves and that it is their own fault that they have these problems. Added to this we hear statistics that show that, nationally, innovation and productivity are lower than they could be and that raising literacy and numeracy levels will solve this. It is only another small step to realise that all these people with low levels of literacy and numeracy are costing us and our country money! Slackers!

A fairly rational reaction, then, would be to go and measure the size of the problem, and so we had the Adult Literacy and Life Skills (ALL) and International Adult Literacy Survey (IALS) surveys that plucked adults out of their workplaces and homes and made them read bus timetables and medicine labels so that their ability to cope with the literacy and numeracy demands at work could be evaluated (interesting logic!)—and this gave us some fairly worrying data to consider.

Under the autonomous model, these statistics lead to talk of a "literacy crisis" followed by a wave of "reforms" and hopefully increased funding. But on the down side you can see how the stigma around low literacy skills developed for people who identify with the autonomous model—which is known as the "dominant approach" to literacy as it is the most common approach worldwide and in New Zealand at the moment.

You can see evidence of it at Workbase who encourage ITOs to create literacy profiles which list all the decontextualised literacy and numeracy skills required for a certain job or industry. From TEC we have the Learning Progressions which list all the literacy and numeracy decontextualised skills that we want all adults to have. When we do our NCALE(Voc)/(Ed) courses we learn to "Know the demands", "Know the learner" and then "Know what to do". We could easily rename these three steps making them "Measure the demands", "Measure the learner to identify what skills are missing (what the deficit is)" and "Know what decontextualised literacy and numeracy skills to teach" and then we can see more clearly the deficit model structure at play. And, of course, with all this measuring required to be done, we now have the National Assessment Tool to tie everything together with our pre-learning and post-learning assessments. These kind of tests to measure "literacy and numeracy skills" are consistent with the autonomous model.

Now we come to the "ideological model" of literacy and numeracy which can also be referred to as a social practices approach. It grew as a result of opposition to the autonomous approach and can best be described by David Barton's six propositions:

- Literacy is best understood as a set of social practices; these can be inferred from events which are mediated by written texts.
- There are different literacies associated with different domains of life.
- Literacy practices are patterned by social institutions and power relationships, and some literacies are more dominant, visible and influential than others.
- Literacy practices are purposeful and embedded in broader social goals and cultural practices.
- Literacy is historically situated.
- Literacy practices change and new ones are frequently acquired through processes of informal learning and sense making.

I used to love the autonomous model. When preparing for the interviews for my current job at the Boating ITO I read all the relevant literature I could find. I read the Tertiary Education Strategy incorporating the STEP and all the Learning Progressions, and it all made sense to me.

We assess our learners, find what literacy and numeracy skills they are lacking, "fill them up" and then measure them again to see the gains made in literacy and numeracy skills. It was great to have a plan to

follow and of course as we increase literacy skills across New Zealand, innovation and productivity will go up which makes us national heroes.

When talking to employers, in my efforts "to raise awareness of the issues around literacy and numeracy", I promised them that if we could assess their staff, identify the ones with literacy and numeracy issues and fix them up, then huge benefits would follow. There would be fewer accidents in the workplace, less rework, fewer mistakes, less absenteeism, higher morale and above all, better profits through better productivity.

In my enthusiasm I decided to get the statistics that could back this up. I called everyone I knew but I couldn't find any figures to support my claims, which I found strange and frustrating.

About this time I started to read about *The Literacy Myth* (Graff, 1982). Early writers had been making all kinds of claims around what increased literacy levels would lead to, even to "a metaphorical 'state of grace'", but more recent researchers have argued that actually there is no evidence that increased literacy levels led to all these benefits, and hence my difficulty in locating the data I was after.

I also started reading about some significant flaws in the ALL and IALS surveys and no longer feel comfortable trotting out their findings as fact.

Another Master's assignment had me interview a literacy practitioner. She described the philosophy of her organisation (Literacy Aotearoa) as basically a social practices model. Learners are very gently assessed when they come in, though at this stage the main focus is to acknowledge and honour them for coming in, and there is no set curriculum. Learners set literacy and numeracy goals for themselves and this becomes the curriculum. Evaluation of the progress of a learner is measured by progress towards the goals set by the learner and when these are reached, new goals are set.

My interviewee also mentioned giving the learners a "voice" and a "sense of entitlement" so you can see there is also an element of critical (empowering) literacy at work there. Her provider is a Treaty-based provider acknowledging one kind of literacy for European/Pākehā New Zealanders and another for Māori. Once you recognise that there is more than one kind of literacy, the idea of multiliteracies mentioned in Barton's second proposition makes sense.

Assessment methods under the ideological model could be learner self-assessment, peer assessment, portfolio-type assessment or assessment against the goals of the programme/goals of the learner.

The more I compare the two models, the shakier the foundations on which I based my support for the autonomous model become, pushing me further and further in favour of the ideological model.

What about the IALS and ALL surveys? We hear that 51 percent of New Zealanders cannot cope with the numeracy demands of work, yet 80 percent of those surveyed reported being happy with their maths abilities. Can we assume that New Zealanders routinely overestimate their own abilities or, in fact, that the surveys are finding issues where none exist?

Does increased literacy really solve a large number of social issues or is that the "literacy myth" at work?

Can you measure all the different aspects of numeracy in a 40-minute test? It is a huge area, after all. What if a learner learns some new numeracy that isn't tested on the next test and so no improvement can be observed? Do we label the training that she did of no value?

Lastly, I strongly believe that literacy and numeracy should be measured and taught in context, a point which is a pillar of the social practices approach to literacy. Consider the sentence "Secure the building". For the Navy, that means the working day is over and it is time to shut all the windows, lock all the doors and go home. For the Army, it means surround the building with armed soldiers and monitor all people coming and going. For the Air Force, it means go to the nearest real estate agency and negotiate a 10-year lease with an option to buy. The importance of context cannot be overlooked.

So, where to from here? I think policy makers will continue to utilise the autonomous model to track their return on investment and practitioners will continue to support the ideological model for the benefit of their learners, and hopefully not start "teaching to the test". As a result, pragmatism must be the order of the day as policy makers and practitioners at the coal face of literacy and numeracy teaching try to meet their particular set of objectives. Let's finish with a couple of quotes that reflect this dilemma:

> It seems unlikely that large-scale, publicly-funded programs will operate in other than a quantitative framework. (Reder, 2009)

> The more those ethnographers explain the 'complexity' of literacy practices, the more policy makers find it impossible to design programs that can take account of that complexity. The more ethnographers demonstrate that literacy does not necessarily have the effects that the rhetoric suggests—improved health, cognition, empowerment—the harder does it become for policy makers to persuade funders to support literacy programs. (Street, 2003)

Libby Reardon

Libby Reardon is an Education Officer with the New Zealand Army and is studying towards a Masters in Adult Literacy and Numeracy Education.

I am in the NZ Army Education Corps, working as the Detachment Commander for the Trentham Military Studies Institute. We work closely with tertiary (and other) education providers, encouraging military personnel to enrol on courses, paid for by the army, and with our support throughout their studies. Currently, our corps is focusing on literacy within the wider army training environment.

Like most other members of my corps, I completed a teaching qualification (primary) prior to joining the army two years ago. After enlistment, appreciating the corps' core focus on literacy, I enrolled myself into the Masters of Adult Literacy and Numeracy Education offered by the Auckland University of Technology (AUT) and I'm still completing this.

In terms of progress so far in the army—it's slow, but it is happening. We have begun using the Tertiary Education Commission tool on our recruit courses and we've just tested 150 recruits. The results are on a par with what we're seeing nationwide for that age group, as 19-year-olds or thereabouts. These results have highlighted that we do need to focus on literacy, and clarified specific areas for focus. As an assessment, it was well received by the recruits, but then they live under a constant stream of formal and informal assessment during their four-month period of recruit training. At this stage of their careers, they just plough ahead—they're not going to complain.

Within my role, I'm also conducting a trade literacy-needs analysis across the 30 trades within the army and looking for embedding opportunities. It is a big task, involving hundreds of instructors and courses scattered throughout the country. The TEC tool and associated results provide support in giving this project strength and assist with getting the buy-in from those making the funding and policy decisions. The military can be a bureaucratic nightmare for implementing change, so having those pretty graphs from the TEC assessment makes explaining the need for a literacy strategy much easier!

The AUT Master's has really helped develop my understanding of what literacy is, giving me a global perspective, as well as a sense of what's happening elsewhere within literacy education. It has also allowed me to gain an understanding of what's involved within the nation's trades, within apprentice programmes, in the vocational arena and within other armies. The American army, for example, is instrumental in the literacy/numeracy movement in the States, so they have developed some models there that we can model.

The assessment component of the Master's is through essays, and the extensive background reading required to complete these gives me a rounded understanding of the content, and therefore develops my confidence in presenting my argument to higher ranks and referring them to articles or programmes completed by more "respected" researchers.

I'm lucky that I'm able to discuss the army approach and literacy project with the Master's tutors and other course members. Through this discussion, my tutors offer support and guidance, and can ensure that I remain on track with my recommendations for literacy development within the army. It ensures I am reflecting on the research read, and consulting other specialists before recommending programmes to the army. The qualification puts a bit more power and voice behind my recommendations about the direction I hope the army takes with embedding literacy, language and numeracy (LLN) within our trades.

So it's a good combination, using the TEC assessment results to inform decision making, and the assessment within the Master's to inform my own understanding. In this way, hopefully, I am combining the assessed and being the assessor to improve army literacy.

Lana Moriarity

Lana is enrolled in the Victoria Graduate Diploma of Teaching (Adult Literacy and Numeracy). She has been working as a Literacy Tutor at Rimutaka Prison.

Tena koutou, my name is Lana Moriarity. I am currently midway through a Graduate Diploma of Teaching (Adult Literacy and Numeracy) at Victoria University of Wellington. I work as a volunteer literacy tutor at Rimutaka Prison. Today I am going to talk to you about my experiences of assessment as a teacher trainer in adult literacy and numeracy. In particular, I will focus on the Tertiary Education Commission's (TEC's) *Starting Points Assessment Guide*, which I use in my work in the prison.

Starting Points Assessment Guide was designed by the TEC for learners below Level 1 on the Adult Literacy and Numeracy Progressions. The Learner Progressions assume that learners already have some specific sets of knowledge and skills (such as phonological awareness, concepts about print and letter identification). For an adult who is not yet operating within the progressions, it is necessary to assess each of these sets of knowledge and skills in order to be able to prioritise areas to work on.

Some of these adults may be learning English as a second or other language. However, others may be New Zealand-born citizens who, for a variety of reasons, have not learnt to read or write. The prisoners I work with fit the latter category. These learners have poor phonological awareness and decoding skills. In many cases, they are also not able to recognise sight words such as "I", "You" or "We".

Prior to the publication of the *Starting Points Assessment Guide* I conducted rudimentary assessments with my learners. These generally provided me with vague and unspecific information. Since the publication of the *Assessment Guide* I am now able to collect more useful information on my learners with which to design more effective teaching interventions.

Here, I would like to provide a brief case study to illustrate my point.

Recently I was asked to work with a young Māori prisoner with very low foundation skills. I was told that he liked rugby and mechanics, could speak Te Reo, and was housed in the pods—the medium–maximum security units.

I visited Rimutaka Prison on a sunny Tuesday afternoon to meet the learner and do a diagnostic assessment. Upon my arrival, I was accompanied to the pods by a prison reintegration team manager and the volunteer co-ordinator.

The pods are difficult to describe. As it is a medium to maximum security unit, it is more secure. Essentially, it contains a central guard house from which prison staff can observe prisoners in the recreation area (a concrete basketball area to the right) and their cells.

Learner X was led out from the recreation area and patted down by a guard as I watched. We then went into the programmes room, which contains chairs, a table and a whiteboard. After we had had a chat and introduced ourselves, I conducted an initial diagnostic assessment with the learner which covered:

- the learner's goals
- prior schooling experiences
- what he already knew, and what he needed help with.

Throughout the assessment, the reintegration manager and volunteer co-ordinator sat and observed.

At this time, I didn't have the *Starting Points Assessment Guide* so I used ideas provided in an earlier TEC publication. From this, I was able to gain a basic insight into the learner's level of literacy skill, and determine that the learner was below Level 1 of the Learning Progressions.

Shortly after conducting this initial diagnostic assessment with the learner, the *Starting Points Assessment Guide* was released. This contains a number of different assessment tools that can be used to measure specific aspects of literacy. It also contains planning, recording and reporting templates. The guide builds on and complements the earlier TEC publication *Starting Points*.

The publication of the guide provided me with the chance to do a more comprehensive needs analysis of the learner. I thought this was important because I didn't yet have enough specific information on the learner's skills and knowledge gaps with which to design effective lessons. Conversely, I didn't want to overburden the learner by giving him loads of tests.

In the end I decided it was a good idea to do all of the assessments located in Appendix A of the guide but space them out over three to four lessons.

Once the learner understood the purpose of the subsequent assessments, he was happy to do them.

Although the entire process was time consuming, doing the assessments helped me to find out detailed information on my learner that I could use to inform a truly effective teaching programme based on a "biliterate" or bilingual approach.

I should also note that throughout this entire process, I myself was in the position of a learner, highlighting the concept of AKO.

To conclude, what I would like you to take from today is that the *Starting Points Assessment Guide* is a useful resource to use with learners below the level of the Learning Progressions. It provides specific information on learners, which can be used to inform a more effective teaching plan.

As noted in the guide: "One size does not fit all: valuable instruction time may be wasted if it is not focused on the specific knowledge and skills needed by a particular learner" (Tertiary Education Commission, 2010, p. 5).

http://www.literacyandnumeracyforadults.com/Assessment-for-Learning2/Assessing-starting-points

Dave Tout

Dave Tout is a Senior Research Fellow in the Assessment and Reporting research programme at the Australian Council for Educational Research.

I'm Dave Tout and I feel like the senior member of this panel. I've gone through a lot of the processes and the experiences that the other three have gone through. Certainly in terms of what Charlie was talking about, they were the sorts of experiences that I've had too.

So a little bit about my journey. I started teaching in the 1970s as a secondary school maths teacher and at the end of the 1970s I moved into teaching adults. I've actually been teaching adult numeracy or involved in adult numeracy for over 30 years so I've been around for a long time. When I first started teaching, in my teacher training we read all of the Freire books and about O'Neill and Sumerhill in England and so on. That was the time when alternative education was really powerful. Those early years of teaching were fun years and I think I learnt a hell of a lot from my students and I think I've learnt more about how mathematics works from my adult students than I had ever learnt before from my studies. But certainly as a maths teacher and initially teaching adult numeracy I really only knew about the autonomous model of teaching. I think we maths teachers are pretty straight in that way, and so I only knew about assessing using tests and teaching through textbooks and worksheets.

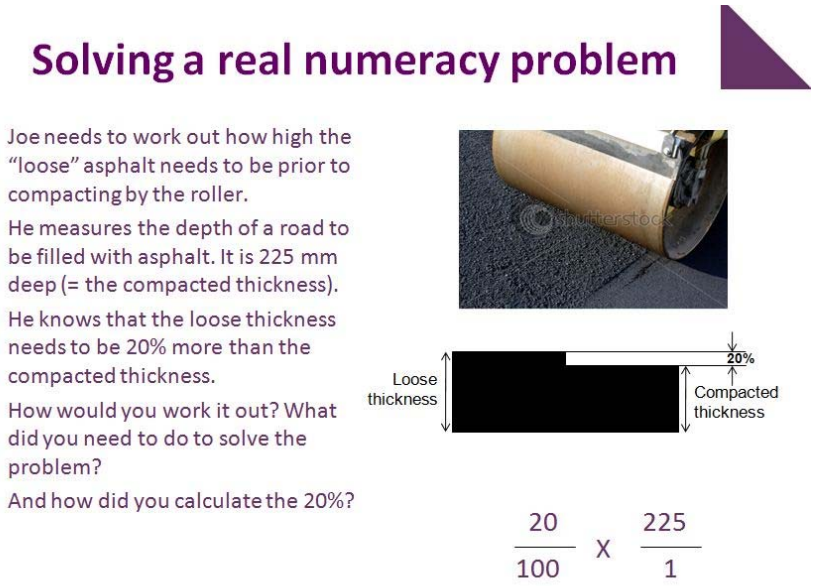
And so it was really challenging for me when I first started teaching adults, to work with literacy and ESOL teachers who were much more innovative than maths teachers. I had to think about what I could do and how I could engage with my numeracy students. And then there was the big question: What was this thing called numeracy? So today I am going to focus a little bit on numeracy because that's my passion.

The 1990s in Australia were a little bit like what you're going through here in New Zealand in the 2000s. That was the time when Australia got a hell of a lot of funding for adult literacy and numeracy and it came out of International Literacy Year in 1990. Australia did a survey in 1989 called No Single Measure which was based on the Kirsch and Mosenthal structure that also sits behind the International Adult Literacy Survey (IALS) and the Adult Literacy and Life Skills (ALL) survey. We had a Labour Government that was supportive and they listened to the productivity issues and arguments, and so a lot of funding resulted. For me the 1990s was an amazing time where I got involved in collaborating with a lot of other people and had to come to grips with this thing called numeracy. I was working with colleagues who were really into the social constructive model of teaching and getting away from the autonomous model. I'd put my hand up and, being a pragmatist, say: How can we do this and take this on board in teaching adult numeracy? I have been involved in most of Australia's assessment programmes, curriculum, you name it. I can be blamed for a lot of things in adult numeracy in Australia!

But for me it was an exciting time. I wrote some of the teacher training programmes for teaching numeracy, taking on board the social constructivist model. I also was involved in writing curriculum like the New Zealand Progressions and our assessment frameworks which was of course always a challenge. The 2000s has seen me move very strongly into assessment. I was involved in the development of the numeracy component of the ALL survey so, again, you could blame me for a lot of the numeracy items that are in the ALL. I was involved with two fantastic colleagues in an exciting project called Rethinking Assessment which was an action research project with about 10 numeracy teachers in my state of Victoria and we produced a publication around that. I've been involved in developing tests for the army, for Corrections in

Victoria. Then I joined the Australian Council for Educational Research (ACER) to work on the New Zealand assessment tool.

And I'm still trying to work out some of the answers to what this thing called numeracy is. For example, I'm always amazed how much numeracy is actually used in the workplace. I haven't yet walked into a workplace where I haven't been staggered by how much maths is actually used and that's pleasing as a maths and numeracy educator, but it's also very challenging as to how you support workers on the ground with their numeracy and their mathematics. I will talk a little bit about that later. But the amazing thing also is that I haven't walked into many workplaces where they actually don't use algebra. That was one of those things as a secondary school maths teacher–students always say "When am I ever going to use algebra?" Well you actually do use it in your work.



So here is a quick example about asphalting. I think there's a question in the assessment tool based on asphalting. I did some work in the asphalting industry and wrote a couple of resources for trainers. Basically, asphalt compacts by about 20 percent so they have to do calculations like this. Workers measure how deep the asphalt has to be and they measure that with their steel ruler and then they have to add 20 percent on to that. So that's a calculation they do on the job all the time. I am not sure how you would have worked it out, but the interesting thing is when you ask people to work out 20 percent of 225, they all do it in different ways. You can divide by five, you can work out 10 percent, double it, there's a whole range of different ways. But when you looked at how these asphalting students were taught, they were told to write down 20 over 100 multiply it by 225 over 1 and cancel down, which is just stupid fundamentally but that was the way they were taught. In other words, this was some maths to do, so let's go back and teach it like it was school maths. But the reality was they would rarely ever do the calculation that way on the job. So to me numeracy is about using mathematics in a really flexible way and teaching the different ways of using maths—in this case, moving the decimal point or dividing by 10 and doubling, or dividing by five (if you're capable of dividing by five).

The thing about numeracy, which is again a bit related to what Charlie was saying about context, is that in the job or in the real world, numeracy isn't about just doing mathematics. The first thing is you've got to decide what you're going to do—what maths do I have to bring to this problem? The first part of solving or

doing a numeracy problem is about literacy—it's about interpreting the problem, making decisions. Then the second step is doing the maths. It might be measuring, it might be calculating, it might be working out 20 percent. The third step is you've got your answer and you need to reflect on it—see if it fits in with the context: "I've got 2,400 mm—I don't want the asphalt up to here, that's way too high. I must have made a mistake; I've got the decimal point in the wrong place."

I could talk for hours about numeracy, about the fact that numeracy to me is about literacy and mathematics and language. To teach and assess numeracy, you've got to be able to assess all of those aspects of numeracy. Why language? To me maths has a language. If I say "one over two plus one over four" it's quite logical to say the answer is "two over six". Fifty percent of your students in adult numeracy class will do that, probably, but it's wrong. But if I say a "one half plus a quarter" most students will suddenly say "well that's three-quarters". It's how you talk about the numbers that makes it make sense, so language is also part of numeracy.

Lynn Steen from the States is a really articulate spokesperson about this. He calls it quantitative literacy or sometimes functional mathematics. I don't care what we call it but it's this thing about making sense of mathematics. He says that mathematics and numeracy quantitative literacy are quite different. Mathematics asks students to rise above the context whereas numeracy or quantitative literacy is anchored in real data that reflects engagement with life's diverse context and situations. And what this means is that numeracy is a lot harder to teach and assess than mathematics.

So I want to now go on my journey through assessment. I was involved in ALL and to me it was useful to have some data. There certainly are questions about what it's assessing but it's assessing something related to numeracy and maths. ALL is a benchmark and that's what it's designed to be, so you can compare change over a period of time, compare performance against other countries, have a look at how big your tail is. The figures are pretty dramatic in themselves, but if you actually look beneath the data you then find out some startling figures. For example, the difference in gender in numeracy is huge. It shouldn't be 10 percent so there's something to look at there.

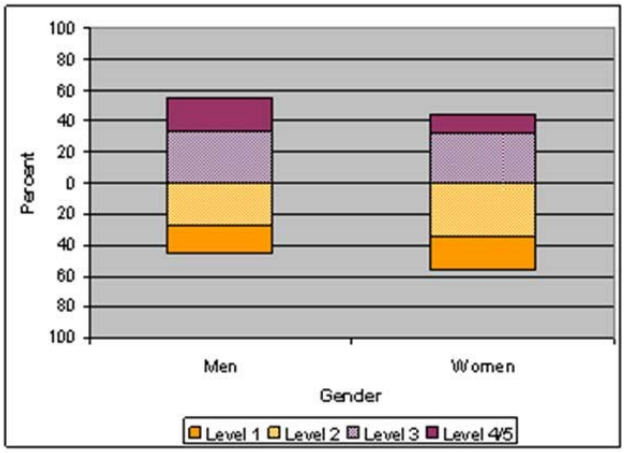
2000's: Benchmarks

Gender:

45% of males are at levels 1 or 2

55% of females are at levels 1 or 2

A difference of 10%!

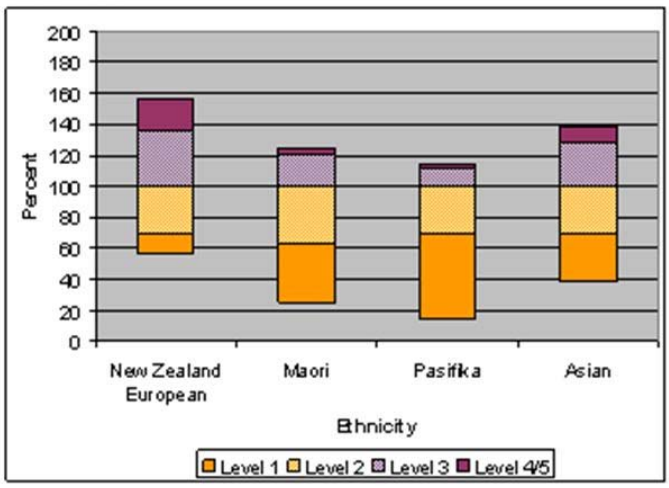


2000's: Benchmarks

ETHNICITY:

75% of Mäori are at levels 1 or 2

86% of Pasifika are at levels 1 or 2



Have a look at ethnicity. There are major issues in terms of who are disadvantaged in this thing that we're measuring. We're measuring something, it is telling us something and we need to do something about it. I'm passionate about that sort of benchmarking because I think it gives us messages, gives us ammunition and we wouldn't be here today if it wasn't for ALL.

Age is another interesting thing in ALL. Why do kids coming out of school not perform as well as people who are in their 20s and 30s and 40s? In Australia, your economic value by going up a literacy level is worth about $200 a week. That's a lot of money. If you have higher levels of literacy you're more likely to be employed, and there's a whole range of social outcomes that are related directly to your literacy and numeracy levels. We do know that investing in improving the literacy and numeracy skills at the bottom end has outcomes that governments and policy makers listen to. The other thing we need to think about and keep in mind is that in the 21st century our expectations in terms of literacy and numeracy are a lot higher than they used to be. We now need to read and fill in a lot more forms than we had to, and in workplaces we're expected to do the same. As well, you could have got a job 10 years ago, 15 years ago without any certification or qualifications. Now everybody has to have a certificate and be trained. Our expectations are much, much higher.

I will put my numeracy hat on again: I think while nobody goes around bragging they can't read and write, people are quite happy to put their hand up and say "I am no good at maths". I think there's a societal acceptance that it's okay to be not good at mathematics and I think that's sad. There's some research from the UK that now shows evidence that numeracy is a much more powerful indicator of social and economic disadvantage than is literacy, especially for women. We just looked at that 10 percent difference between how women perform compared to men on the survey and then you've got this research that says you can have good literacy skills, but with poor numeracy skills you can still end up being more disadvantaged. So I think there are some issues to hammer on the table about how important numeracy and maths are.

To me, a good teacher always assesses in a whole range of ways. Assessment is really important to good teaching. What that assessment is, and what tools we use, is the issue. What aspects of our assessment can we use the new New Zealand assessment tool for? It can't do all of our assessment, and I think we also need to keep separate the differences between teaching and assessment. They are different beings and we need to fit them in and make them work together. But some aspects such as the appropriate placement of learners, or in identifying where student errors might be, would be suitable for the sorts of examples that Libby was talking about, and the tool would really fit in to help with these tasks. So it's thinking about: Where do I use the tool? What information, what parts of those assessment aspects is it going to be really helpful for? It's not an either/or situation. In your classroom the assessment tool is only one part of your toolkit of assessment approaches and student feedback and how you work with your student.

I mentioned the Rethinking Assessment project earlier. In that project we came up with a diagram for what this group of experienced numeracy teachers thought were the important characteristics of assessing students' numeracy skills. The left-hand side is more the cognitive side. The skills and knowledge, the ability to transfer and apply skills, was seen as a really important aspect of students proving that they knew something about mathematics. The task process cycle was about getting into the problem and solving it and the skills and knowledge was another part of the jigsaw puzzle. Confidence was seen as really important—everybody said confidence was crucial. How do you assess confidence, though? And then the right-hand side is the more affective side—students being autonomous, their awareness of the mathematics in the world around them and making personal connections between some of their maths from their home life, their other lives, where else the mathematics was really important. That was the basis on which the group tried to work through the numeracy assessment framework. There was the task process cycle about identifying the problem, choosing your maths strategy, applying it and then reflecting on the outcomes. So, to me, when you're talking about assessing numeracy, it's not just about assessing mathematics. It's also about assessing the ability to get to the mathematics.

We need to have a look at the fact that the assessment tool gives you the ability to do both. There are both items in the assessment tool that are not in a context and items that are set in a context. Although we tried to reduce the reading demands as much as we could for the numeracy items, you can actually use the assessment tool to look at how your student was able to do both aspects—assessing the maths itself and assessing the ability to get to the mathematics. Is the literacy—the reading—for getting to the problem the major issue that the student is facing or is it the mathematics or is it both? So the tool has that ability behind it—that you can use it as part of your diagnostic information in your classroom.

I think one other issue in Australia—I'm not sure about here, but it's probably similar—is that a lot of adult numeracy teachers in Australia actually aren't trained or qualified in mathematics. I know when I started to teach adults I didn't know a lot about the beginning levels of mathematics. I knew about how to solve quadratic equations and differentiate, but I didn't know why you subtracted by borrowing and paying back or by decomposition. I had to go back and learn that. I didn't know why that, when you divided by a fraction, you turned the fraction upside down and multiplied. You just did it, it worked. So to teach numeracy you do need a lot of mathematical knowledge.

I think the assessment tool has diagnostic information in there that is very useful in that respect and the Progressions are very useful in that respect too.

3. Balancing choices and tradeoffs between different assessment purposes

Rosemary Hipkins

Dr Rose Hipkins is a Senior Researcher at NZCER.

Introduction

In traditional assessment practice, learning and assessment tended to be treated as sequential activities—you learn something and then you get assessed. Recent developments in education have resulted in the recognition of much closer relationships between the two types of activity and, as a consequence, assessment is now seen as integral to learning, not just a subsequent measure of its outcomes. I hope we would all endorse this shift in principle! However, in practice it creates a number of dilemmas that require smart solutions if promised improvements in outcomes for learners are to realise their potential. This presentation will explore the tensions and issues associated with making assessment integral to learning, concluding with some thoughts about possible solutions to the dilemmas identified.

A key issue is that assessment has different purposes and we often expect it to serve several of them simultaneously. This can make life difficult for tutors when it necessitates skilful juggling of competing priorities, both in initial task design and in the assessment judgements subsequently made. As the report on *Assessment of Foundation Learning* (AFL) identified (NZCER, 2006) the assessment choices that tutors must make have an additional poignancy when the learners being assessed are adults whose earlier experiences of assessment may have been damaging to their views and expectations of themselves as learners. As the personal talks given to introduce the conference showed, no matter how robust our own sense of ourselves as learners, we all react emotionally when we perceive something to be "not right" about the assessment we experience.

As a first step in our discussion of assessment of adult literacy and numeracy learning, we do need to be clear about the purposes we intend for the assessment activities we plan to carry out. There are different ways to categorise possible purposes and in this talk I plan to use those from the AFL report, with which you are likely to be familiar (and which I endorse and on which I will expand). Drawing on a framework developed by a Canadian education researcher whose own work has included assessment of adult learners (Aikenhead, 1997), three main purposes for education can be articulated as follows:

Systems accountability: This entails making judgements about the success and efficiency of learning, and is often linked to the idea of summative assessment, or assessment that reports on what has been achieved at the end of a learning cycle and/or progress made towards specified goals or benchmarks. This purpose foregrounds reliability of assessment. If comparisons are to be made between different learners, learning conditions, programmes of instruction etc. then the data gathered must be, and must be seen to be, fair to all involved stakeholders. Of course assessments must still be valid for their intended learning or they will lack any useful meaning, but this validity cannot be achieved at the expense of reliability. As we will explore, this is likely to entail some tricky balancing of assessment priorities and conditions.

Improved teaching and learning: This purpose is often linked to ideas of diagnostic and formative assessment. Assessment is carried out with the intention of determining current levels of achievement, and this information is then used to work out specific learning needs and potential next learning steps—often personalised to individual students' learning progress. This intention foregrounds the validity of the chosen assessment activity—it must have something meaningful to say to the tutor and the student about achievements and about current learning challenges. Reliability is likely to be more in the background here, unless of course the assessment is intended to serve dual accountability/assessment for learning purposes.

Lifelong learning: This purpose extends formative assessment purposes to "fully involve students and empower them to continue learning" (NZCER, 2006, p. 21). The assessment must not only be seen to be valid by the tutor, the learner must also understand what has been assessed and why, what the results mean and why and the nature of next learning steps and why these are now appropriate goals. That's a big ask! As recent policy advice to the Ministry of Education makes clear, it demands high levels of assessment literacy for everyone involved in the education system, from learners and teachers/tutors right up to programme managers, and those who maintain accountability systems on behalf of taxpayers (Absolum, Flockton et al., 2009).

The AFL report (NZCER, 2006) describes six assessment design principles and lays these out visually in a format that suggests that the Lifelong Learning purpose can subsume the Assessment for Learning purpose, which in turn, given careful assessment design, can subsume the Accountability purpose. Table 1 below shows the six principles, as they are laid out in the AFL report, and, in the interests of brevity, includes a selection of the advice provided for enacting them. (The full table in the report spreads over two pages.) As I have already indicated, meeting more than one purpose with the same assessment activity requires the assessor to successfully juggle both validity and reliability, and to apply large doses of "assessment literacy" to all the decisions made. Another tension highlighted by the selection of advice included in the table concerns the recommendation that assessment should allow for reporting against a continuum of progress, not just episodes of achievement (Principle 2) yet should also assesses specific (in context/authentic) competencies (Principle 4). The tension to be balanced here is between situated and more generic assessments.

Table 1 **Assessment principles and purposes (abbreviated from AFL Report, NZCER 2006, pp 27–28)**

	Systems accountability	Supporting teaching and learning–assessment for learning	Supporting lifelong learning

Assessment principles	Selected features of assessment (there are more in full report)
There are transparent assessment goals and clarity of purpose.	There is a shared understanding of the assessment process and criteria. The students' prior learning and current competencies are recognised Programme recording and reporting is consistent with the stated goals and purpose and, if appropriate, provides external recognition with a common currency.
Assessment aims to improve learning and pays attention to the needs and interests of the learner and to the process of learning.	The needs of the learner and an understanding of how students learn are central to practice Assessment is designed as a continuum of performance; it is ongoing rather than episodic. Assessment involves constructive, honest, feedback to the learners; enables reflection by students; and provides teachers with the information needed to plan the next learning steps. Assessment develops learners' capacity for self-assessment so that they become reflective and self-managing.
Assessment is valid, reliable, ethical, fair, and manageable.	Assessors use appropriate tools and sources of evidence that are fit for the purpose. The assessments provide teachers and learners with the evidence needed to make quality interpretations. There are systems to ensure that consistent judgements are made about performance.
Assessment is authentic.	Assessment places an emphasis on both the generic competencies of the area as well as specific competencies of the workplace, community, and everyday life situations.
The assessment is credible to all relevant stakeholders.	A systematic approach is taken to accountability and programme improvement based on meeting students' needs and programme goals. The teacher and provider participate in ongoing quality assurance processes, such as moderation systems. The assessment provides useful information to report credibly.
Assessment is undertaken by tutors with experience and assessment practice is supported by ongoing professional development.	Assessment results are used to improve programme design and teaching approaches. Teachers are experienced in the use and interpretation of the assessment tools and are able to use the information to promote learning. Teachers share their teaching and assessment practices and undertake regular professional development.

Revisiting commonplaces of assessment

Before we can explore ways of managing tensions and tradeoffs such as those between validity and reliability or between situated and more generic assessments, I want to return to the idea that different theoretical paradigms underlie the three purposes for assessment I have outlined. Accordingly, my next task is to lay out some ideas and beliefs that underpin assessment practices as we currently understand and enact these.

First we need to visit our beliefs about **what learning is**. These beliefs are often held tacitly yet they impact powerfully on what we think evidence of learning should look like, which in turn shapes the nature of the assessment tasks we offer, and subsequent judgements we make. The audience for the assessment information we generate is likely to share these commonly held tacit beliefs, with the consequence that current practices "make sense" to them, and this in turn reinforces their legitimacy. The "common currency" needed for accountability purposes (see Principle 1 in Table 1 above) reinforces the challenges we face in making our theoretical assumptions more explicit.

For all of us, one important yardstick of "robust" assessment tends to be those practices that offered us challenges in our learning pasts, especially if we were successful in these assessments, with all the positive consequences that followed. It is important to remember that the audience for assessment includes the learners themselves, especially when we are considering adult learners who may not have experienced the assessment successes we enjoyed in our own pasts. They, too, are likely to share our beliefs about what learning is, but I would argue that the present neat alignment between views of learning and common assessment practices has worked against their interests and has to change if we are genuine in our desire to help them make the most of the second chance they have been brave enough to take up. My argument points to the need to revisit our **attitudes to assessment** and specifically the work we expect it to do in the world.

Equally daunting is the associated challenge of revisiting our views of the learning that "counts". Here we need to explore our views—again, almost certainly tacitly held—of **what knowledge is** and what we expect it to do in the world. One possible purpose of assessment is to act as a predictor of the likely success of using current learning in new contexts—the so-called "competencies" agenda. It can also be used as a predictor of likely success in future learning—the "lifelong learning" agenda. Here our views of the **transfer of learning** must be brought into clear view, so that we can examine their adequacy in the face of future-focused learning challenges.

In order to put some practical flesh on the bones of these rather theoretical arguments, I'm now going to describe three different ways of going about making assessment judgements. Each of them addresses the assessment purposes outlined above in a quite different way, and with very different consequences. I have deliberately been somewhat provocative in the way I have shaped Tables 2 and 3 because I hope they will trigger some spirited discussion about those tacit assessment commonplaces that we must bring critically into view. To begin the process of comparing this type of assessment with other possibilities, I have shaped several deceptively simple questions:

- How is success defined? (The views of knowledge question)
- What does evidence of successful learning look like? (The outcome question)
- What does this evidence imply about what learning is? (The theory of learning question)
- What does successful assessment predict? (The competency/transfer question)
- What does the learner learn about their learning "self"? (The learning to learn question)

Traditional assessment practices

Let's begin with the assessment practices that sorted us out as successful learners in our own pasts. Many of you, like me, will have experienced the need to pass traditional written examinations, likely beginning with School Certificate. If we apply the above questions to traditional examinations, we might answer along the lines shown in Table 2:

Table 2 **The "commonplace" assumptions of traditional norm-referenced summative assessments**

How is success defined?	Success is determined relative to others. There have to be "winners and losers" to create a "normal" distribution.
What does evidence of successful learning look like?	Typically, evidence of learning comes from successful recitation of valued knowledge: the more of it, and the more difficult/abstract, the greater the success.
What does this evidence imply about what learning is?	Learning is acquisition of knowledge and ability to recall and use it appropriately on demand.
What does successful assessment predict?	Success predicts likely future success in similar assessments.
What does the learner learn about their learning "self"?	We learn to know our place!

The turn to "standards-based" assessment

Internationally, standards-based assessment has become more common in recent years, and is typically associated with the development of National Qualifications Frameworks (NQFs) that broadly specify learning outcomes, often shaped as some sort of "standards" or benchmarks to be achieved. Here in New Zealand we see this trend reflected in the development of the various National Certificates registered on our own NQF and awarded for a wide range of types of learning. These qualifications include the NCEA (National Certificate of Educational Achievement) and, perhaps more relevant for some second-chance learners, the NCES (National Certificate of Employment Studies), with their associated achievement and unit standards. Ponder a moment on how you think this method of assessment might change the ways we would answer the five questions from Table 2 (if indeed it does). Table 3 shows how I would respond to my own challenge.

Table 3 **The "commonplace" assumptions of standards-based assessments**

How is success defined?	Success is defined relative to a specified standard—the "bar" is set the same for everyone (but normative judgements are likely to have been implicated in the determination of where the bar should sit).
What does evidence of successful learning look like?	Success is defined as evidence of satisfactory performance of tasks that meet the specifications of the standard, as determined by the professional judgement of the assessor.
What does this evidence imply about what learning is?	The nature of learning is tacitly defined by a combination of the standard and the tasks seen as appropriate to make the necessary judgement. Evidence could still be gathered via traditional written test-type tasks (traditional views of learning), or it could require some practical demonstrations of skills (see below for a discussion of views of learning and knowledge that include "doing", not just "knowing that …").
What does successful assessment predict?	Success predicts the ability to do similar tasks in future—there is some potential here for competency development if tasks allow for that.
What does the learner learn about their learning "self"?	We learn whether our performance meets the standard (whether or not we understand why is another matter).

Assessment as inquiry

For my third example, I want to move beyond these familiar assessment processes and speculate on what might be if we could transcend current theoretical frameworks and begin a process of reimagining assessment.

Table 4 is based on Ginette Delandshere's idea that the most important assessment question is "What does it mean to know?" (Delandshere, 2002). This deceptively simple question can't be answered without active involvement of the learner in the assessment process because they are the person who has to make the meaning that allows them to use their knowledge in the world. Here the focus is on the learner and teacher working together to enact assessment as an investigation into the learning that has occurred and the potential worth of what has been achieved. Both parties have work to do if a student is to succeed in lifting their performance in the areas targeted, and they need to do this work together as willing and knowledgeable partners.

I hope that as you read the next table, you recognise resonances with the idea that assessment can impact positively on any person's view of themselves as a potential "lifelong" learner.

Table 4 **Reframing assessment as a process of collaborative inquiry**

How is success defined?	Success resides in meeting clearly specified personal learning goals, which have been co- constructed with the tutor or another learning mentor.
What does evidence of successful learning look like?	Evidence should clearly inform the learner about their progress towards meeting their goals (with the proviso that we do actually know what progress looks like across a range of learners and contexts).
What does this evidence imply about what learning is?	This ideal fits more comfortably with sociocultural theories of learning (see discussion to follow).
What does successful assessment predict?	Successful learners display an enhanced ability to act in the world. The skills and dispositions they demonstrate also predict likely success in ongoing learning—that is, the ability to transfer that learning to new contexts.
What does the learner learn about their learning "self"?	Learners know from direct experience that they can be successful learners and meet their goals—they are aware of the strengths and attributes they can build on, as well as areas that may need strengthening.

Relationships between theories of learning and knowledge and assessment approaches

What theoretical ideas lie beneath our tacit thinking about the commonplaces of assessment? What might we need to consciously rethink as we juggle assessment purposes while shifting our practice towards the widely-promoted goal of fostering lifelong learning? At this point in my talk I want to delve just a little into different ideas about the nature of learning and of knowledge.

Theories of learning

Much has been written elsewhere about the limited range of learning theories that are compatible with traditional assessment practices (see, for example, Delandshere, 2002). For example, behaviourist learning theory assumes that what is taught can be learnt exactly as intended, as in rote learning. Here the targets for traditional assessment are easy to identify and it is possible to assume that every student in the class has the same opportunities to learn, given exposure to the same experiences (Moss, Pullin et al., 2008). Cognitive learning theories render such assumptions somewhat problematic because they highlight the individual and active nature of meaning making (as in some forms of constructivism). However, these theories also tend to work as if learning resides primarily in the brain of the individual and is essentially a mental activity. While assessment might seek varying degrees of "understanding" the focus is still essentially on the retrieval and re-presenting of declarative knowledge, which fits comfortably with traditional assessment practices.

I am not saying here that these learning theories are "untrue". All of us have found rote learning useful in some circumstances. And I'm sure we are all aware of times when we have understood something differently from others, and perhaps found ourselves doing some uncomfortable rethinking. My reason for raising these tacit theories is to highlight their inadequacy for the assessment challenges, and indeed life circumstances, that now confront us in the 21st century. They simply do not account for all that is now known about how humans learn, or indeed need to learn to learn in order to thrive in the complex societal conditions in which we are all immersed.

One limitation to behaviourist and cognitive theories of learning is that they assume a brain/body binary that is incompatible with insights from modern biology about how the brain actually works (Capra, 2002 provides a clear explication of this point). I'm sure we can all think of times when our learning has been thoroughly embodied and how hard it can be to unlearn/relearn in such circumstances. For example, a few years ago I needed to readjust my walking gait after knee surgery. Even with the help of a skilled physiotherapist this was a frustrating and lengthy learning challenge. A related limitation is that behaviourist and cognitive theories of learning do not take account of the complex emotional ecologies of learning (see, for example, Zembylas, 2007) where our individual feelings and prior experiences impact powerfully on what and how we learn. Yet another limitation—again, one with which I am sure you are all quite familiar—is that behaviourist and cognitive theories of learning do not take account of the powerful role that contexts can play in learning. So let's briefly turn our attention to a broad family of learning theories that do address these challenges.

Sociocultural learning theories do not assume a brain/body binary, nor do they assume that individual learning can be separated from the contexts in which it takes place. I have found the following four broad characteristics of sociocultural learning theories particularly useful for rethinking assessment practices and challenges (see, for example, Hipkins, 2009):

Learning is situated: It happens in contexts that include other people, artefacts of various sorts and a history that will entail particular ways of being and working (how we do things here), as well as some sort of agenda for the activity taking place (what this learning is really about).

Learning is distributed: The action is stretched over the various resources of the situation and hence entails multiple interactions between the learner and these other dimensions.

Learning is mediated: The various resources and cultural tools available, including other people, and the routines and social practices present, need to be deployed in ways that support each individual's opportunities to learn.

Learning is participatory: Learners engage in activities that have compelling meaning for them and their communities, and that have the potential to open up into bigger picture frames and concepts and to help them construct positive futures for themselves and others.

It's interesting to think about the possibilities of standards-based assessment in the light of these different learning theories. All of us will be aware of assessment tasks, developed for one unit or achievement standard or another, that enact SBA very much as "business as usual", maintaining the traditional reliance on written tests or examinations, workbook completion or the like. Changing the method of assessment is demonstrably insufficient to challenge views of what matters in learning! However, SBA certainly has the potential to positively account for the situated, distributed, contextualised and participatory nature of learning when it opens up a wider range of assessment activities, as has happened in some cases. Why hasn't this potential been more widely embraced, given what we now know about learning and the obvious potential benefits for learners? I suggest that we've got another powerful group of tacit theories to unsettle before innovative methods of SBA, let alone ideas of assessment as inquiry, will be more widely embraced.

Theories of knowledge

As we've already noted several times, traditional assessments typically require learners to write about things but not actually do them. The tacit theory at work here is that knowledge is a "thing" that you can get and give. Recent theories position knowledge in more participatory ways—it is what we do that

demonstrates what we know—knowledge in action, not just knowledge as saying so. Without digressing into the long history of the traditional view of knowledge (see Gilbert, 2005 for a discussion of this) I want to draw attention to the deep academic/vocational divide that rests on, and is in turn sustained by, this valuing of recited, decontextualised, abstracted knowledge. Since it is possible to trace this idea right back to the thinking of the ancient Greeks it's not surprising that it is so deeply embedded in our culture as to be all but invisible.

I'm sure we can all think of a range of challenges associated with the academic/vocational divide, not least of them the systematic undervaluing of learners whose aptitudes lie in practical tasks and manual dexterity. The foolishness and waste of this should be evident on several scores. We rely on people with practical know-how to keep things working in the world. We certainly do value specific practical knowledge when we don't have it ourselves and need access to it. We should also note that practical knowing is highly complex and getting more so as technologies evolve. As with the brain/body binary I mentioned earlier, we need to unsettle the deeply entrenched belief that academic and practical knowing can be easily differentiated. Complexity thinking and sociocultural learning theories would suggest that they are thoroughly enmeshed in each other.

Many of the adult students whose learning challenges we are exploring today are likely to fit the description of "vocational" learners, so one challenge in working with them lies in undoing damage to their beliefs in their own abilities to be successful as learners. If what they have learnt about themselves from their past experiences of traditional assessments is to "know their place" in the world, we need to help them see a different set of possibilities. But in order to believe they can learn, and hence to want to keep on learning, adult learners also need opportunities to rethink their ideas about the nature of learning and of knowledge. Changing assessment practices can certainly help here, but changing what we do with assessment results—how we talk about and use them—is equally important. Since our actions and interactions are underpinned by our own tacit beliefs, the old adage, "first change ourselves", seems pertinent here.

Theories of competency/transfer of learning

To round out this section I need to briefly deal with another set of ideas that lie behind assessment practices. Claims about the potential benefits of learning typically relate to what people ought to be able to do with what they know at some future time—the so-called transfer of learning. Traditional teaching and assessment methods tend to take transfer for granted; it has been assumed rather than specifically addressed (Engle, 2006). Sociocultural learning theory doesn't necessarily help us here because some researchers use the situated nature of learning to question whether it is actually possible to assess transfer at all, since specific contexts have such an impact on what can be learnt, and indeed whether or not previous learning can be displayed (for a discussion of this, see Taylor, Ayala et al., 2009).

In the light of the recent move to a focus on learning for competency development we cannot ignore these challenges. The OECD's key competencies, which were intended to underpin learning at all levels of the New Zealand education system including tertiary learning (see, for example, Ministry of Education, 2005) emphasise being able to adapt knowledge and skills for use in new or more challenging contexts, with an associated focus on the dispositional challenges associated with demonstrating knowledge-as-action. I would argue that the assumption of transfer is integral to such competency definitions and this has huge assessment implications. How would we know transfer has occurred? What would the evidence look like and what assessment methods would help us to seek it? Here are some of the challenges I can foresee:

If knowing is seen as residing in action, and if competency is seen to require learners to demonstrate ways they adapt what they know and can do to appropriate use in different contexts, can we rely on evidence from "one shot" assessment events? If not, how might we determine sufficiency of evidence and what processes should we adopt to co-ordinate evidence from different episodes to make overall professional judgements?

Can we assume that one student's transfer challenges and achievements are the same as any others? (We certainly act as if we think this is the case when we assess them all with the same task, judged in the same way and reported to the same metric.) If not, how do we keep track of and link assessment episodes so that learning trajectories for different individuals, in all their potential idiosyncratic variety, can be documented in manageable ways that make their personal progress clear? (How do we even know what progress could/should look like in order to make this comparative judgement?)

How do we allow for knowledge-in-action that emerges in highly engaged group learning? NCEA allows for this in some performance contexts such as ensemble making of music but this is seen as an exception rather than something we should be aiming for in a wider range of learning areas. What about adult learners who are expected to use their new skills in the workplace? Where do we draw the line about provision of assistance? How might we determine what different individuals have achieved in co-working situations?

Let me point to a glimmer of light in the complex dilemmas I've just posed. Who *potentially* knows best if a supposed piece of new learning really is a stretch for them or not? Who *potentially* knows best if something they have just done really was a new use of existing knowledge and skills (i.e., an instance of competency development) or didn't really extend what they could previously do? Who *potentially* knows best if they really have made an effective contribution to a group action, or just reaped the benefits of others' hard work? I've highlighted the word "potentially" here to signal that the critical self-knowledge needed to assess one's personal learning gains certainly cannot be assumed. But this sort of critical self-assessment literacy certainly can be modelled and taught, given that tutors have the necessary knowledge, skills, attitudes, values and access to appropriately supportive assessment tools and strategies that support their assessment work. The ball is in the tutor's court first and foremost.

As I've just outlined, the active involvement of students in their own assessment has the potential to solve practical problems and make assessment processes more manageable for assessors. However, it is likely that the greatest benefits reside in what learners learn about themselves and their learning. In turn, this idea aligns very nicely with more recent theoretical thinking about the nature of transfer. For example, Ferrence Marton (2006) highlights the key role played by recognition of both similarity and difference as an enabler of transfer. In his view learning to "see" similarities and differences is a matter of increasingly refined perception. Conceived of this way, teaching for transfer is literally a matter of helping students to "see" their learning differently and having done so to also think about the action possibilities bound up in their new insights and skills. It is about learning to learn, or preparation for future learning (Bransford & Schwartz, 1999) as much as about learning something specific.

Armed with this broad sweep of theory, it's time to return to Aikenhead's idea that different purposes for assessment align to different theoretical paradigms. Table 5 is the summary of his framework, as published in the AFL report.

Table 5 **Three purposes of assessment (from AFL report, NZCER, 2006, p. 21)**

Assessment purpose	Most compatible paradigm	Nature of assessment
Systems accountability and reporting	Empirical–analytic	Empirical methods based on psychometric principles, yield "robust" comparative data
Improving teaching and learning	Interpretive	Evidence of achievement against specified standards, may combine descriptive and data-based components. Judgements made by others, not students
Lifelong learning	Critical–theoretic	Extends features of interpretive paradigm—collaborative methods fully involve students and empower them to continue learning

I hope you can see the resonances between the idea of transfer as preparation for future learning and the "lifelong learning" assessment purpose. However, if this purpose is to be achieved and not simply paid "feel good" lip service, we really do need to make substantive changes in how we think about learning, knowledge and competency/transfer, with associated shifts in how we think about and enact assessments. I don't for one moment think there will be one right way to address these challenges. I can see great potential in developments such as the use of e-portfolios and other such new electronic recording tools. However, the challenge we have set ourselves today is to take account of the theoretical shifts outlined yet still come up with manageable methods of assessment that are able to help juggle the competing priorities of the three different assessment purposes. Is this an impossible ask? Let me explore one possibility to round off my talk.

"Smart Tools", accountability and lifelong learning

One dilemma that assessors face is how to make valid yet reliable determinations of progress such that differences between individuals can be allowed for and acknowledged. We can't ignore the need for a measure that clearly describes progress when the accountability purpose for assessment requires empirical reports of that progress. Yet we also know that individual learners will start at very different points and make progress at different rates and in different ways. Can we juggle competing priorities so that one assessment process can provide the reliable but essentially generic data that managers and regulators need, yet also provide useful and more specific information that individual learners can validly use when reflecting on what their learning means for them in terms of realised achievements and next learning goals (i.e., assessment as inquiry)? What is needed is a measure of learning that is generalisable, so that results can be aggregated and reported, yet also potentially specific to any one individual. Are these demands incommensurate or can this mix be achieved?

Thinking about this question, I can see a valuable place for Smart Tools (computer-adapted assessment tools) such as the new Literacy and Numeracy tool. The scale developed for each tool allows for generic collation of assessment results, with confidence in their reliability because individuals are assessed in the same way and assigned a "place" on the scale. Assuming that the assessment items are valid, this potentially satisfies accountability requirements, particularly when shifts over time can be demonstrated for whole learning cohorts. However … so far, so traditional! Each scale is based on a typical order in which learning unfolds (as are the Learning Progressions to which the tools are aligned). It also assumes that

individuals will respond to the tasks in the manner intended by the assessment designers. Nevertheless there are several interesting points of departure from traditional assessments that I find very interesting and potentially useful for thinking about the lifelong learning challenge.

Let's begin with the tricky emotive challenge. As I noted earlier, normative judgements can discourage ongoing learning when students find themselves at the bottom of the learning heap. They learn to know their place! I'm not suggesting we should fudge feedback about relative progress—indeed it is an essential cornerstone of lifelong learning that we embrace and address personal learning challenges, and robust feedback is obviously helpful when used in this positive, targeted, personalised way. It is the negative comparison with other learners that is so discouraging in traditional normative reporting. Smart Tools help address this issue, in part by adjusting the mix of questions to which the learner responds. Because responses are instantly assessed, learners who are getting most items correct (i.e., the test is too easy) will have more challenging questions selected to follow up. Conversely, if most early responses are incorrect, the program adjusts by selecting some easier items. Of course good tutors make these sorts of judgements and adjustments to their own assessment tasks all the time. The difference here is that selections are made from a generic bank (reliability is addressed, not just validity) and they are made in the moment in a manner that is invisible to the learner as they experience the assessment activity. That's clever!

I hope you are at least broadly aware that the common scale for each tool has been statistically derived using a psychometric process called Rasch modelling. This process is used to analyse many students' patterns of answers to a range of questions. I know Charles and Juliette will have more to say about this in their talk but I want to draw attention to the implications for making reliable professional judgements. A process such as this can do something we humans cannot do unaided. No matter how much we might wish to do so, our attention will inevitably be drawn by interesting details in answers to individual items and it is very easy for these to carry an unequal weight in our minds that distorts our overall judgement of achievement. By contrast, the process used to convert the raw test score to a "scale score" (i.e., an overall judgement against the common scale) is based on empirically derived measures of how difficult each item is for adult learners in New Zealand. Again, that's smart!

One trade-off for this smart computing is that the items need to be broadly generic because the tests will be taken by a wide range of learners who work in very different contexts. Good tutors rightly pride themselves on their skills of contextualising assessment tasks to motivate and encourage second-chance learners and it's really important to keep those informal assessments in play. Smart Tools complement rather than replace existing tutor assessment expertise, but they do also make new demands. I hope it will be obvious from what I've already said that I see these new demands as residing in the ways tutors and those who work with them learn to use the new tools to support assessment as a powerful and empowering inquiry process. In the spirit of inquiry, we should be encouraging learners to explore the meaning of the scale scores they achieve, both in terms of successful progress, and next learning challenges. No tools could ever replace tutors' expertise in helping students unpack what their assessment results mean—but well-designed tools can certainly add to such learning-to-learn conversations. The challenge is for all of us to work together to achieve the potential of smart assessment to turn around the chances of success for learners who must unlearn their old place in the world in order to be able to fully engage the more participatory ethos of ongoing and lifelong learning in this century.

References

Absolum, M., Flockton, L. et al. (2009). *Directions for Assessment in New Zealand (DANZ).* Wellington: Ministry of Education.

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

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

Capra, F. (2002). *The hidden connections: Integrating the biological, cognitive, and social dimensions of life into a science of sustainability*. New York, London, Toronto, Sydney, Auckland: Doubleday.

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

Engle, R. (2006). Framing interactions to foster generative learning: A situative explanation of transfer in a community of learners classroom. *Journal of the Learning Sciences*, *15*(4), 451–498.

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

Hipkins, R. (2009). Determining meaning for key competencies via assessment practices. *Assessment Matters*, *1*, 4–19.

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

Ministry of Education. (2005). *Key competencies in tertiary education*. Wellington: Author.

Moss, P., Pullin, D. et al. (2008). *Assessment, equity and opportunity to learn.* Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo, Delhi: Cambridge University Press.

NZCER. (2006). *Assessment for foundation learning: The importance of purposeful assessment in adult literacy, numeracy and language courses*. Wellington: Ministry of Education.

Taylor, M., Ayala, G. et al. (2009). Understanding learning transfer in employment preparation programmes for adults with low skills. *Journal of Vocational Education and Training*, *61*(1), 1–13.

Zembylas, M. (2007). Emotional ecology: The intersection of emotional knowledge and pedagogical content knowledge in teaching. *Teaching and Teacher Education*, *23*, 355–367.

4. Inside the assessment tool (Part 1)

Juliette Mendelovits

Juliette Mendelovits is Lead Principal Researcher, Humanities and Literacy, at the Australian Council for Educational Research

Introduction

The Literacy and Numeracy for Adults Assessment Tool has several purposes: to provide good evidence to learners, educators and other stakeholders about the level of proficiency attained by learners at points in time; to show where the gaps in a learner's or group of learners' knowledge and skills are that might require intervention; to indicate to stakeholders whether individual and group proficiencies are improving over time. But the overarching purpose—the one that guided the development of the 2,650-odd items that were generated for the assessment tool in numeracy, reading and writing—is to improve the competency of adults in these essential skills through providing sound evidence of what they know and can do. Good assessments are very carefully constructed, so that each item contributes a piece of this evidence, to tell part of the story about learners' proficiency. In this paper, some of the thinking and processes that have gone into developing the items for the assessment tool are described.

Starting point: frameworks and blueprints for the assessment

An assessment framework guides the work of test development and provides an explanation for users and other stakeholders of what the assessment is about. A framework also ensures that the test material is developed systematically so that it will yield results that can be interpreted and generalised in a coherent and intelligible way. For the Literacy and Numeracy for Adults Assessment Tool, the primary specification was to address the adult learning progressions for literacy and numeracy that were published by the Tertiary Education Commission in 2008, so the progressions formed the basis of framework development.

Only some of the progressions were selected for the first phase of the assessment tool. Those selected were considered to be both among the most essential, and the most amenable to assessment through an online system. Table 1 shows the strands and progressions targeted for the first phase of the assessment tool's development.

Table 1 **Strands and progressions included in the first phase of assessment tool development**

Strands	Progressions
Read with Understanding	Vocabulary
	Language and text features
	Comprehension
	Reading critically
Write to Communicate	Purpose and audience
	Spelling
	Language and text features
	Planning and composing
Make Sense of Number to Solve Problems	Number sequence
	Place value
	Number facts
	Additive strategies
	Multiplicative strategies
	Proportional reasoning
Measure and Interpret Space and Shape	Measurement

Every task in the assessment tool is designed to reflect one of the steps of one of these progressions. In the case of the literacy strands, the selected progressions for reading and writing were the focus, respectively, for single Reading and Writing assessments. For numeracy, all seven progressions were included in a general Numeracy assessment. In addition, two subsets of progressions were identified to underpin assessments of Number knowledge and Number strategies. The assessment of Number knowledge combines Number sequence, Place value and Number facts; and the assessment of Number strategies and Measurement combines Additive strategies, Multiplicative strategies, Proportional reasoning and Measurement.

In addition to the progressions and steps, the frameworks for the assessment tool set out other important variables for shaping the instruments. In the Reading assessment, for example, a key consideration was including a range of texts in different formats, in recognition of the fact that adults read not only continuous text or prose (for example, in newspapers, magazines and books) but also noncontinuous texts such as tables, graphs, maps and forms. Therefore, each reading item is categorised as either Continuous or Noncontinuous.

Another important feature included in the assessment tool framework was contexts: the situations in which adults typically draw on literacy and numeracy skills and knowledge. Adults need to read, write and work numerically in the workplace, in community activities, at home and in further education and training. For all of the assessments (Reading, Numeracy and Writing), a set of context categories was adopted, with minor modifications, from the international Adult Literacy and Life Skills (ALL) Survey, in which New Zealand had participated in 2006. The contexts used in the assessment tool are Work, Home and family, Health and safety, Consumer economics, Community and citizenship, Leisure and recreation, and Education and training. In Numeracy, it was decided that it was important also to include a "context-free" category of tasks. Contextualised tasks are set in real-world contexts, where part of the challenge might be

for the learner to figure out what the numerical problem is that needs to be solved. Context-free tasks, on the other hand, assess whether learners can perform operations and calculations in a "pure" form.

The frameworks set out the rationale of the assessment and the variables that are to be covered. In order to define the relative importance of the various categories identified for each framework variable, a test blueprint was devised. Table 2 shows an extract from the Numeracy blueprint, with the targeted proportions for each category of the context variable.

Table 2 **Blueprint for Numeracy context categories**

Contexts	Percentage of items, numeracy
Work	33
Home and family	9
Health and safety	6
Consumer economics	9
Community and citizenship	6
Leisure and recreation	7
Education and training (and context free)	31
Total	100

It can be seen that a large proportion of items in the Numeracy assessment is in the Work context, as this was judged to be the most relevant context for the adult learners who would be using the assessment tool. In the Numeracy blueprint, the context-free items referred to above are included in the Education and training category, since "pure" numeracy items are likely to be encountered in an educational setting.

Making good items: science and art

The science of test development is that every item contributes a piece of evidence that will ultimately contribute to describing (quantitatively and qualitatively) what a learner knows and can do. It is thus essential that items are developed deliberately so that an inference about proficiency in the learning area can legitimately be drawn. This is done scientifically by ensuring that each item addresses specified aspects of the framework, and by ensuring that the set of items as a whole conforms with the blueprint. There is also, however, an art to test development. Finding an interesting and appropriate piece of stimulus as test material, and framing and honing an item to maximise its validity, are matters of craft, experience and judgement as well as technical skill.

In fact, science and art are inextricably in play in test development. In the case of the assessment tool, the test developers began with an orientation to the progressions and the development of the framework and blueprint, and then looked for suitable stimulus on which to base the items. Some of the questions in the test developers' minds when looking for stimulus were:

- Does the material relate to the test specifications (for example, the defined contexts and, for reading, the text formats)?
- Is it at the right level, in terms of the expected ability, knowledge, skills and understanding of the learners?
- Is the material interesting? Is it worthwhile? Is it of some importance?

- Is it coherent? Unambiguous? Clear?
- Is it self-contained? Or does it assume other prior knowledge and, if so, is this appropriate and fair?
- How would this material stand up to public scrutiny (including that of project stakeholders and the wider community)?
- Does the material breach any ethical, cultural or other sensitivities?
- Is it likely to result in bias: that is, is it likely to be easier or harder for certain subgroups in the test population for reasons other than differences in the ability being measured?
- Is the reading load as low as possible? (unless it is to be used in a reading assessment!)
- How will New Zealand adult learners perceive this material?

While some of these questions are technical, many of them require the test developers to exercise imagination and empathy, to apprehend the cognitive and cultural demands of the items from the point of view of learners. The stimulus may be found—typically, in reading, it is an authentic text. Alternatively, it may be constructed by the test developer: for numeracy and writing, more often than not the test developer conceives of a scenario that will support tasks addressing a particular set of skills and knowledge, and then collects images and words to create the scenario.

Having selected or constructed stimulus, the test developer then embarks upon the initial creation of the items themselves. The original drafting of an item must address the stimulus in a meaningful way that does not trivialise or distract from the context. The choice of item format is also critical. Since instant reporting is built into the assessment tool for the online Numeracy and Reading assessments, all items in the adaptive assessments are scored automatically, precluding the kinds of constructed response items that need the judgement of a professional marker. (A small number of constructed-response Numeracy items that require hand marking are stored in the assessment tool for use in nonadaptive printed tests.) The available item formats for the adaptive assessments in the assessment tool are multiple choice, highlighting (for example, selecting a phrase from a piece of prose in the Reading assessment), clicking to select (for example, clicking on an icon on a map) and closed constructed response (for example, entering a number in a text box for a Numeracy item). The Writing assessment, unlike the Reading and Numeracy assessments, was in the form of extended written responses that would need to be hand marked by educators. With the available technology, and in order to present an assessment that was authentically matched to the learning strand, an adaptive or instantly scored Writing assessment was not judged feasible for the first phase of the assessment tool. For each strand, therefore, the range of item formats was selected to best suit the demands of the strand, within the constraints of the available technology.

Part of the task of developing an item is drafting the question that the learner will see. Another part of the task involves developing the scoring rules for the item. For multiple choice items, this requires simply identifying which of the (usually) four alternatives is to be given credit. For a closed constructed response or click item, the correct response (or small set of responses) must be defined. For other item formats, such as highlighting, the scoring can be a little more complicated: the marking scheme specifies the minimum and maximum parts of the text that must be highlighted for credit. Most complex are the marking schemes for the Writing assessments, where the criteria upon which the written response will be judged must be specified in the form of a rating scale, with a score and description for each category of response within each criterion. An example of the scoring scheme (also called marking guide) for one criterion in the assessment tool's Writing assessment is shown in Figure 1.

Figure 1 Extract from the Writing assessment's marking guide

1. Purpose and Audience

Skill focus: the writer's capacity to orient and engage the reader through provision of relevant information and consistent and appropriate register.

Note - consider register, language choices, information and communicative effect.

- Register: is the address of the reader appropriate to the task? Is there a written or oral 'feel'? Which is appropriate?
- Language choices: are there formal or informal language choices (sentence structures, vocabulary)? Which are appropriate?
- Information: is there enough? Is it relevant to the task?
- Communicative effect: are task goals achieved?

Category score	Category descriptor	Additional information
0	writer is unaware of the need to orient the reader	• no goals of the task are met
1	shows basic awareness of audience needs and expectations • attempts to orient the reader by providing some relevant information	• register inconsistent or inappropriate for task goals • gaps in information
2	reader is oriented to task • supports reader understanding • has an appropriate tone	• register generally consistent and appropriate for task goals • adequate and relevant information provided
3	controls writer/reader relationship • writing influences reader • takes readers' values and expectations into account • information is clear and easily understood	• appropriate register maintained • relevant information supplied • reader's needs catered for (clarity of expression) • task goals achieved

When assessing a response to a writing task, the educator selects the category description that best matches the learner's response; this is done on each of several criteria: for example, Purpose and audience (as shown in Figure 1), Spelling, Punctuation of sentences, Quality of ideas and Text Cohesion.

Alongside the item as seen by the learner and the marking guide, original test development involves defining the item in terms of framework characteristics: most importantly for the assessment tool, the progression and step. There was often a challenge in achieving this since the progressions are designed as descriptions of learning rather than as assessment frameworks. Some of the indicators of learning that are described are difficult to operationalise in an online assessment. For example, both the Reading and Numeracy Progressions often refer to "using strategies". In a learning environment the educator can observe the kinds of strategies a learner uses; in a formal assessment context, however, more often the strategies have to be inferred. So, for example, in designing an item to reflect a step description such as "use comprehension strategies to assist in understanding information or ideas in longer or more complex texts" (Read with understanding, Comprehension progression, Step 3) the item in the assessment tool will necessarily focus on collecting evidence of understanding of information, from which the possession of the appropriate comprehension strategies must be inferred.

As well as defining progression and step, the test developer defined the relevant categories of context, text format and text type (for Reading). A number of other pieces of metadata were also generated at this initial stage of development. For Reading, Writing and Numeracy, the test developer also wrote a "question

intent" for each item, describing in specific terms what the item was intending to assess. This piece of information would appear in the report on learner performance as part of the diagnostic output.

Review and refinement of items

A crucial stage in test development for the assessment tool was the "panel", sometimes called "cognitive walkthrough" or "shredding". At this stage, a group of three or four professional test developers met to discuss test material, viewing from every possible angle both the stimulus and items. The test development team tried to anticipate every kind of criticism: conceptual, contextual, linguistic. They looked at the layout of the stimulus, the formatting and fonts, the choice of any illustrations. They also reviewed the metadata categories and the questions' intent. Some of the questions that were regularly asked in panels for the assessment tool were:

- Does the item reflect the intended learning progression?
- Does the item relate to the essence of the stimulus or does it focus on trivial side issues?
- Is it clear what would constitute a good response to the item? That is, will learners understand exactly what they are being asked to produce (whether or not they can produce it)?
- Are there any "tricks" in the item that should be removed?
- When there is more than one item attached to a piece of stimulus, are there dependencies between the items? Does one item give a clue to the next one? Would a different order of items within a unit make a difference? If a response to one item is incorrect, does this affect possible responses for other items in the unit? (To gain maximum information from each item, it should provide independent evidence about the learner's proficiency.)
- Is the key (the correct answer to a multiple choice question) indisputably correct?
- Are the distracters (the incorrect options to a multiple choice question) plausible but indisputably incorrect?

Typically, it is rare that an item emerges unaltered from the panelling process. Taking on board the critique of peers, after the panel the originating test developer discarded the item if it had not passed scrutiny, or refined it, ready for further reviewing.

As well as internal review, there were two external filters for the items. First, a percentage of the Numeracy and Reading items was submitted for checking to consultant experts in numeracy and literacy. Second, a reference panel of Māori and Pasifika educators and researchers was asked to review items, especially those that were considered potentially sensitive. These two quality checks were valuable means of ensuring the appropriateness of the material for the intended users.

Out in the field: trial testing the items

The process of test development up to this point was a matter of craft and judgement, underpinned by technical experience, on the part of test developers and expert consultants. A crucial next step in quality test development is a field trial, the main purpose of which is to collect data about each item from large samples of test takers similar to the target group, so that the validity and reliability of the items can be inspected statistically.

Items that had survived the initial processes of internal and external review were desktop published and assembled into printed booklets. Even though the assessment tool is an online system, the trial test was administered as a paper-and-pen assessment because the delivery platform's development was concurrent

with the test development. Accordingly, sets of items in test booklets were administered to learners in New Zealand polytechnics, industry training organisations and workplace literacy and numeracy programmes. The trial testing took place from November 2008 to September 2009, with the co-operation of institutions large and small throughout New Zealand. In all, about 20,000 trial tests were distributed across New Zealand, comprising about 1,450 Reading items, 1,200 Numeracy items and 20 Writing tasks. Up to 600 learners were administered each item. The trial tests were carefully designed so that sets of items appeared in more than one booklet, forming links across the whole set of items in each domain so that all could be calibrated on a single scale for each of Reading, Numeracy and Writing.[1]

As part of the trial test, learners and educators were asked to give feedback on the items with regard to their interest, difficulty and cultural appropriateness. This information was fed into the selection process. However, the main thrust of the trial test was to collect quantitative results from the responses to the items themselves.

Analysing the statistics from the trial test

Once the booklets had been returned to NZCER, teams of markers scored those items that required judgement (some Reading and Numeracy items, and all of the responses to the Writing prompts) and then those scored responses and the multiple choice responses were entered into a database and analysed using various statistical programs.

Several kinds of output were generated to allow analysis of whether each of the items was performing well enough to be included in the final set for the assessment tool, and whether the set of items for each strand provided the required coverage, as defined in frameworks and blueprints. Some of the questions to which the statistical analysis provided answers, at the item level, are listed below. The term in square brackets at the end of each question indicates the statistic that provides this information.

The first set of questions pertains to the extent to which the item contributes in a consistent, nonrandom way to measuring proficiency in the strand of interest:

- Does the item appear to be measuring the same latent trait (for example, number knowledge) as the other items in the pool? [fit statistic]
- Does the item help to discriminate between more and less proficient learners—that is, is there evidence that getting credit for the answer is not a random event? [discrimination and point biserial correlation]
- Is the mean ability of the learners who got credit for this item higher, across the whole assessment, than the mean ability of learners who did not get credit? [mean ability]

Other pieces of information provided by the item analysis relate to the difficulty of the items in relation to the proficiency of the learners. In an assessment of this kind, the guiding principle is not so much to ascertain whether learners have mastered a prescribed part of a curriculum or learning programme, but to estimate what point along a continuum of development they have reached, and what consequently might be done to help them redress any gaps in their learning, and to progress to the next level along the continuum. The full set of items in each assessment pool needed, therefore, to cover a range of difficulty, to match the range of proficiency of the learners. Such a range would provide even the most proficient learners with some challenge, and even the least proficient with some success. The following questions related to item difficulty can be answered using the statistical output:

[1] Charles Darr's paper discusses this part of the process more fully.

- How difficult is this item compared with all the other items in this assessment? [item threshold and item delta]
- How many learners got credit for the item?
- What percentage of the learners who were administered this item gained credit for it? [percentage correct]
- What percentage of learners gave an incorrect answer, and what percentage omitted to answer altogether? [percentage incorrect, and percentage of missing responses]

Figure 2 shows the trial test analysis for a single multiple choice reading item, with an explanation of some of the key features below.

Figure 2 **Item analysis for one item from the assessment tool trial test for Reading**

```
================================================================================

Item 24
-------
item:24 (R000014)
Cases for this item     230   Item-Rest Cor.  0.42
Item Threshold(s):    -0.68   Weighted MNSQ   1.03
Item Delta(s):        -0.68
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   WLEAvg:1 WLE SD:1
--------------------------------------------------------------------------------
   1       0.00        30      13.04   -0.15    -2.24(.026) -0.64     1.04
   2       1.00       145      63.04    0.42     6.93(.000)  0.51     1.23
   3       0.00        36      15.65   -0.15    -2.37(.019) -0.64     0.78
   4       0.00        19       8.26   -0.35    -5.60(.000) -1.49     1.00
================================================================================
```

Item: 24 (R000014)

This is the 24th item in the set of Reading items analysed in this output. R000014 is the unique code for this item in the assessment tool.

Cases for this item **230**

This item was administered to 230 learners during the trial test period; thus the analysis is based on a statistically reliable sample.

Item Threshold(s) **-0.60**

This statistic indicates the difficulty of the item compared with all the other items in the entire trial test for Reading (1,468 items). The mean difficulty of the items was set at 0, so it can be seen that this item, with a difficulty threshold of -0.60, was easier than average.

Item-Rest Cor. **0.42**

This statistic is an indicator of the discrimination of the item: that is, whether students who do well on the test generally got credit for the item, and students who did less well on the test generally did not get credit for this item. A number above 0.20 is considered satisfactory, so at 0.42 this item has good discrimination.

Weighted MNSQ **1.03**

This "fit" statistic shows whether the item appears to be measuring the same kind of thing (reading) as the other items in the assessment. Items with figures around 1.00 fit well: they are contributing to the whole test the optimum amount of information expected from any one item.

Label and Score

The numbers under the Label caption identify the four alternatives of this multiple choice item; under Score, the number indicates which of the alternatives is the correct answer (indicated by 1.00). In this item, alternative 2 was the correct answer.

Count and % of tot

The figures in these columns show respectively the number of learners and the percentage of learners who selected each of the alternatives. For example, 145 of the 230 learners who were administered this item (63.04 percent) selected the correct answer, alternative 2.

Pt Bi s

The point biserial correlation gives a statistic similar to the Item-Rest Cor. described above; in fact, in a multiple choice item, the point biserial correlation of the correct answer and the Item-Rest Cor. are identical. The key should have a positive value, typically above 0.20, while for the distracters (incorrect answers), figures below 0 are optimum. In this item, all of the distracters' point biserial correlations are below 0.

WLEAvg

This statistic indicates the mean ability, in relation to the whole test, of the learners who selected each alternative. Ideally, the mean ability of those who selected the correct answer is at least 0.5 above the mean ability of any other group. At 0.51, the mean ability for alternative 2 is well over the nearest mean ability statistic for any of the other alternatives (-0.64).

In summary, the analysis of this set of item statistics indicates that item R000014 has robust psychometric characteristics.

Each item included in the trial test was analysed in this way, and items were only selected for the assessment tool if they met established statistical standards in terms of reliability and discrimination, of the kind outlined above in relation to Figure 2.

Conclusion

This paper has focused on describing how the material for the assessment tool was developed, and demonstrating that it was a rigorous and multi-faceted process. Each item included in the assessment was submitted to several qualitative and quantitative filters. Users of the assessment tool should thus be reassured that the development and selection of items included for the purpose of measuring adult learners' reading, numeracy and writing was based on sound theory and practice.

The companion piece to this paper, Charles Darr's Inside the Assessment Tool Part 2, discusses some of the larger, test-level aspects of the assessment tool's development.

4. Inside the assessment tool (Part 2)

Charles Darr, NZCER

Charles Darr is Manager of Assessment Resources and Development at NZCER.

Introduction

Assessment is concerned with developing an understanding of our learners. When we assess, we hope to generate information that can lead to productive decision making around teaching and learning. As educators we have a number of techniques we can use to generate this information.

Different assessment techniques offer different types of information and all come with different strengths and weaknesses. Some techniques are very focused, others are more general. Some are quite formal and use objective scoring techniques, while others are more casual and scoring (if done at all) is much more subjective. Some use tasks that breathe authenticity, while others are confined to pencil-and-paper approaches. Ultimately, we weigh the costs and opportunities different options present and make choices about assessment techniques to suit the purposes for which we assess.

The Literacy and Numeracy for Adults Assessment Tool adds to the educator's collection of assessment techniques. The tool harnesses a number of technologies to gather rich and useful information about the literacy and numeracy competencies of learners. These include the application of modern test theory and computing power. This paper begins with a short discussion of these two technologies, before describing how they are utilised in the adult assessment tool.

Testing and Modern Test Theory

Testing is concerned with measuring the level of competency a learner has reached in a learning domain; for instance, how well a learner can comprehend what they read or apply ideas about number.

The technique of testing involves administering questions or tasks to learners that require the knowledge and skills that characterise the learning domain being assessed. The learner's ability to respond correctly is seen as an indicator of the learner's level of competency. A high number of correct answers is associated with a higher level of competency. Conversely, a low number of correct answers is assumed to indicate lower levels of competency.

In traditional test theory the number of questions a learner gets correct (the test score) is taken as the measure of competency. However, test scores (often called raw scores) depend on the test taken. A score of 92 percent on a difficult test suggests a different level of competency than 92 percent on an easy test. In traditional test theory the measure of competence is bound to the test. This makes it difficult to compare learners' test scores when different tests have been used to generate the scores.

By contrast, Modern Test Theory (MTT) focuses on producing scores that are independent of the particular test used to assess the student.

MTT, which is often referred to as Item Response Theory (IRT) begins with the idea that there is a probabilistic relationship between the level of competency a learner has and their ability to answer test

questions. According to MTT this probabilistic relationship can be modelled by a family of mathematical functions (referred to as item response functions). The item response function used in the adult assessment tool is known as the Rasch Model.

MTT makes use of this probabilistic relationship to construct a scale that measures the competency we are assessing. Both learners and questions can be located on the scale. For learners, higher locations on the scale indicate higher levels of the competency being assessed and are associated with greater probabilities of being able to correctly answer questions that involve that competency. Similarly, questions that are located high on the scale indicate that they require greater levels of the competency to be answered correctly. These questions are most likely to be answered correctly by learners who have high levels of competency.

The development of an IRT scale starts by locating a bank of possible test questions on the scale. To do this, questions are administered to learners in trials and the response data analysed using the mathematical model. Checks are done to see that the modelled relationship between success on each question and its location on the scale is confirmed by the trial data. Questions where responses don't appear to fit the item response model are examined and sometimes excluded from the final bank of questions.

Once we know where the questions are located on the scale, we can convert the learner's score to a scale location.

Once constructed, an IRT scale allows us to:

- assemble tests at different difficulty levels
- compare assessment results even when the assessments are made up of different questions
- estimate the probability of a learner at a known scale location correctly answering questions located at different points on the scale.

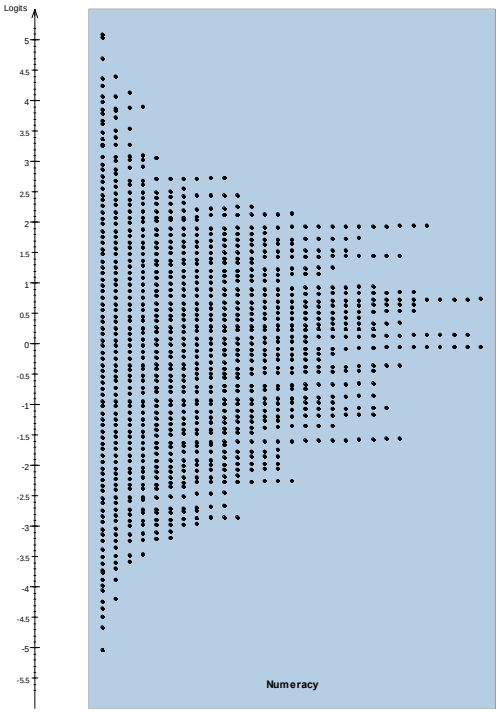
Calibrating the bank of questions

A big part of developing the adult assessment tool was concerned with locating the questions on a scale. A trial design was used that made sure each of the trial assessment forms contained questions that were included in other trial assessments. In this way the difficulty of any question could be connected and compared to the difficulty of every other question that had been trialled.

Once all the trial data were collected they were analysed using the Rasch Model. Fit statistics and graphical indicators were used to show how well item responses fitted the model. Poorly fitting items became candidates for exclusion. The analysis resulted in the confirmation of a scale in each of the learning domains—reading, writing and numeracy.

Figure 1 shows all the numeracy questions on the numeracy scale. Each dot represents a question in the numeracy bank. The questions were written to represent the increasing sophistication of knowledge and skill described by the steps of the Learning Progressions for adult numeracy. As can be seen, the questions are spread up and down the scale.

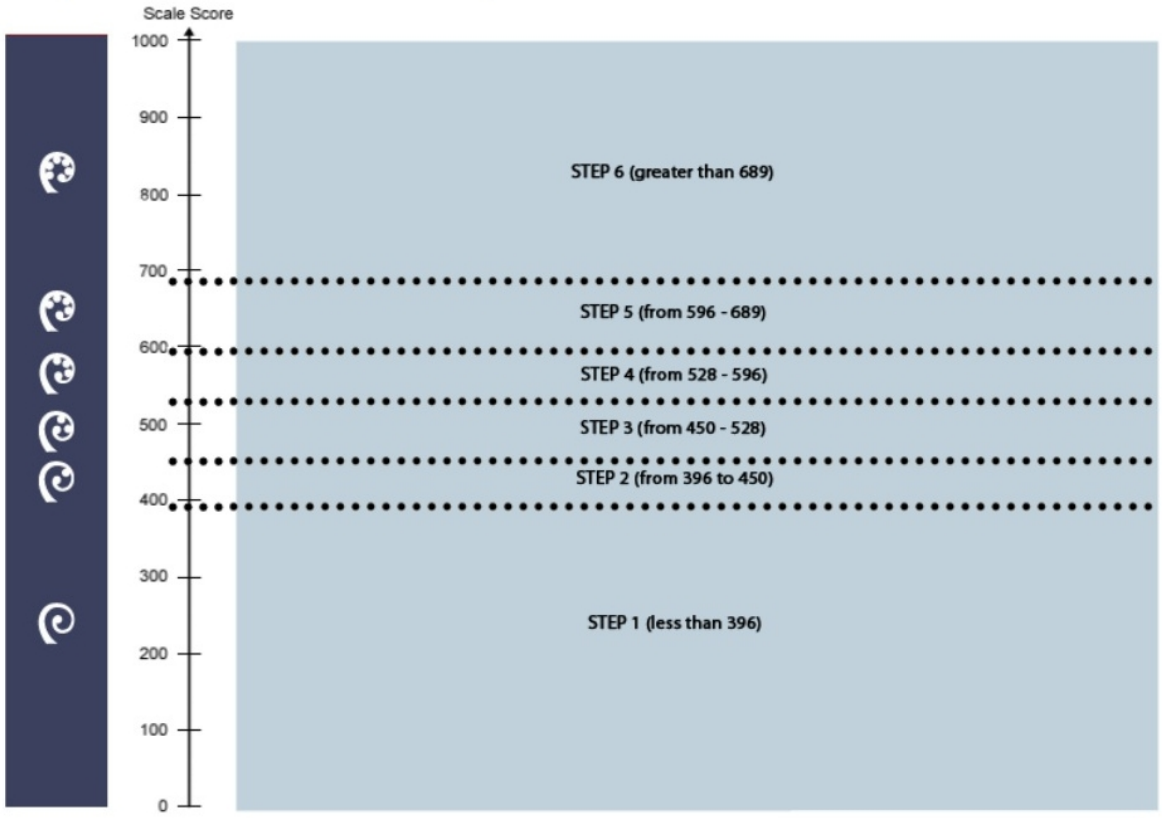
Figure 1 **The numeracy scale showing question locations**



Linking the scale to the adult numeracy progressions

Because all the questions are located on the scale, it is possible to link scale locations to the step levels of the adult literacy and numeracy progressions. In a step-setting exercise, a panel of judges worked their way up the scale looking at examples of questions. The judges' task was to find the locations on the scale where the level and type of competencies required to correctly answer the questions moved from being best described by the descriptions of the competencies typical of one step level to being best described by the descriptors for the step level above. The exercise made it possible to divide the scale into different regions with each region representing the performance level required to be at a step. In the figure below we have taken the numeracy items and drawn cut points showing these step regions. When a learner's achievement falls in one of these regions it is described by the step level for that region.

Figure 2 **Scale locations associated with each of the numeracy steps**



Estimating learners' locations on the scale

Locating questions on the scale like this is often referred to as item calibration. When a bank of items is calibrated we can select a number of questions (say 30 or so) from the bank, and administer them to learners. Once the learners have responded we can find the place on the scale which best fits the level of success they had on the items.

Notice at this stage, however, that we don't necessarily have to use a computer to make use of the bank. Theoretically, we could simply combine the items into a paper-and-pencil assessment and use it with learners. We could then mark the assessment and use a look-up table to find where each student was located on the scale.

So where does the computer come in?

Computer technology

First, a computer is a place to store the questions. Each of the questions can be stored in a database and retrieved as necessary. For instance, in the adult assessment tool, there are over 1,100 numeracy questions that can be called upon. A computer can retrieve and display these on demand.

The computer can also administer the items to learners. In many cases, it can add an element of interactivity and engagement.

A computer can combine items together to assemble assessments. Assembling assessments involves selecting a set of items for learners at particular competency levels. The adult assessment tool can assemble assessments at different difficulty levels. These can then be printed out or done online.

A computer, however, can go one step further than assembling a collection of items. It can optimise an assessment as it proceeds. This is often referred to as Computer Adaptive Testing or CAT and, like test assembly, is a function offered by the adult assessment tool.

CAT is a collection of techniques that exploits a computer's ability to store and retrieve information and make quick calculations. It involves the computer tailoring the assessment to the individual learner who is being assessed.

Before we look at how the assessment tool creates a CAT, let's look at why it is advantageous to do this.

Quality testing

As has already been noted, tests gather information by providing a selection of tasks (items) that require the kinds of competencies and skills that make up the learning domain we are interested in. Success or failure on the tasks is used to indicate the presence or absence of these competencies.

A good-quality test seeks to:

- provide a balanced range of tasks that represent the cross-section of skills and competencies involved in the learning domain being assessed.
- match tasks to the achievement levels of the learners.

When we don't succeed with the first of these objectives we have a validity problem. When we have problems with the second we have a precision issue.

Validity refers to how well we can justify the inferences we make or actions we take on the basis of an assessment result. Our ability to make valid inferences and decisions about a learner's level of competency is enhanced when the assessment contains tasks that draw on the range of competencies that characterise the learning area.

One way to think about this is to imagine a learning domain as a loosely bounded set of skills and knowledge. Our aim is to infer how well a learner can work with the ideas and skills that the bounded set contains. If our assessment focuses on only a small subset of the skills and knowledge involved, then our ability to say something useful about the learner's competency in the learning domain as a whole is very limited.

Another way of thinking about this is to imagine an assessment as an attempt to uncover a painting by repeatedly shining a focused beam of light onto parts of the canvas. Our ability to make valid inferences about the picture as a whole is constrained if we only concentrate on one part of the canvas. A much better technique would involve sampling across the canvas to build up a strong idea of the picture it contains.

The assessment tool is able to call on a rich bank of items which have been clearly designed to elicit the kinds of skills and knowledge that define the learning areas being assessed. Each item is designed with the intent of making learners apply aspects of the skills and knowledge that make up the learning domain. In order to develop a comprehensive picture, it is important that items are selected as a representative sample of the entire domain.

What about the precision issue?

If we stick with our painting metaphor then precision in testing is a little like the resolution of the picture we come up with when we shine our light on the canvas. Low resolution means the pixels that make up the picture are large relative to the picture as a whole. As a result the picture is blurred—it is difficult to discern what we are looking at, even when we have sampled across the whole picture. When we use a higher resolution we have more information to work with and the picture becomes clearer.

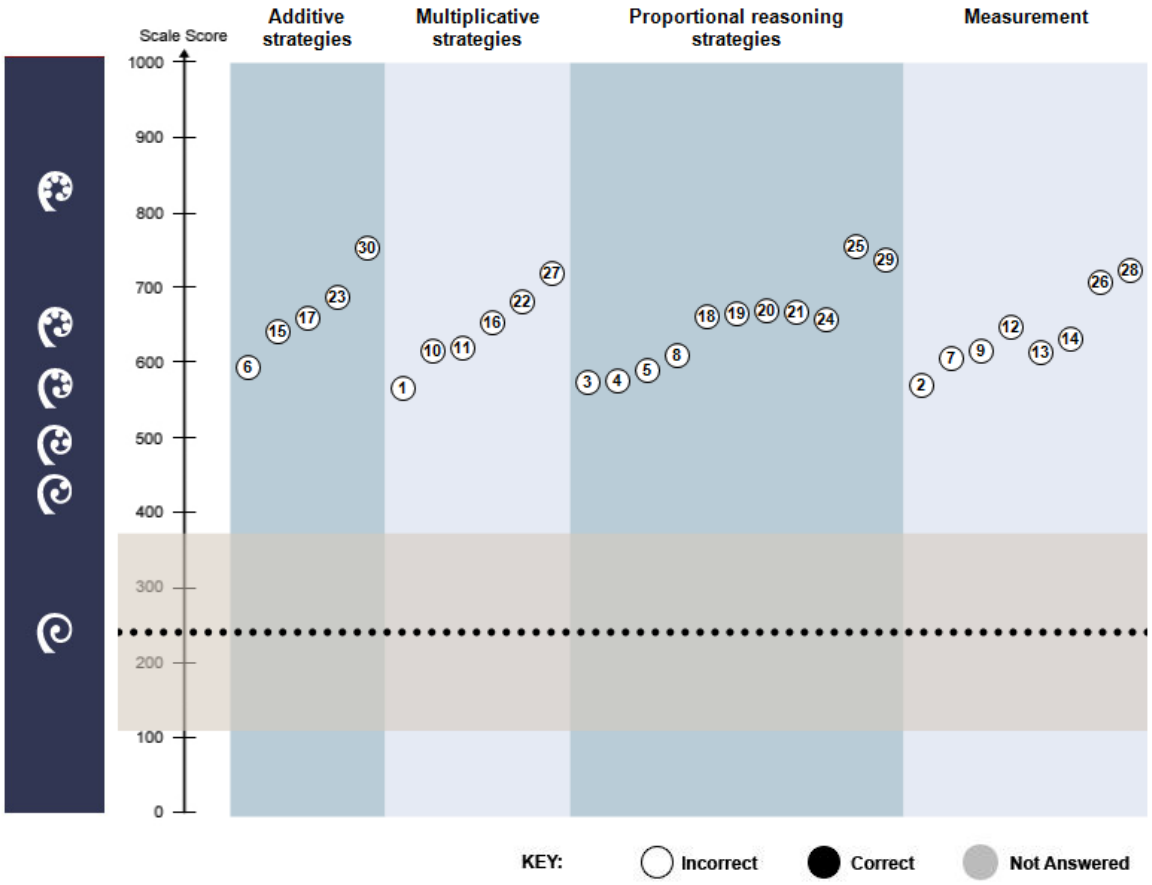
The precision of an assessment is affected by how many questions we ask and how well those questions target the learner's level of competency. To maximise precision we must target the questions that can provide the maximum information about where the learner's achievement level fits on the scale.

If the questions in an assessment are too easy, the learner will get most, if not all, of them correct. When this happens, all we can tell is that the learner is higher on the scale than most or all of the questions, but we have very little information about how high. Similarly, if a set of questions is too hard the learner is liable to get them all wrong or nearly all wrong. We can tell that the learner is lower on the scale than most or all the questions, but we have no idea how low.

In traditional paper-and-pencil-based testing situations the same test is often used with large groups. Because the competencies in a group vary, test developers will usually write the test to suit the majority of people who fall in the middle of the achievement distribution. This means that low and high achievers can be faced with a test that doesn't assess their competency level very well. The scores for these learners are unreliable—they will tend to get notably different scores if they are assessed again.

The figure below shows a report for a student who has taken a paper-and-pencil test that is not well targeted. The questions have been too difficult for the learner. As a result, they have got none correct and the estimate of where they are on the scale (shown by the dotted line) is very imprecise (shown by the grey shading). The estimate of imprecision is often referred to as the measurement error. It provides a plus or minus range around a test result. With the Rasch Model we can calculate the measurement error for every possible score on a set of questions.

Figure 3 **A graphical report for a poorly targeted assessment**



Optimising an assessment then involves using tasks that characterise the learning domain well, and that are targeted at the competency level of the learner. The first of these objectives can be achieved by carefully developing a bank of rich assessment items. The second presents more of a challenge. It involves somehow personalising the test to each learner to make sure the items that each learner is administered best target their achievement level. This is where a CAT comes in.

Computer Adaptive Testing

In a CAT, an internal algorithm is able to calculate interim scale locations for a learner as the assessment proceeds, and adapt the assessment so that questions are administered that best fit the emerging achievement level of the learner. At the same time the algorithm can register the types of questions that have been administered and work to maintain a balance of question types across the learning domain.

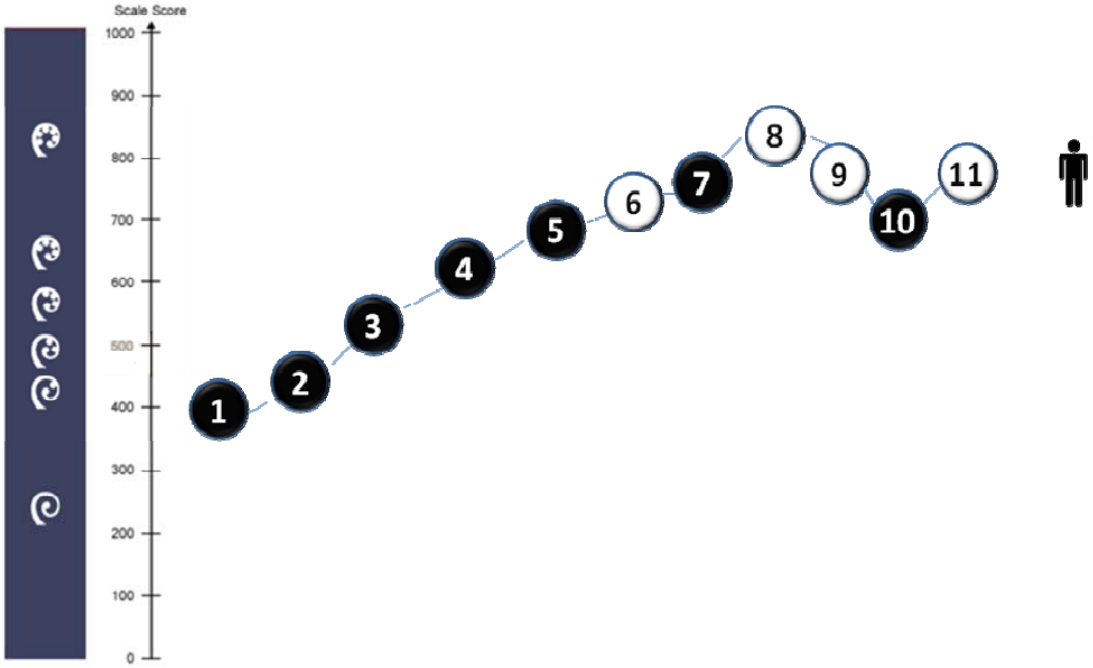
We can use a graphic representation to journey through an assessment with a learner and see how the algorithm works question by question to optimise the test. The assessment we will use is one administered by the adult tool designed to assess a cluster of four numeracy progressions: Multiplicative Strategies; Additive Strategies; Proportional Reasoning Strategies; and Measurement.

When the CAT starts, the tool has access to a large bank of items which have all been calibrated on the same scale. It also has a large collection of information about each item: for instance, which of the learning progressions it represents; what the correct answer is; and whether the question involves a context. This kind of information is often referred to as the item's metadata.

At the start, however, the tool doesn't know anything about the learner's level of competency. So it begins by choosing at random an item from the mid to lower levels of the bank and administers it to the learner.

Figure 4 represents this by placing a circle with the number "1" inside it next to the scale. The circle is presented in black to indicate that the question has been answered correctly.

Figure 4 **A map of a learner's journey through a CAT assessment**



The computer now has a piece of information to work with. The first item has been answered correctly, so the best course of action seems to be to administer a slightly harder item. It does this and again the learner answers correctly.

The computer now has two pieces of information. It decides to administer a slightly more difficult item. Once more the candidate answers it correctly.

Figure 4 shows that this continues until the sixth item. For the first time the learner answers incorrectly.

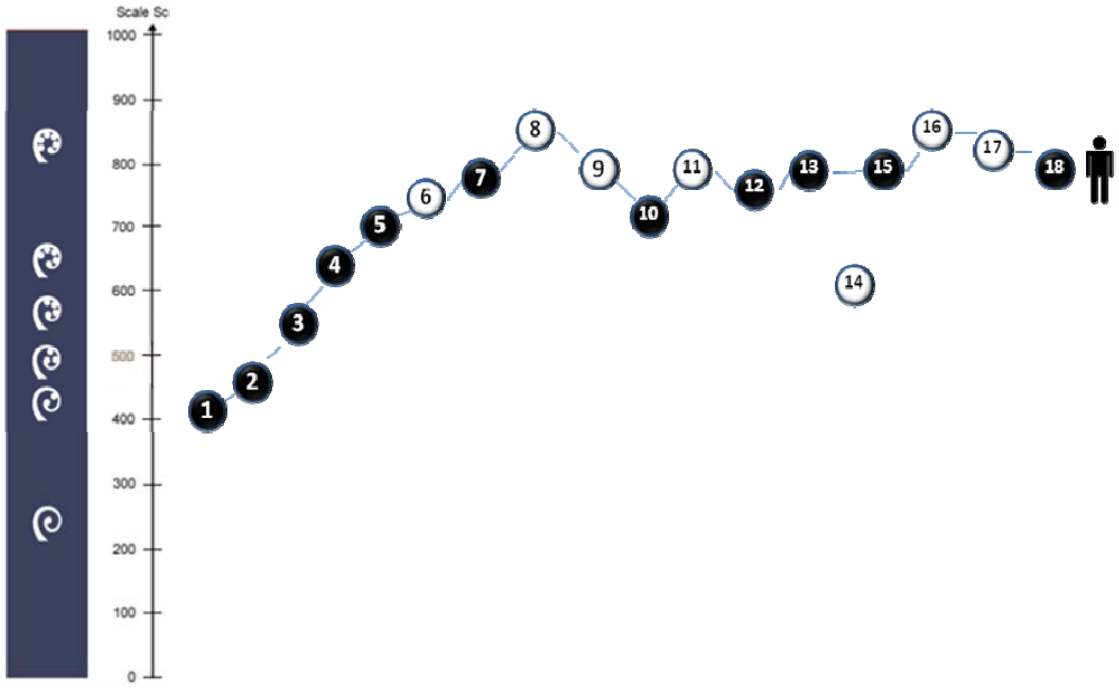
The computer considers the information and calculates where on the scale the learner is most likely to be if they have answered five of the six questions administered correctly. It uses this estimate to choose the next question that best targets this location.

When the computer selects an item it always selects items randomly from within a band of items centred on the best estimate it has. This helps ensure that items are not overexposed.

At the same time as the algorithm is selecting items to administer it is also keeping a record of the kind of items it is using. It knows the assessment it is delivering is an assessment of four of the Numeracy Progressions. It also knows that some questions are represented in contexts and some without. It wants to make sure that a minimum number of both types of questions appear in the assessment. If a minimum of one question type is reached the algorithm will try to select questions from another area before considering selecting that type of question again.

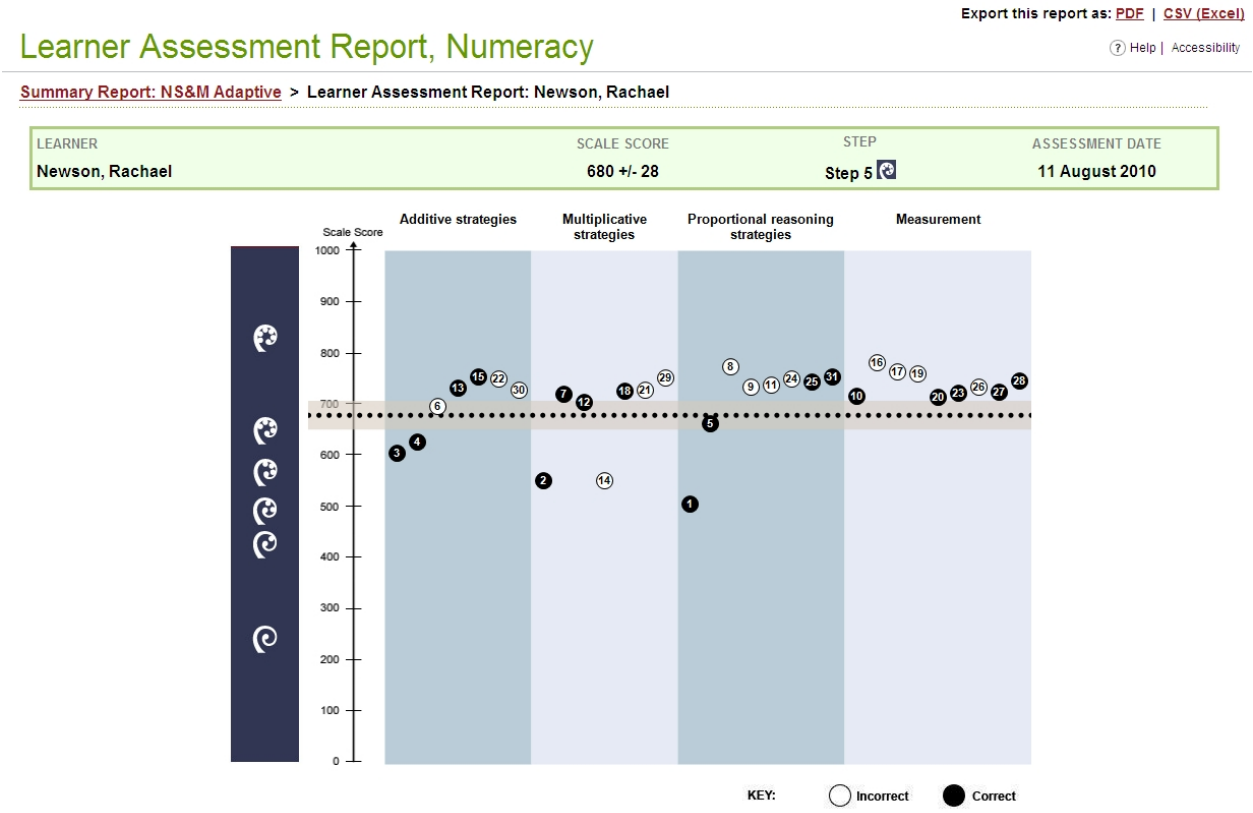
Figure 5 shows the assessment after 18 items. As can be seen, the learner is now having much more mixed success and the steep trajectory representing the learner's interim achievement estimates has flattened off. The learner seems well targeted by the questions being asked and the estimate of where the learner is on the scale is relatively stable.

Figure 5 **A learner's CAT journey continued (Q1 to Q18)**



Once a learner has completed an assessment, the assessment tool allows us to examine a report which shows how the learner has done on the whole assessment. An example of the report is shown in Figure 6. The report is very similar to the map we have been using to illustrate the journey through a CAT. There are several features worth noting.

Figure 6 **A report for the assessment generated by the assessment tool**



First, we can see how the student has done on each question. Incorrect questions are represented by blank circles, while the ones answered correctly are shown as filled-in circles.

All the questions are organised according to the Numeracy Progression they represent. In this case, we can see that the learner has answered a mix of questions from across the four progressions assessed.

The line is used to show where the learner is located on the scale. In this case, 680 units on the scale. The scale has been designed so that a learner's location on the scale indicates where they would have a 70 percent probability of answering questions located at the same place. A 70 percent success rate is an indication that the learner is likely to have a strong understanding of the kinds of tasks located at and below where they are on the scale.

It is also possible to see how the learner's achievement connects to the step level. In this case the learner is classified as working at Step 5.

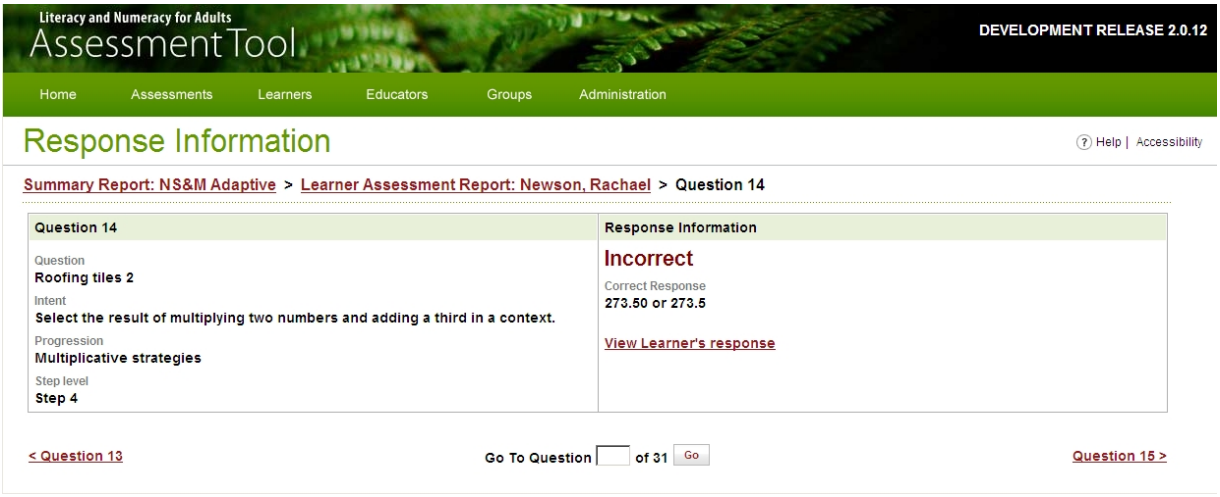
It is important to notice the fuzzy region that surrounds the line used to indicate the learner's overall level of achievement. This is used to indicate the plus or minus region that represents the measurement error. We can be reasonably sure that the learner's true location on the scale falls within 28 points of 680.

If we scroll down on the report we can find some more information. On the second page we have a record of which questions the learner got correct and incorrect as well as a short statement about the intention of the question.

The assessment tool is designed so that we can interact with the report. It is interesting to note, for instance, that the learner has got Question 14 incorrect. This is a little surprising given that it is located much lower on the scale than the learner's overall achievement level.

The assessment tool allows us to click on the question number and navigate to another report about that item. This report shown in Figure 7 provides extended information about the item, including the correct answer for this question and information about used and common misconceptions associated with any of the distractors (the incorrect options). One more click on this report and we are taken to a view of the item itself. Here we can see the option the learner chose and get a feel for what the question was assessing.

Figure 7 **An item report**



Functionality built into the adult assessment tool makes it possible to surf between reports and explore the learner's journey through the assessment. Other reports let us look at how the learner has progressed from an earlier assessment or how well a group of learners has performed as a whole.

Final thoughts

At the start of this paper, I referred to the adult assessment tool as an addition to the educator's collection of assessment techniques: one that harnesses the power of computer technology and MTT to provide opportunities to learn more about the learner. As we have seen, the CAT technology built into the tool coupled with the extensive bank of items provides advantages not easily duplicated with traditional paper-and-pencil testing.

There are often two very different reactions to the harnessing of technologies. One can be characterised as that of the "super-adopter". A kind of wholehearted, uncritical acceptance that is all too aware of the possibilities and blind to any limitations or possible negative consequences. At the other extreme is the Luddite position. Here, like the original Luddites who at the start of the Industrial Revolution destroyed the new machines that threatened identities and people's livelihoods, any possibility of a use for good is outweighed by an overwhelming sense that the ultimate end is negative and perverse.

The same perspectives can be applied to the tool. For those enamoured with data—managers, policy makers, statisticians and psychometricians—it holds boundless possibilities. For others, some of whom may be educators or theoreticians, it can suggest a mechanisation of assessment that disempowers the educator and depersonalises the learner.

Without taking the Luddite position, I would suggest that the tool must be used professionally and with care. All technology has its blind spots and doubly so when it can be used to measure and categorise people. I would encourage the super-adopters, and particularly the policy makers amongst them, to make sure teaching and learning are front and centre in the way the tool is deployed, particularly in terms of the stakes it represents for learners, educators and organisations.

Without falling into the first camp of uncritical acceptance, I would like to suggest the new tool is full of possibilities for educators and learners. Properly utilised it can be a starting point for more powerful learning. As such, I believe it is an investment that can pay back into people's lives and the economy of New Zealand many times over. I encourage educators to explore the possibilities and add it to their assessment tool kit.

The Literacy and Numeracy for Adults Assessment tool was developed by a team of people committed to developing a tool that would make a real difference. We will continue to support its day-to-day use and we look forward to hearing your stories, the stories of educators and the adult learning community, about how the tool is being used to improve learning and teaching. We hope the tool will continue to develop, evolve and improve, and truly add value to the adult education project in New Zealand.

5. Looking at assessment through a Māori lens

Cheryl Stephens, Tamati Waaka, and Aroha Puketapu-Dahm
Te Whare Wānanga ō Awanuiārangi

Introductory remarks from Jessica Hutchings, Manager of Te Wāhanga

Te Wāhanga, the Māori Unit at NZCER, is really thrilled and delighted that we can privilege some of the space today in this conference to engage with Māori scholarship, Māori academic thinking, kaupapa Māori thinking, matauranga Māori thinking, thinking that's based on our tikanga and to have researchers from Te Whare Wānanga ō Awanuiārangi based in Whakatane, come and share whakaaro with us today about literacy through a Māori lens—a kaupapa Māori way of thinking or a way of looking at it. Just before I do introduce the panel I want to remind us of the bigger context in which this korero this afternoon around kaupapa Māori assessment or assessment that engages Māori is framed. Last week we had the privilege, we in the Māori Unit as well as quite a few people in the audience, to attend the Ako Aotearoa hui for Māori who are participating across the whole tertiary space. And congratulations to Ako Aotearoa, to Peter and Ngahiwi and the team there for pulling together such a successful hui to engage in the broader issues that Māori are facing, not necessarily just in the tertiary sector—although that hui was focusing on the tertiary space—but within the context of education in Aotearoa and more broadly within the context of continuing to live in a country which is colonised. So I do want to remind us of the wider context in which participation for Māori in education happens.

Last week at Pipitea Marae at the Ako Aotearoa hui, there was a lot of talk about second-chance learners, of Māori as second-chance learners and a point was made that actually for many Māori, engaging in tertiary study or engaging in certificate-level courses is our first opportunity to learn because the schooling system has failed us, that we were never given that first chance to learn as learners. So I just want us to think about the korero today within that broader context of education for Māori in Aotearoa. I also want to come back to the point that I made in my opening remarks at our conference today around models of assessment being culturally located. We need to hold this whenever we are developing models of assessment or frameworks, that these are embedded within the cultural assumptions from those who are framing or who are developing those assessment frameworks.

I also want to tutuku your korero Charles and the point that you raise that the "tech talk" can measure certain aspects but it cannot measure all aspects. When we talk about assessment and measurement and wellbeing, that's really pertinent for us as Māori. We need to think about what it can measure but also what its limitations are. I have just been told by the panellist that Te Whare Wānanga ō Awanuiārangi is trialling the tech assessment tool to see whether or not it has the ability to actually be able to support Māori educational aspirations. So really it's a nice flow on from the last presentation to be able to engage with the whakaaro of the speakers from Awanuiārangi.

I also too want to acknowledge the leadership of Awanuiārangi in thinking about assessment from a kaupapa Māori space, from a culturally responsive space and just to remind us that it is not always Māori's responsibility or our responsibility as Māori to be the ones doing this thinking.

I'd like to welcome to the panel our three members: from Awanuiārangi, Cheryl Stephens, Tamati Waaka and Aroha Puketapu-Dahm. Kia ora.

Cheryl Stephens

Cheryl Stephens is the Director of the National Institute of Māori Education, Te Whare Wānanga o Awanuiārangi.

Our presentation is in three parts: three people, three parts, good numeracy. Our Chief Executive Officer was actually asked to speak and he chose to pass it to me and when Jenny rang I said, "I'll do it as long as Aroha can do it with me" and then Aroha said, "I'll do it as long as Tamati can do it with me." Consequently, there are three of us. So culturally we feel really safe in the context of this panel today.

I want to do a little interactive exercise with you all, so bear with me. I want you to look at someone near you and I want you to lift your eyebrows without touching them and face a person who is near you. Go. Okay, now tell that person what you said.

So for us, that's a particular literacy, a cultural literacy that some of you in the room can relate to. Here's another one. Different person if you like. Shrug your shoulders at somebody. There's some laughter that tells me the literacy. Now tell that person what you said.

So those are particular literacies that suit different cultural contexts, different cultures and some people know the messages, some people will find out the messages in different ways. We want to share some of those different cultural literacies with you. And I learnt a word last week at a Māori literacy hui that we ran in Whakatane—paralanguage, paralinguistics.

It means body language. But it can also mean surroundings or particular icons or particular visual elements which, for us, is text. How many of you have been to a marae? There are whakaaro, there are carvings in that context. Those are paralanguage items. They tell a story, they have a voice, they are visible and they have a whakapapa. Those are some of our cultural literacies.

Awanuiārangi is the youngest wānanga of three: Aotearoa, Raukawa and us. The Education Act is what gives us our context, our content and who we are. So it's about matauranga Māori and tikanga Māori. We're located in the North Island, in Whakatane. We're located in an iwi context—Ngati Awa is the iwi that's located there but we're in a wider context, our mataatua waka. This is our context.

What does this picture tell you?

Let's see if you can write down three words that that picture tells you. Now, I'm making the assumption that you can write quickly and think quickly. Who'd like to share? Anybody want to give us any words that come to mind based on that literacy?

Audience: Colonisation.

Audience: A behaviourist view of learning.

Audience: Resentful.

Audience: Native school.

Native school. Unhappy tamariki. Some other words we could use: regimented.

Audience: Disciplined.

Audience: Imperialistic control.

Imperialistic control. Here is another slide which says that 10,000 acres of prime Ngati Awa land was confiscated and given to endow what was called the University of New Zealand at that time which is now Auckland University. Auckland University—and I know some colleagues are here in the room—is on Ngati Awa land. So instead of Ngati Awa saying we want our land back, Awanuiārangi has a Memorandum of Understanding (MoU) with Auckland University which is around, "We helped you to get started, you help us to get started." So we have an arrangement for them to help us in terms of our academic programming. Ngati Awa went through a Treaty process for that land confiscation and Awanuiārangi, our campus, is on some of that Ngati Awa land which came back as part of that Treaty process.

We began with a very simple idea which came from Hirini Moku Mead. He wanted a wānanga, he wanted Ngati Awa people to be educated. He got a whole lot of other people thinking that same idea and no money. There were lot of people with a whole lot of ideas, a lot of commitment and vision and passion to start Awanuiārangi and this is one of the photos of those very first buildings—they're actually still there.

The idea was to have a pot of paint, hammer, nails, a piece of board, put a name on it, put it on the wall, start. That's how Awanuiārangi started. Our critical literacies have a social conscience, a social justice focus; it's about citizenship, and a kaupapa.

One of the other elements of our work that reflects that is that we have now passed Te Tai Tokerau and Northland in terms of health deprivation statistics. Our region is the most socially deprived community in the country now and, whilst we're not proud of that, it actually underpins a lot of the work we do in terms of education in our community. So this literacy/numeracy work that we're doing is pivotal in changing the colour of that map. Our community is really, really important to us. In particular, our kaumatua. We have an annual celebration of our kaumatua—kaumatua day—which is usually in November. We have a kaumatua ropu that interacts with us and helps guide us in the work we do as part of our education process. Awanuiārangi's title is Te Whare Wānanga ō Awanuiārangi. The Whare part is really important. It distinguishes us from the other two wānanga in terms of higher learning. And so Whare has a reflection in terms of the Whare Wānanga, the traditional Whare Wānanga. One of our goals is around growing Māori scholarship and leadership, and literacy and numeracy are really important as part of that process. We have Level 1 to Level 10 programmes. Level 1 on the National Qualifications Framework is certificates, Level 10 is PhD. We're the only institution outside the university sector with PhDs. When we started this work 18 months ago we said to the Tertiary Education Commission, "We have literacy and numeracy needs in all our programmes from Level 1 to Level 10." That's really important for us to acknowledge. So the work that Tamati and Aroha are going to talk about is around the certificate programmes. But at the PhD level we still have people who need some assistance with literacy and numeracy, at a different level of course.

We have three schools. The first one is our School of Indigenous Graduate Studies which is where our Master's and PhD programmes are based. These are just some of the statistics in terms of the doctoral students that we're working with and these are some national statistics that we're using as an aspirational tool in terms of growing that capacity and capability of our institution, and of Māori and across the board. I do want to acknowledge our head of our School of Iwi Development, Te Waka, who is here today and also Elder Te Reo, who is one of our regional co-ordinators.

So our work is based in our School of Iwi Development and it's marae-based and Tamati will talk further about that work. There's lots of work in the programme end and once again Tamati will follow that up. We have a School of Undergraduate Studies which is mainly Bachelors' programmes. One of them is our teacher education programme and we've done some trials with the assessment tool with our education students as well as our iwi development students.

Te Reo Māori is one of our core programmes. Te Reo Māori is in the Level 1 to level 10 in our programmes and Te Reo Māori language excellence is one of our key focuses. The institute that I'm the director of is the National Institute of Māori Education. We have a national focus in terms of the work we do. It's very young—born this year—so it's still in its infancy, but it's around adult literacy and numeracy, educational research and development. It's also indigenous leadership and teacher education and adult education. So we have some education programmes, early childhood, early years, primary. We're working in collaboration with Waikato University on He Kākano which is a three-year school leadership project that we have just started to roll out into 100 schools.

There are some pivotal people in Māori society who have had a strong influence in education across this country. You've got Graham Smith, our Chief Executive Officer, you've got Anne Salmond, Bruce Biggs, Ranginui Walker, Hirini Moku Mead, Hugh Kawharu and Roger Green. People who've been pivotal in Māori education and Māori aspirations and Māori development over time. Some of those people taught in our Master's programme when the wānanga just started, you know, with the hammer and nail sort of kaupapa. They came in and taught for free. These are some examples of literacies and the critical literacies, and I was pleased to hear Paulo Freire talked about and quoted this morning because that's a pivotal element of the work we do in Awanuiārangi . Those critical literacies, the politicising of the work we do. Teaching is a political act. These are some of those aspects of that work and, of course, it goes into Parliament as well. I was pleased to see this on your slide this morning.

Computer literacy is another element of our work. "Tech pa" is a new innovation that we have begun which is about bringing school leavers who are at high school into the wānanga to have a tutu. (Those people who know what that means can explain that to the people who may not.) The tech pa is an innovation but it's really around getting young people engaged with education using technology. So we have an after-school programme where they come in to use the computer technology, use the resources in terms of human resources, to create and innovate in whatever way they want. Here's another literacy. There was a series of paintings done by Cynthia Kapene, one of our staff, and this is one of those paintings.

Can anyone see any relevance to what's on that painting to the work you do? Does anyone know some of those words and where they might come from? What does that picture tell you? Anyone know some of the quotes that are there, where they come from? That was a series of six paintings and they were actually painted for a charity ball we hosted last year for the Child Cancer Society and for the local Eastern Bay Hospice. And so they were auctioned as part of that. So we have an aspiration for educational excellence and I was really pleased to hear the either/or, that it's not "either/or" mentioned by a number of speakers this morning. We have that same philosophy. It's not about compartmentalising our Te Reo Māori Matauranga Māori and leaving the English medium on the side. We're about collaborating and pulling those elements together. And you'll see some of that through the next two presentations. For our graduation we have a theme called "Gown in Town" and we literally take the gown into the town. We've had two years of it and it's been phenomenal in terms of engaging with the community—the wider community—and letting them know what we're doing, because there is a perception that Awanuiārangi is only for Māori people. It's not. So we're at a stage where there's a sense of renewal. Ngati Awa has settled in terms of the post-Treaty settlement process and we say that that's a beginning, not an end. The beginning of the work to go through in terms of what we do now that we've settled. And a number of iwi are also thinking in that way at the moment. So we have a post-Treaty settlement centre that's going to be launched here in Wellington at the end of this year and that's about providing some education focus, providing some support to those people, to those iwi who have settled through the Treaty process. So we're looking at the depth and breadth of knowledge and, as I said, not either/or, English/Māori, it's complementary in terms of the work we do and you'll hear that from our next two speakers. So thank you and I look forward to some of your feedback. Kia ora.

Aroha Puketapu-Dahm

Aroha Puketapu-Dahm is the Numeracy and Literacy Project Leader at Te Whare Wānanga o Awanuiārangi.

My name's Aroha Puketapu-Dahm and kia ora. I don't know how we got roped into this but I'm really nervous actually. I'm not normally nervous and I've sometimes got stuff to say but it must be the company I'm in.

At Awanuiārangi we have been, as Cheryl has said, trialling the national assessment tool under the shelter—whakaruruhau means to afford shelter—of the Tertiary Education Commission. When we were first asked to do some consultation work with this new assessment tool we were thinking, "Oh, man, do we have to do that stuff?" you know, because from a Māori perspective it's like it's new and if it's got anything to do with higher education it's possibly boring and so we were sort of thinking, I was thinking, "How am I going to convince these fellas to, you know, get around the table and trial it?"

So that's exactly what we did. We were in the School of Iwi Development and we were in the main room which is where they have all their hui, which is the kai room. Tamati rocks on in with the national assessment tool, puts it up on the PowerPoint and we're sitting down and we're going through the items and not far into it we have an argument and we're all educators, there's no students in the room and the question was something about the police roading budget. They had given us a figure and after discussing the picture which had a police car in it and then trying to work out the formula for figuring out the roading budget for the next year, we had ended up in a political debate about who the roading budget should be allocated to and who should be contracted to fix the roads up in Whakatane. Anyway, Tuwaka and I won because we had the loudest voices in the room but actually we were wrong. And we were all off the point and from our perspective that was a bit of a distraction because we missed the point—this is about competency. This is about competency and assessment from a Māori perspective. There's this thing called kaumatua tanga. I thought to myself, "Well, if we're going to use this in the wānanga we'd better be able to use it out on the marae."

The Iwi Development team are experts at going out and taking education to the marae. So I thought I'd go and have a look and see what they do. Tamati's going to talk a little bit about that but it's not an easy experience gathering that sort of information. It's not an easy experience having to go out there because you need to be authentically pono in your approach to them. They will see you for what you are and basically they do want to read your paralanguage—what is this person here for? They want to look you in the eye and listen to what you've got to say and tell you how much they know about you. They're determining whether you're wasting their time or not. That's the approach we took when looking at the assessment tool because, if we're going to be putting this in front of Māori learners, we'll scare them away if it's not a safe place for them to be. It's hard enough to get them into the tertiary education sector and participating; we don't want to scare them away. So we need to take that considered approach in this exploratory year, make some decisions, help NZCER and whoever we can to work with the tool and to reframe it.

My initial reaction to most of the items was, "Why are all the pictures of Māori people like pictures? They're actually not images, they're just pictures, sketches." And then, because one of them's called Aroha and the other one's Cheryl and the other one's Tamati, my second question was, "Why are they all ugly?" I was going, "Man, can't someone get out there and just take a photo of a whole lot of Māori and a whole lot of Pacific Islanders" and, you know, that was sort of one of my initial impressions eh, which sort of turned

me off. But actually it's good that we're having this conversation because this is a hard conversation that Māori won't tell you after the fact when they don't participate. So we're just sort of giving the heads up and we're really happy and actually really humbled that Jenny's sort of taking a lot of this because Jenny Whatman's bearing the brunt of all of those types of initial whaakaroa or initial thoughts and she very graciously does that and then writes a report back to us, which is really good. This is a korero and this is a discussion that we need to have about these sorts of things because there's so much at stake for our workforce. There's so much at stake for not only your superannuation, for mine. I'm highly aware of that and for our families. Because what other option do we have? Kia ora.

Tamati Waaka

Tamati Waaka is an Academic Advisor to the School of Iwi Development at Te Whare Wānanga o Awanuiārangi.

My korero here today is relating to assessment—assessment for Māori, how to assess Māori, what Māori think of assessment and if you were at the Wairakei conference or symposium the other month, if I repeat the same joke, don't laugh or don't give it away and if I speak Māori and tell a Māori joke, ask this table why they laugh so hard.

I like to always go back to the Learning Progressions. Know the learner, know the demands, know what to do in terms of embedding literacy and numeracy into our courses, our programmes, our institutions. Those three factors also relate to assessment and creating assessment for Māori that is culturally responsive for all our learners. We have to know the learner. We have to know their demands and then know what to do or what to use as a form of assessment. So first of all before you can understand me, before you can get what I'm talking about, you've got to know me.

So I'm from the Eastern Bay of Plenty, stronghold of Māoridom. If you are a Māori, you came from Whakatane somewhere along the line. Very strong at performing arts, oral literacy. I know that's an oxymoron folks—oral literacy—but there it is. People from the Eastern Bay of Plenty have a great sense of humour. Most people grow up in the Eastern Bay of Plenty, they grow up in and around and on the marae, the hub of the community, the nightclub of the community, the gossip, the Woman's Day of the community, and of course one of my legs is from the coast where the highest form of currency is kina. Down south it's oysters but in the East Coast you've got the kina. So that's me. My first language is Te Reo Māori. A lot of you may know the East Coast from watching Boy the movie. I was just hoping it would get above Whale Rider as the highest-rating New Zealand-made movie in New Zealand, There's a rivalry between this part of the East Coast and Nga Iwi's part of the East Coast. When it jumped over Whale Rider I was stoked, I was stoked. Those Ngati Porou got a lot to say but I've got a comeback—Boy the movie. Rumour has it they're trying to make one called Girl. But anyway, I digress.

Knowing the learner. Knowing me. The word "me", in Te Reo Māori is "ahau". We're taught in Māori class that ahau is me. But ahau also means me and everyone behind me. Everyone where I'm from. Everyone who thinks like me. Welcome to my world.

People like Joan Metge, Ranginui Walker, John Rangihau talk about Māori people being communal. It's about whānau. Rangihau in particular says a whānau means blood ties so everyone says, "Ah, here's my whananga" but has he got your blood? No. He's not your whananga, he's your friend. That's whānau. Hapū, the iwi, again communal. In English if I was going to think of a better English translation then what

Metge, Walker, Rangihau could come up with is interdependence—mahitahi, sharing, and that the whānau is dynamic and fluid. And there was an article in one of Ranginui Walker's books, an introduction called "Being Māori", and some of them capture Māori experiences in different parts of life. For example, for education he says, "Being Māori is going to school to eat someone else's lunch." So this is not me, it's just the insight. Some Māori go to school not to learn. This is true. "Being Māori is punching someone in the nose for calling you dumb." To me I like to translate that in saying, "Actions speak louder than words." But they also have a strong sense of pride. "Being Māori is spending all your money to service the marae at whanangatanga manaakitanga." "Being Māori is having a waka that is a hundred times better than the Queen Mary." Quite proud. And, of course, "Being Māori is getting your Pākehā spouse to rent the apartment." I try to pass for Pākehā but then I open my mouth and I get found out. So, context, context Māori. When looking for Māori stimulae (a new word I heard today) we're looking for Māori context. One problem or issue which is hard to overcome is that in Māoridom there is more than one truth. Jeez, I sound like a pastor. There is more than one truth. More than one way of looking at and approaching things. More than one answer. And a lot of these answers stem from the environment. So because it stems from the environment you see those words again, they keep on coming up in everyone's presentations, it's experienced and it's practised. The Māori world sometimes is totally opposite to any other world. We flip it upside down. So those of you who understand Māori, help me out here. How do I say, "White horse" in Māori? Hoiho ma. The horse is before the white. That just a little example of how we're backwards, well, sorry, you're backwards, I'm the right way up.

In the old culture there was a lot of early emphasis on whakarongo korero and, by the way, even writing, reading—writing and reading need oral language. When you read it, you say it out loud. It cannot exist if it wasn't oral in the first place. We've had paralinguistics, we've had visual language. Now these to me are devices. These are mnemonic devices. Mnemonic devices of course aid recall.

Traditional modes of assessment, traditional learning—titiro Whakarongo.

- Mahia te mahi kei tamariki ana (practise when young)
- Tangata i akona ki te kainga (learn in a safe environment)
- Ka noho te mātāpuputu (succession)
- Hinga atu, ara mai
- Roles of kaumatua
- Observer.

So these are the things that a Māori learner takes with them when going on a course. This is what they know about assessment. Some of them have had bad experience of formal Pākehā education because this is what they knew and they couldn't do an exam because you don't do an exam on the marae. You don't do an exam in the community. So for a lot of people they've had bad experiences.

I do kapa haka. We learn, we practise. The word is practise, we don't write it. We train as a whole, not as individuals. These are just some examples of assessment in a Māori world. The role of the tutor is to lift everyone up to a competent level. Once everyone's competent then we strive for excellence. Once we get excellence we try to lift them up to be exponents.

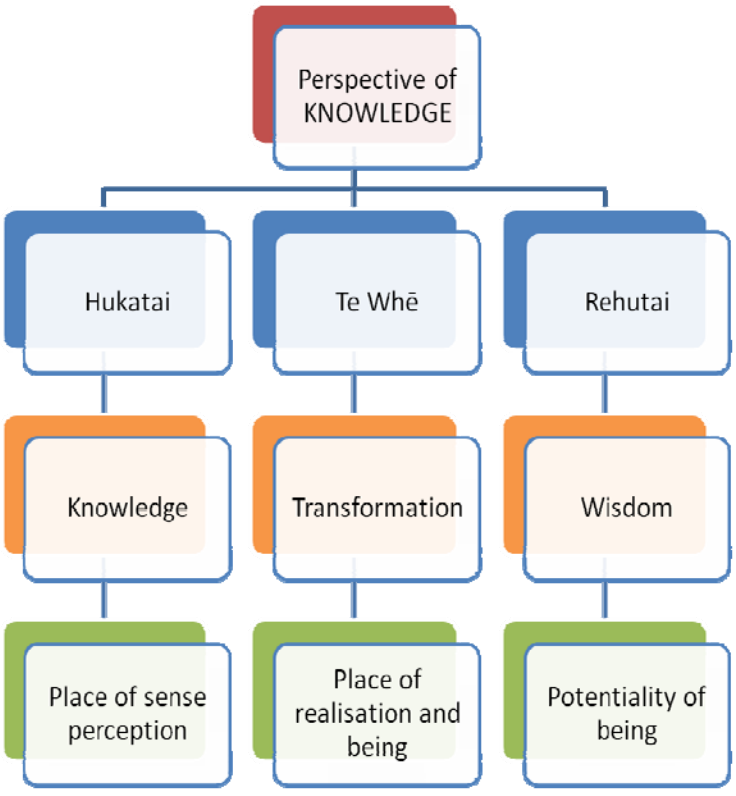
There is no audition, there is no test, unless you're one of those flash towny groups but if you're in the wop wops like me it goes back to the observer. The tutor observes and sees when someone is ready. Ready. Which is most probably the better word for assessment. Are they ready?

So I went to Ako Aotearoa website which I always go to and look at all the time. I was trying to do some reading about assessment for Māori or in a cultural context and I came across this document here, *Hei Tauira—Teaching and Learning for Success for Māori in Tertiary Settings*, which has five points. The main one I wanted to concentrate on is tikanga. To me, tikanga is a way of doing things, that has been done that way for a long, long time. Then it becomes a tikanga. It is sort of like a tradition but also you can manipulate it from time to time. It needs to be lived and practised and not just a theoretical construct.

So I really like the first presenters. I could see that you've had a good, awesome time teaching people and I can see you've done more than the theoretical. You've actually practised it over and over.

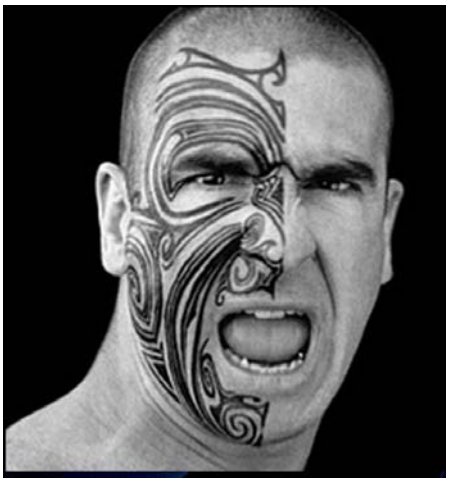
I came across another document done by Ako Aotearoa quite recently, and it provides some guidelines for the development of literacy, of embedded LLN.

A lot of research for adult learning is quite thin on the ground so we always go back to early childhood because there's a lot and they've got beautiful pictures in early childhood. For example, this Master's thesis (*Exploring Whakapapa as a Tool Towards Kaupapa Māori Assessment Framework in Early Childhood*, Vanessa Anne Paki, University of Waikato, 2007) which comes up with a "theory of transformation from knowledge to wisdom as a basis that forms the assessment framework". I just replicated off her.



What I liked about it was to me she put on her Māori spectacles. Hukutai is the beginning, Rehutai is the end test and what we do in the middle to lift them is called Te Whē. So, from a Māori perspective, I can see it. Those terms are familiar to me because I know my Māoritanga. This is one way to contextualise to Māori the importance of a national assessment tool. They say they don't want to do the tool but if you go with a picture that says "You are here and I want you to get here but first you have to do this", it's way more beautiful.

Again, I'd like to thank the Tertiary Education Commission for allowing us the opportunity to work with NZCER to provide feedback and advice using my world view in creating questions or stimulae.



Here is a picture of Eric Cantona with a moko on his face, a half moko at that, drawn on with a felt. He's a Frenchman, played in the English Premier League. I call these distractors. A Māori would see something like this they'll get mad straight away and they'll get put off straight away and you wonder why no Māori play soccer.

I've got three names now: learning humps, distractors or noise. Māori learners will not move on or do anything until they can get past the noise. When there are stimulae, questions or something inherently wrong or un-Māori in the questions or content that they consider to be ridiculous, they will not participate and this is not just for Māori, this is most probably any culture, any person. So lessons—be careful when trying to add Māori context, sometimes you're digging a bigger hole.

Kina—like I said, the greatest form of currency on the East Coast. One of the questions asks you to calculate how to share 12 kina among three whānau. A Māori will not answer this question. They will look at you and wonder, that's like sharing our 12 pebbles to three families and each family has 20 people in it. So it's just some of those, we've just got to be mindful of these things. Perhaps if you see this, you won't even bother carrying on because it's wrong or you'll move heaven and earth just to try and find out who made that mistake so they can fix it. And of course the next question in the list was about pipi and they're even smaller than kina. But we had great fun with this. I did anyway. The other people were kina connoisseurs and they could not get over it for a week.

Another part we contributed was the word bank. They gave us a list of 3,000 words and we went through them to see if there are any words that were stale or not culturally appropriate for Māori. And there were a lot of old English words there—imperial, like really old. Māori have never seen any of these things unless they had been to London in the 1860s or something. So those got the scratch. And then you also need to be mindful of words that make sense in both languages and be aware that Māori may think you mean the Māori word. For example, "take". Māori, straight away, they're gonna say "take" but of course, if it's written in the English word bank it means "take". Māori's going to fail. How come there are no Māori words in the word bank? These words are in the dictionary. Mana is used every day in the newspaper—any great rugby player has got a great mana nowadays. Haka kai whānau. Also there are some foreign animals on the list. So if you're like Tama Iti and you're a terrorist in the middle of the Uruwera, there aren't too many hippo or hyenas. And, again, we've got the dual languages, words that have an English and Māori meaning, such as "tame", "pure" and "take".

Māori learners like all learners had different modes. Many are visual or visual helps them. So some of the images that Aroha talked about—they were bad-quality pictures. Unfortunately, I didn't bring any audio clips but just also be aware when doing listening tests, accents, tone, colour, pace, pitch. Sometimes all people, all cultures, have difficulties in understand the pronunciation of words by certain people. I mean, imagine a person like Bill McLaren reading out a listening test and think if Boy the movie character can understand him.

I suppose the main message I want to get across is again thanking NZCER for letting us participate in the process, having some say. It was like the rugby analogy—instead of standing in the stands we got to play on the rugby field and every Māori loves to play. We don't want to be on the bench. We didn't come here to call the assessment tool useless and rubbish it. Quite the contrary. We came here to help so that it can be a good tool because we're trialling it and we want to use it. The Learning Progressions tell you to know your learner. Even though they tell you to know where your learner is at in their learning, you've got to actually know them. You've got to drive down their street. You've got to wear their shoes. That is the primary tikanga Māori whenever a Māori meets another Māori—who are you? And I don't mean just you. Who are you? Who is everyone behind you? Where are you from? What are your interests? And of course you have to tell them who you are.

5. Synthesis of the day

Peter Coolbear and Bronwyn Yates

Peter Coolbear is the Director of Ako Aotearoa, the National Centre for Tertiary Teaching Excellence.

Bronwyn Yates is the Chief Executive, Tumuaki, of Literacy Aotearoa.

Peter Coolbear

Kia ora Koutou. When Jenny first asked me to be on this summary panel, my first response was, "Ask somebody else because I don't know much at all about assessing literacy and numeracy."

So what you're going to hear from me are some general thoughts about assessment of adult learning and a bit about what I've learnt or what I've half learnt today as I have been thinking through things.

It's been a really absorbing day and I have learnt a lot. I think it was great to start with a panel talking about practice because ultimately that's where you've got to start.

I've got a list of thoughts and questions here and the first one for me is how do we actually move this debate forward? We started in Rose's excellent presentation from a point I thought we'd got over about 30 years ago, about the transmission model of education, yet we still seem to use that as a starting point for the debate. Why do we do that? Because I think we've still got a lot of teachers in the tertiary sector who lapse into that as a default mechanism. So if nothing else, having debates like this starts us thinking beyond that. That's one part of a kind of system context we have to work in. You know, we have these sessions, we have great discussion, great debate. The big challenge for us is how we get it beyond the enthusiasts, how we get it beyond the people who are prepared to come for a day and talk about this. So that's a challenge for Ako Aotearoa that we're trying to work on.

The second thing for me is you can't ignore the political context in these debates and we're in a time here where everybody is talking about value for money. Government is talking about value for money, learners are talking about value for money, communities and whānau are talking about value for money. What's it going to cost me to participate in tertiary education? Why am I doing it? What am I going to get out of it? And some of the statistics—and Dave alluded to one or two this morning—are really quite worrying. If you look at this from a return on investment by New Zealand society, look at the number of people who are studying Level 1 to Level 3 qualifications. For the year 2000, data from the Ministry of Education for people who started studying at Level 1 to Level 3 in a tertiary education institution or in a private training establishment (PTE), show there are about 50,000 of them that year. What have they done in the last seven years? That's measured up to the end of 2006, so seven possible years of study. For every 1,000 who enrolled in 2000, only 300 successfully completed the qualification for which they enrolled. Worse than that, it seems that only 10 percent of that 300 have subsequently successfully completed a higher-level qualification at Level 4 or above. Something's wrong somewhere. These were some of the things the afternoon's presentations addressed specifically. We've got to pay attention to those statistics. They are so

bad I don't actually believe them. But, we've checked with the Ministry of Education and they're really worrying.

We talk about Levels 1 and 3 as being entries into participating in New Zealand society, the New Zealand economy, and yet what we're doing at the moment is failing most of those people who actually want to take the chance to do it. So there are some real concerns there. For that reason alone the work on literacy and numeracy, progressions and assessment is hugely important. It's looking to see how we can begin to address that problem. It will only be part of the solution but it will be an important part of the solution.

The other thing that I've taken from today is that assessment strategies need to be purposeful and fit for purpose. One thing I didn't think was emphasised enough is that part of getting that fitness for purpose is that the learners involved need to understand what that purpose is and ideally have a say in the way they're assessed. Now, we touched on the learner voice occasionally but I don't think that discussion got very far. You've got to remember that I went to university in the early 1970s in the UK: we got whole curricula changed, never mind the odd assessment.

We need to realise there's a huge risk around any assessment strategy that it becomes an instrument of control and that's a control at the system level, an instrument of colonisation, an instrument of controlling the individual and making them compliant with what's happening in their particular institution. So we've got to be very wary of that. On the other hand, when we talk about assessments and assessment strategies, sometimes I think we start to get into the argument of trying to make assessment do too much. Sometimes we talk about assessment as though it should have multiple purposes and do everything—measure soft skills as well as practical skills, as well as knowledge and understanding. Perhaps sometimes we have to say this is too complex to be dealt with by one instrument. Maybe we should be less ambitious about any assessment strategy as well.

I have two final things. One is a question really that I think I'd like to talk about with NZCER another time, and I think we need to debate it at Ako Aotearoa as well. I get the sense that the definition of "competence" is slipping around and changing on me quite a lot at the moment. Competence as was defined in unit standards in the early 1990s is not the same as how we talk about competence now, yet nobody has really explained to me what the difference is. And that's a worry and I think we need to pay attention to that.

And the other one is we've got a huge challenge ahead of us very shortly. We've got the New Zealand Qualifications Authority targeted review of qualifications which is looking to reduce the proliferation of qualifications at pre-degree level, at certificate and diploma level. We are talking about a new model—which I actually believe in and fully support—that allows multiple programmes to lead to the same qualification. If ever there was a challenge for assessment, that's it. We're going to have to start moderating between different assessment strategies, between unit standards-based programmes and between programmes that are assessed in other ways. That's going to be an enormous challenge. We've got a huge moderation issue to deal with there.

So put that all together and the only conclusion I can come to is we need more days like this: thank you very much NZCER for organising it.

Bronwyn Yates

E nga mana, e nga reo, rau rangatira ma, tēnā koutou katoa.

Today has been a fantastic opportunity to listen and learn, and for me (and possibly others) most of all to think, reflect and question. At the end of the day the valuable input of others leads me to recognise the care and brilliance that has led to the development and implementation of the adult literacy strategy and the corresponding infrastructure.

It's also led me to question what the BIG GAME PLAN is.

Young's *Framework for Analysis* notes that in the consideration of policy there are some essential questions to ask, including:

- Who is this policy [strategy] for and for what purpose?
- Who benefits from this policy [strategy] and for what purpose?

The presentations have led me to seek further clarification about the national adult literacy strategy. Is it a strategy to:

- create a literate New Zealand?
- only increase productivity?

Furthermore, will the strategy be based on models of teaching that:

- are autonomous or ideological?
- are about banking or transformative education?

And does the strategy aim to reflect the power of "and", with the aim to benefit all. And if this is the case, how will the current strategy achieve this, and how much will that really cost?

A key recurring theme of today has been that "assessment must value the learner". Everyone agrees that appropriate assessment methods and tools are critical to effective teaching—they need to be diverse and they need to be useful to match learners' needs and circumstances. I found myself asking today: "Where is the learner's voice?" And it led me to wonder whether we need to be more open to seek advice and input from literacy learners about the current assessment resources and any future developments.

I believe it would also be worthwhile to consider whether it is useful to create the connections to those aspects of literacy learning and use that we have yet to connect assessment to—that is, the "soft" outcomes; the critical differences on people's lives; the impacts on intergenerational issues. We may decide not to pursue how these connections are measured, but I believe exploration of these outcomes will increase our knowledge about what is meant by critical literacy learning. When we do this, it will be essential to have the conversation with students and LLN practitioners.

Hopefully we can build on the NIACE study of Catching Confidence—adapting to "I CAN" statements, such as I can read to my children; I can fill out the forms at work; I can drive my car legally; I can take messages on the phone; I can make conversations at social events; and I can make a difference to not only my life, but to others'—perhaps establishing a new process for self-assessment, that leads to an outcomes-based performance framework. This framework, if based on individual statements about what students can do, can then be culturally located, contextualised and student-centred. But, however we create this new

framework, it will need to be computer-collatable, so that we can crunch the numbers and from the data, get a closer feel of where we take ourselves as learners and where we take ourselves as practitioners.

We have a rich history of adult literacy in this country and internationally. We add to the names of Freire, Street and O'Neill—Wally Penetito, Colin Lankshear, Kathy Irwin and Kuni Jenkins to name only a few of New Zealand educationalists who have informed and discussed adult literacy issues in this country for many years before recent times.

So, where to from here? I believe it is time to grow an outcomes-based performance framework. I think we can recognise that assessment in itself is a progression and so the development of the tool and other assessment processes should be ongoing. I look forward to further debate and growth of our knowledge of how to honour learners' progress and achievement in ways that continue to be uniquely right for Aotearoa and all our diversities.

Kia ora ki a koutou katoa.