

Trends in assessment: Five big questions and associated dilemmas where research insights could be useful

Rosemary Hipkins and Marie Cameron

In mid-2018, the Ministry of Education (MOE) commissioned the New Zealand Council for Educational Research (NZCER) to scan national and international literature for recent trends in assessment policy and practice. MOE expressed an interest in any evidence, or arguments grounded in evidence, that the national assessment principles might need to be updated, including consideration of whether any new principles should be added, especially given the rapid evolution of digital technologies in the intervening years since 2011 when the current assessment policy was published.

Specific questions for the review included consideration of:

- how teachers are using assessment to support teaching and learning in New Zealand, with a particular focus on the compulsory schools sector;
- any shifts in the ways educators are using assessment for learning; the system conditions necessary to support effective assessment for learning practices;
- the extent to which key competencies/skills for global citizenship are being assessed and how they are being assessed;
- the extent to which digital technologies are being used to personalise learning and give students, teachers, parents and whānau rich information about learning; and use of effective strategies that engage parents and whānau and the wider community about different approaches to assessing and credentialing achievement.

Combinations of key words from these questions were used in the literature search. Over 200 individual items were added to the Zotero data base we created for the project.

Our findings in relation to the questions just outlined are addressed in the full report, due for release in October 2018. This document is not a summary of that report, nor are references included. Our aim was to write a series of discussion prompts for the NZARE SIG day in September 2018. To this end, we foreground interesting research questions and dilemmas, which we have organised under five overarching questions.

1. What should we do about growing assessment capability within our education system?

The argument that we need to build assessment capability at all levels of our education system was central to the Directions for Assessment in New Zealand (DANZ) report commissioned to inform development of the current assessment policy. It was subsequently picked up in two of the six principles in the 2011 national assessment policy:

- Building assessment capability is crucial to achieving improvement
- An assessment capable system is an accountable system

Only a small amount of research addresses the idea of assessment capability per se, but plenty addresses assessment for learning (AfL). We found we needed to clarify the relationship between these sets of ideas: broadly, assessment capabilities are those that are needed to enact AfL pedagogies.

The centrality of AfL to an assessment system fit for the 21st century is a very clear theme in the literature, along with calls for changes that will support more consistent and sustainable AfL practices. Yet there is a considerable body of international research which suggests that creating and sustaining effective AfL pedagogies is easier said than done. Quite a lot is also known about potential reasons but there is still plenty of scope for local research.

Involvement of students in assessment decision-making: building *students'* assessment capabilities was central to the DANZ argument, where it was linked to curriculum challenges such as fostering agency and building students' learning-to-learn capabilities. But the available research evidence (and there is quite a lot of it) points to a need to better understand why this also seems to be easier said than done. What is getting in the way? One possible answer might relate to a lack of clear progress indicators in many learning areas—this is addressed next. There also seems to be a need to understand more about teachers' sense-making when enacting assessment in the classroom. We are aware that this is an area where international research involving New Zealand as a partner is underway. If this is correct, (how) could we leverage this research effort to address the challenge more widely within our education system as a whole?

An ongoing need to build teachers' data literacy: there has been quite a lot of local research activity in this space especially, but not only, in ITE contexts. (One dilemma here is that beginning teachers need the support of skilled mentors to further build and strengthen their fledgling assessment pedagogies). The general thrust of this body of work suggests that there has been no overall improvement to teachers' data literacy since the assessment standards were published in 2011, and this remains an area of need for professional learning and support. One recent report says that that lack of assessment literacy is preventing teams from making more of enquiry projects funded by TLIF. Would it be fair to assume that Kāhui Ako will face the same challenges? How would we know? What sort of research evidence would be most helpful if we really want to gain traction on what currently looks to be a fairly intractable challenge?

Washback from high-stakes assessments: there are clear indications in the international literature—and in local commentary about NCEA and National Standards—that accountability trumps AfL when the two are seen to be in tension. The need for accountability isn't going to go away, so how might these tensions be managed in ways that allow AfL practices to thrive? What might the research community contribute to resolving this dilemma?

Access to resources: some systems (e.g. a recent review of Ontario's assessment system) pin hopes—at least partially—on provision of more/better resources. Is this an area where we need to know more? We have very little systematic evidence on appropriateness and actual use of resources we already have (e.g. assessment resource banks; TKI assessment pages; PACT resources, and now the Coherent Pathways tool for Kāhui Ako). Could links between an AfL focus and these resources be developed or strengthened? What else might we need to know to make the case for strengthening support for their effective use?

2. How should we respond to the shift from reporting point-in-time achievement to reporting progress made over time?

The clear international trend is away from an emphasis in summative reporting of point-in-time achievement towards reporting of progress made over time. This emphasis is evident in the so-called “Gonski 2” report to the Australian government. It is being supported and socialised by researchers from ACER, but also has its critics. Here in NZ, the development of the digital technologies curriculum also reflects the trend, with “indicators of progress” replacing outcomes as the organisers of content at the different curriculum levels. The drivers behind this trend each bring interesting potential research questions along with interesting dilemmas.

Equity: students learn at different rates, in different ways, etc. Reporting of progress allows the learning gains of every student to be acknowledged. An implication is that this will be more motivating—critics of the Gonski 2 report say this is an untested assumption. How could we test it, should we decide to try and address this challenge?

Reporting to parents: the concept of progress is easy to understand. Models of progression are needed to underpin use of digital technologies for reporting meaningful learning gains, e.g. by use of dashboards etc. There are indications that SMS currently in use in NZ are *not* configured in ways that readily allow actual progress to be reported. But we found little systematic data about use of SMS and what might need to change. This challenge points to similar questions to those re personalized assessment (see below).

Reflecting NZC: in particular we know very little about how students actually make progress in developing more complex learning outcomes being signaled in recent curriculum development work (both in NZ and internationally). Researchers at ACER describe a “chicken and egg” dilemma where researchers need to work in classrooms to collect evidence of students’ capabilities, but teachers won’t focus on these until there is evidence of what they should be working towards (i.e. clear progressions). The Coherent Pathways tool makes a first attempt to address this dilemma but those involved in its development feely admit that it is a work in progress. Could our research community make a contribution in this space? What might we focus on and why?

Building assessment capabilities: in particular, involving students in assessing their own work might be better supported if teachers had better access to likely trajectories of learning progress. Again, this surfaces questions about priorities and resourcing. Would we want something similar to the level of detail in the PACT progressions for every learning area? Assuming it was even possible to do this (bearing in mind the chicken and egg dilemma) what would be the pros and cons and how should we balance these? How do teachers use the progressions we already have? Could timely research insights help make the use of them more effective?

Teacher professional learning and ownership: The most sophisticated progressions possible won’t achieve their intended benefits (e.g. building students’ assessment capabilities, equity, clarity for parents and whanau) if teachers won’t, or don’t how to, use them as intended. Some jurisdictions have addressed this ownership challenge by making teachers central to the process of building new progressions. There is a clear risk here—semantic incrementalism can all too readily creep in. (A more sophisticated process was used to build the NZ digital technology progressions.) The literature is clear that moderation of assessment tasks can be helpful for building shared understanding and ownership. Yet we have almost no local moderation research. Why not? Should we? (Making OTJs

for National Standards is an exception to this comment. But the message from a range of research projects only underscores the concern. The evidence suggests that teachers did not do this well, or consistently.)

3. When and where (if at all) do our current assessment practices reflect the richness of NZC?

Reflecting NZC has already been identified above as a challenge. Meaningful integration between the back and front halves of NZC is one important contributor to this dilemma. (The “more complex” outcomes now understood to be needed have already been mentioned.) But the major factor is that assessment conveys strong “washback” effects that tend to determine curriculum choices—the cart is in front of the horse. This is a dilemma internationally. Notwithstanding different assessment systems, the “how to” dilemmas that follow are common.

How to assess key competencies: (or their equivalents elsewhere). We found plenty of instances of question-asking, but no convincing answers as yet. Some research is underway in Australia, using problem-based learning as the opportunity that allows students to demonstrate a complex of interrelated capabilities. Whether the rubrics they intend to design will be workable and adequate to the challenge remains to be seen. Would similar case studies in NZ be useful?

How to carry out robust performance assessments: the literature is clear that performance assessment should be used when assessing students’ competencies/capabilities to show their learning. However the term appears to be somewhat taken for granted. What is in scope and what is not? Is this something many teachers are already doing well in NZ—but we just don’t call it that? Would it be productive to develop exemplars for professional discussion? One international review suggests that use of performance assessments, with associated moderation conversations, strengthens teaching and learning and hence lifts achievement.

How to allow for emergent outcomes: this is an interesting dilemma where questions are being asked but we found no answers as yet. How do we know what students might be able to do if we have never documented certain types of outcomes in traditional assessments? (This is one half of the ACER “chicken and egg” dilemma.) Emergence—or at the very least unpredictability—is possible in a number of ways: seeking complex competency-based outcomes as already noted; seeking outcomes that emerge at the intersection of different disciplines when integrating the curriculum; and seeking deeper “knowledge” outcomes such as conceptual coherence.

How to make best use of micro credentials: these are coming ready or not! NZQA has just introduced guidelines for the tertiary sector. Research and critique of this rapidly growing field has strong messages about the design of microcredentialling systems, and the risks to credibility if multiple different systems proliferate. Stackability is identified as important (sets of microcredentials purposefully build to a bigger whole). Can our research community make a contribution here? Where and how are microcredentials already being offered in NZ? How, if at all, are issues of credibility and rigour being addressed? Do we need a degree of regulation? If so, what would it look like? What if we thought about NCEA as a system of microcredentials—could that change processes and supports around it?

4. Can we learn more about equity challenges in New Zealand assessment contexts?

Equity concerns have already been highlighted as one of the drivers underpinning the trend to shift from reporting point-in-time achievement to reporting on progress. The learning sciences assert that every student is capable of learning and making progress, albeit at different speeds, in different ways, and to different extents. This draws fresh attention to a complex of interrelated issues in our system.

Universal design for learning (UDL): there is advocacy for using UDL principles to design assessment and learning experiences that are more accessible to more students (e.g. a dedicated section of TKI). But we found no research-based case studies of this happening in practice. What is actually happening: when, where, and how? Would it be useful to know?

Reporting progress when students work longer term within one curriculum level: the literature is clear that all students are entitled to access the full breadth of the curriculum. How might this be assessed when students work at a slower rate than their peers? Does it matter that the traditional assessment focus for students with the most challenging learning needs has been predominantly on social progress? (One MOE funded project is underway in this area—the aim is to develop progressions with a much finer grain size, but the question of how many such progressions can be developed before teachers become over-burdened is relevant here.)

High expectations for all students: one Auckland-based team has made a systematic contribution to research in this area. Their most recently published research implies that teacher judgements are influenced by their *perceptions* of students' academic ability (with indications that this might be influenced by their family backgrounds and where they go to school). Is this a concern that we need to know more about? Could more of us be more proactive in this space?

The role of assessments in constructing inequalities: this is a lively topic in critiques of international assessments such as PISA (which we ultimately decided was outside the scope of our review). We found little recent research or critique that has directly addressed equity impacts of actual assessments in the New Zealand context. (One very thorough paper was contributed to the DANZ process and is on TKI.) What might new research look like? Who might be best placed to actually do it? Should we/ how could we, as an assessment community, help shift public understanding that “one size fits all” is the fairest way to assess and report on learning and achievement?

Finally, here is a more speculative worry, albeit informed by some NCEA research, where there is “parity of esteem” for all types of learning. Is there an equity risk in the move to progressions? What if there was a degree of complacency when students were seen to be making progress, but nevertheless remained well behind their peers? Would that matter? Why or why not? How would we even know? What sort of research might be helpful in this space?

5. How—and to what extent—is personalized assessment happening?

The assumption that we should aim for active involvement of every student in the assessment of their own learning progress is associated with most of the ideas we have already addressed: building assessment capability; reporting individual progress over time; striving for greater equity in our assessment system; assessing students' demonstrated capabilities—what they can do with their learning—and so on. With these connections in mind, our final question addresses the personalization of assessment. This theme is especially evident in assessment research set in the context of digital technologies.

What does personalized assessment *practice* look like? This idea is closely related to personalized learning practices, with a focus on how undertaking the act of assessment is supported for each student. There is a gap in rich case studies in the literature. Research in this space tends to use standardized assessments to measure the impacts of personalized assessment practices (so that treatment and control groups undertake the same actual assessment). We think this is an interesting dilemma but does it actually matter? Why or why not?

What do we mean by personalized assessment? In the digital technology context there appear to be a continuum of ways to personalize the learner's actual assessment experience. These range from computer adaptive testing to fully embedded (stealth) assessment in which there is no evident separation between assessment and learning. How much of this type of assessment would we like to see in our schools and in what curriculum contexts? Who should choose? Do we need an element of regulation, given the high costs and diverse offerings? There are concerns about commercial products that assess "nicely packaged knowledge" at the expense of building knowledge that reflects the depth of the curriculum, and of learning interactions with other students. There are also equity concerns, specifically the high likelihood of unequal access to rapidly evolving digital learning and assessment resources (and personalized neuro-therapies at the cutting edge of commercial offerings).

How should we prioritize? The literature is clear that this is an area where digital technologies will soon have a big impact. For example, game-based learning environments "assess" as the learner interacts inside the resource. So-called stealth assessments can collate and report formative assessment data to the teacher. Researchers in this space emphasize the need for rigorous design—new validity and reliability issues have been a focus of careful research. This makes such resources expensive to develop and test. Should we let choices simply unfold (knowing that there will almost certainly be vested commercial interests in play) or should we be proactive in ensuring these opportunities are used in fair and appropriate ways? If the latter, what might the research community do to help?