CONCORE TŪHURA

Assessment and Consistency Measures

Good practice in the development and implementation of skill standards-based qualifications

Karen Vaughan and Andrew Kear | September 2024

Hummingbird EFFECT

With the support of

WAIHANGA ARA RAU
Construction and Infrastructure
Workforce Development Council

NZQA SUPPORT FOR THE GUIDES

NZQA supports ConCOVE Tūhura's approach in developing these guides to help standard setting bodies and end-users develop a deeper understanding of skill standards.

As the building blocks of vocational qualifications and micro-credentials, skill standards have huge potential to support consistent graduate outcomes and meet industry needs. We envisage that the toolkit approach to the good practice guides will be particularly useful to the target audiences, some of whom may be new to standards-based qualifications and programmes. – NZQA

ACKNOWLEDGEMENT

This guide was commissioned by ConCOVE Tūhura and part-funded by Waihanga Ara Rau Workforce Development Council. The work was delivered by Hummingbird Effect.

We extend our sincere thanks to the many individuals and organisations who contributed their time, expertise, and insights to the development of this guide, including:

- Waihanga Ara Rau Workforce Development Council
- Building and Construction Industry Training Organisation (BCITO)
- Te Pūkenga
- Apprenticeship Training New Zealand (ATNZ)
- Eastern Institute of Technology (EIT)
- Competenz
- Ara Institute of Canterbury
- Site Safe New Zealand Incorporated
- New Zealand Qualifications Authority (NZQA) Policy Team

Technical Advisory Groups for:

- Rigging
- Core Construction
- Temporary Traffic Management
- Glazing
- Painting
- Structural Detailing
- Ringa Hora Services Workforce Development Council
- Toi Mai Workforce Development Council

Their collective knowledge and experience have been invaluable in shaping this resource to support assessment practice in vocational education and training.

CONTENTS

AN INTRODUCTION TO THIS RESOURCE

Good practice toolkit

In late 2023, skill standards began replacing unit standards, becoming compulsory components, where they exist, of the programmes leading to New Zealand national qualifications. The first skill standards were approved and published by the New Zealand Qualifications Authority in 2024. National qualifications and programmes of learning developed from 2024 onwards will be based on skill standards.

This is the introduction to the 'toolkit' of good practice guides. The toolkit audience is those involved in the process of developing and implementing skill standards-based credentials and qualifications:

1. standard setting and qualifications development;
2. learning programme development;
3. learning programme delivery;
4. assessment of learning; and
5. moderation of outcomes.

These roles sit within Work Based Learning Organisations (WBLOs) , Institutes of Technology and Polytechnics (ITPs), Private Training Establishments (PTEs) and wānanga for programme development and delivery (including assessment); Standard-setting Bodies (SSBs) for standards and qualifications development and moderation of outcomes; and industry and other stakeholders for advice on qualifications and programmes.

There are six guides in the toolkit:

1. A Background to the Emergence of Skill Standards
2. An Overview of the System
3. Standards and Qualifications Development
4. Programme Development and Delivery
5. Assessment and Consistency Measures
6. Industry Stakeholders and Advisory Work

The guides were developed with research and input from sector entities and teams. Each guide discusses the most challenging issues and sets out guiding principles, illustrated with practice interpretations. It is designed to help standard-setting bodies (SSBs) and tertiary education providers deepen their understanding of what skills standards should be and do in order to achieve the quality and consistency that will meet the needs of industry and learners.

FOCUS AND AUDIENCE

This guide focuses on what is involved in assessing the learning outcomes in skill standards and assuring the consistency of those outcomes (i.e. moderation). The intended audience is:

- Assessors in work-based, online and distance, or provider-based settings; and
- Quality assurance or moderation staff in SSBs.

The guide is also useful for those working in other parts of the vocational education system, whose work sits 'upstream' or 'downstream' of the main audience. For example, those who:

- Develop qualifications and standards
- Provide qualifications or learning programme developers with subject matter expertise as industry or stakeholder representatives
- Implement qualifications by developing learning programmes and/or delivering through workplaces or provider-based settings
- Have an interest in the vocational education system.

HOW TO USE THIS GUIDE

We recommend that the following NZQA documents (or any updated versions of these) are referred to alongside this guidance.

1. NZQA's Guidelines for listing skill standards on the Directory of Assessment and Skill Standards.[1] We refer to this document as the NZQA Guidelines.
2. NZQA's Aromatawai and the principles of assessment[2] which explores culturally sensitive assessment practice.

[1] New Zealand Qualifications Authority, 'Guidelines for Listing Skill Standards on the Directory of Assessment and Skill Standards' (Wellington: New Zealand Qualifications Authority, August 2024), https://www2.nzqa.govt.nz/assets/Tertiary/Approval-accreditation-and-registration/Standards/Skills-standards/Guidelines-for-listing-skill-standards-on-the-DASS.pdf.

[2] New Zealand Qualifications Authority, 'Aromatawai and the Principles of Assessment Supporting Aromatawai and the Development of Quality Assessment Practices' (Wellington: New Zealand Qualifications Authority, August 2022), https://www2.nzqa.govt.nz/assets/About-us/News/aromatawai-and-the-principles-of-assessment.pdf.

SKILL STANDARDS AND THE PRINCIPLES OF GOOD ASSESSMENT PRACTICE

All structures and practices associated with the assessment of learning contain a dual possibility. They can constrain learning by directing attention to trivial credit accumulation or they can support (deeper) learning by helping make it meaningful in relation to meeting real challenges.[3]

The emergence of skill standards does not introduce new forms of assessment. However, their form can be understood as more encouraging of good practice in assessment. Standards-based assessment, and the topic of this guide, may be new to some tertiary education providers and/or assessors. For everyone, regardless of experience, there is an opportunity to reconsider and up one's game in this area.

Regardless of the tools used, educational assessment that demonstrates good practice forms an element of a system that supports and respects learners and meets industry needs. Such a system:

- Recognises and addresses learners' individual values, histories and needs
- Recognises and respects learners' pre-existing knowledge and skills
- Recognises and fosters learner progression and motivation
- Ensures access to learning and assessment opportunities
- Reduces barriers to success
- Provides learning resources and assessment activities that are appropriate and fit-for-purpose

Institutional realities regularly come into conflict with these principles. Good practice seeks to support them. For example, standards-based assessment prompts a focus on learners as individuals that may clash with a cohort approach. Standards-based assessment reframes one cohort of 30 learners as 30 cohorts of one learner. This is because each learner develops competence at different rates (not in lockstep). Therefore, it is good practice to assess via naturally occurring evidence rather than testing everyone at the same set time. Assessing individuals should of course be balanced with attending to the needs of groups (e.g. underserved learners and those directly impacted by inequities in the system). Notwithstanding this, some trainers, tutors, teachers and assessors may find a standards-based approach challenging if they are not familiar with it or if operating within business models that prioritise through-put of cohorts.

Equally, the institutional context can discourage the concept of assessment as a process and instead focus on it as an event (e.g. a test). Scheduling a test is much more straightforward administratively than observing, recognising and recording individual achievement where and when it occurs. Providing learners with many opportunities to demonstrate what they have learned can interfere with the flow of the course presentation. It may even be regarded by some trainers, tutors, teachers and assessors as an administrative nuisance that increases workload, or something that imposes perceptional challenges (e.g. in settings or organisations that routinely promote tests or exams as more rigorous, regardless of what needs to be assessed).

Ensure assessment is transparent

It is sound practice to ensure that learners know what they are going to be assessed on, when and by what method or methods. In many cases this requirement has been satisfied by giving learners the unit standard and the dates of the tests.

Learners are not the intended audience for assessment standards for several reasons. First, the language of standards may be at an inappropriate level. Second, standards also inevitably contain information that is irrelevant or unclear to learners. Third, it is the job of the course or programme design, and those associated with delivering it, to provide clear and appropriate information to learners. The goal of transparency is not well served by using the standards themselves to bypass this responsibility.

[3] Vaughan, K., & Cameron, M. (2009). Assessment of Learning in the Workplace: A Background Paper. Ako Aotearoa.

Ensure assessment is valid

Ensuring valid assessment means assessing what we say we are assessing.[4] As an extreme example, testing someone's capability to hang wallpaper by asking them to write about it in a worksheet or test is clearly not valid because it does not actually test the hanging of wallpaper. Instead, the assessor needs to observe the learner hanging paper, or at the very least view the finished product. Assessors should also receive formal reassurances that the work is the learner's own, probably supplemented by some discussion with a supervisor about aspects of productivity such as wastage or timeliness.

This example begs the question about whether it's valid to test the learner's understanding of the process of paperhanging by a written test. For example, this might be a test that asks things like: "In what order do you undertake the various tasks and how do you decide that? What are the various processes involved? What are the different types of paper and adhesive, and what are the advantages and pitfalls with each?".

However, the answer to the question of how to test a learner's understanding is more complex than it looks. For example, is it that we are really testing the learner's competence in reading and writing? Depending on the circumstances, it could be their computer literacy we're assessing, or possibly their imagination. Possibly we are inadvertently testing their vocabulary. In the example used above, are we sure that our learners all know what a 'pitfall' is?

The best assessment is the simplest assessment

In an ideal world a single assessor would have time to sit with an individual learner and say, "Tell me everything you know about laying bricks" and then ask clarifying questions throughout to ascertain the learner's understanding. On site, it might go like this: "Show me what you're doing and explain what you're doing and why." Allowing learners to choose their own words and their own sequence of explanation pays dividends in terms of validity: the assessment process hasn't contaminated the way the learner thinks about it or explains it.

Of course it's not an ideal world. However, the 'simplest assessment' rule of thumb is a valuable guide that should underpin all assessment planning.

Opportunities are not rationed

Good practice assessment is best regarded as a process rather than as an event (or series of events). Learners who know (or can do) things before the 'test date' shouldn't be held back by the timing of the event. Likewise, some learners benefit from the assessment process and additional learning, and further opportunities to demonstrate competence should be extended to them. Offering more opportunities to succeed can help learners strengthen their self-efficacy (self-belief that in turn improves performance)[5] if accompanied by an approach that emphasises learning and effort to overcome challenges (rather than innate ability and a 'fixed mindset').[6]

Learning is a continuous process. Each learner's journey through the process is unique. To the extent that it is possible, assessment should match those realities.

[4] Validity can be further broken down into ensuring content validity (the skills being assessed); criterion validity (how performance will be judged); construct validity (overall method of assessing); and consequential validity (what happens as a result of the assessment). See Shelley Gillis and Andrea Bateman, 'Assessing in VET: Issues of Reliability and Validity' (Adelaide: NCVER, 1999).

[5] Albert Bandura, 'Self-Efficacy: Toward a Unifying Theory of Behavioral Change', Advances in Behaviour Research and Therapy 1, no. 4 (1978): 139–61.

[6] Carol S. Dweck, Mindset. How You Can Fulfill Your Potential (Great Britain: Robinson, 2012).

Evidence is gathered where it occurs

If good practice assessment is a process then it follows that evidence of achievement that is observed outside the 'test' cycle should be duly reported and used to form final judgements. That is, assessors should use 'naturally occurring evidence' to support their decisions, arising out of the learner's programme or everyday work. This doesn't mean that evidence emerges spontaneously and magically; opportunities for learning (not simply rote task completion) still need to be designed.[7] But the designed opportunity should not require a separate test as well because it should nest within the everyday work context (or within a learning context that simulates a work context).

The terms 'formative assessment' (assessment for learning) and 'summative assessment' (assessment of learning) have become relatively commonplace. It can be useful to distinguish between the different purposes of assessment – e.g. occurring 'during' learning processes for the purposes of making learning explicit and deepening it (formative) and at the 'end' of learning for the purposes of recognising and reporting achievement at a moment in time (summative). However, that distinction has not always assisted in supporting the principle of naturally occurring evidence. The terminology has potential to encourage assessors to impose an unsophisticated discrimination between formative and summative assessment.

This has led some assessors to ignore, or downplay, and/or to fail to record what is undeniably valid and appropriate evidence of achievement against a standard. This is often done on the basis that it wasn't what they were looking for on that occasion or based on the perception that the learner is "not supposed to be ready for that yet". Perhaps it is that the assessor or the organisation has no administrative process for recording out-of-sequence evidence.

In all cases it makes good sense to think of assessment of an individual's progress, seeking evidence of achievement whenever and wherever it occurs. This does require sophisticated skills in assessing.[8] There are examples by which assessors and assessment can be professionalised[9], and the system can be strengthened with a focus on professional conversation for learning within an assessment 'community of practice'.[10] Professional development for assessment capability is further discussed in the Programme Development and Delivery guide.

Inference as an assessment tool

The concept of naturally occurring evidence supports the assessor's exercise of inference as a legitimate means of assessment. The assessor who asked the apprentice brick-layer, "Show me what you're doing and explain what you're doing and why" was overtly using the practical exercise to assess knowledge and understanding. There was no need for a separate 'theory test'.

Equally legitimate, in some contexts, is where an assessor sees the practical work and is able to infer the knowledge without further questioning. "You couldn't have done that job if you didn't know how." Each vocation has its examples of this principle. In Carpentry it might be scribing corner joints in complex mouldings. It is inconceivable that an apprentice could have produced a tidy, well-fitting joint without knowing how to. This makes assessment much more about judgement than measurement, as assessors use a 'web of reason' to build a complex picture of the learner's capabilities.[11]

NZQA's encouragement to SSBs to combine theory and practical in standards development "wherever possible" reinforces this aspect of good assessment practice. The Assessment Criterion "Apply active listening strategies to understand and respond to customer requirements" is an example of this approach. The assessor could administer a test that would establish the learner's understandings of active listening strategies. On the other hand, the wording of the Assessment Criterion specifically encourages assessment that infers the knowledge from its application through practical demonstration.

In some cases, the need to infer (or test for) the existence of knowledge is hidden deeper in the skill standard. The Assessment Criterion "Components are checked and fixed using correct equipment and fixing methods" appears to be simply about the observation of a skill. Indeed, all Assessment Criteria in the standard quoted (skill standard 40000, Manufacture a timber flooring cassette) are about skill demonstration.

Nevertheless, the Indicative Content of this standard refers to "Fixing hardware types, specifications, purpose and fixing method" and "Glue types, purpose and application method". These make it clear that the course of learning is intended to have a significant theory component and therefore that assessment judgements must draw out that knowledge. We discuss Indicative Content and its challenges in more detail later in this guide.

Assessment is a specialist capability

There is a recurrent discussion about the capability of assessors that suggests technical, or subject matter expertise is more important than education and assessment expertise. This sidesteps the crucial point: it is vocational education. While the needs of industry drive much of vocational education's programme offerings, training and assessing learners is primarily an educational endeavour. Good vocational education and good learning outcomes demand good teaching and assessment. This is especially so in the standards- (or competency-) based model, which moves beyond a time-served model of master-apprentice craft with tacit teaching.[12]

Training people of course requires technical or subject matter expertise. Nobody wants to see, for example, flooring apprentices trained by people who are not deeply experienced in doing flooring work. However, the capability to train others well is not inherent by virtue of having technical expertise. It is not uncommon for technical experts to struggle with teaching. It is hard to remember what it is like to not know something once you yourself know it. Teaching involves a different set of skills. It means someone can thoughtfully utilise strategies beyond simply reproducing their own experience of learning (for better or worse).

When it comes to assessment, often carried out by someone other than the trainer in workplace-based learning contexts, education (assessment) expertise is as much as a priority as technical or subject matter expertise because assessment is about interpretation of standards and qualifications.[13] Assessors need to be assessing at the level of the industry's national standard. They can consult with the trainer and any other workplace evaluator on any technical or subject matter issues. It is most ideal if assessors have assessment expertise and technical knowledge or at the very least a strong affinity for it alongside others with technical knowledge that they can consult.

That assessment requires education expertise could be, and regularly has been, overlooked where assessment was designed directly off individual unit standards. Assessors could then take a mechanistic, event-based approach – for example, administering multi-choice tests or tick-box marking learners' worksheets against each and every (small credit-value) standard. This is not only undesirable from a learning and outcomes point of view, but it is less possible with well-developed skill standards.

Assessors need to be recruited and trained for education capability, specifically in assessment so they can gather naturally occurring evidence, make professional judgements and record the nature of evidence and the reasoning for their judgements.[14] This may be particularly challenging for workplace-based assessors (typically contracted to WBLOs) as well as for some campus-based staff (e.g. trades tutors recently 'off the tools' into tutoring roles). Their organisations need to provide them with appropriate support.

[7] Rosemary Hipkins, 'Assessment of Naturally Occurring Evidence of Literacy', Assessment Matters 4 (2012): 95–109.

[8] Selena Chan, 'New Zealand's Move to Graduate-Profile Framed Qualifications: Implications, Challenges and the Occupational Identity Solution', International Journal of Training Research 14, no. 1 (2016): 5–18.

[9] Karen Vaughan, Ben Gardiner, and Jan Eyre, 'The Transformation of Industry-Led Assessment of On-Job Learning in the Building and Construction Industries' (Wellington: Ako Aotearoa, 2012).

[10] Karen Vaughan, Andrew Kear, and Heath MacKenzie, 'Mate, You Should Know This! Re-Negotiating Practice after a Critical Incident in the Assessment of on-Job Learning', Vocations and Learning 7, no. 3 (2014): 331–44.

[11] M. E. De Vos et al., 'Unravelling Workplace Educators' Judgment Processes When Assessing Students' Performance at the Workplace', Journal of Vocational Education & Training 76, no. 3 (26 May 2024): 517–36.

[12] Jeanne Gamble, 'Why Improved Formal Teaching and Learning Are Important in Technical and Vocational Education and Training (TVET)', Revisiting Global Trends in TVET: Reflections on Theory and Practice 204 (2013).

[13] In fact for many sub-fields that have few people (e.g. specialist trades such as flooring), it is rare to have an assessor who is also a subject matter expert in that field.

[14] Karen Vaughan and Marie Cameron, 'A Guide to Good Practice in Industry Training Organisation Structures and Systems For On-Job Assessment' (Wellington: Ako Aotearoa, 2010).

Existing knowledge and skill are recognised

Learners come to new courses of study along their individual pathways and with varying skill and knowledge. As a general rule it is poor practice to make learners 'do the course' or 'sit the test' if they already know or can do the things that the course is about.

A core principle of the New Zealand Qualifications and Credentials Framework is that its Record of Achievement provides each registered learner with a permanent record of the assessment standards and the qualifications or credentials that they have achieved. Once the outcome is achieved and recorded, the individual has no need to prove competence again.

But there are inevitably complexities to this. What if an individual who wishes to sign up for a programme to achieve a qualification already has 15 of the required 50 credits? How, in a real world, does the institution concerned realistically exempt the candidate from some of the learning programme? And which parts would it be, especially when their programme does not slavishly follow the construction of the assessment standards that make up the qualification?

In a similar vein, what if it becomes clear during the course of the programme that the learner is in fact not competent in the skills or knowledge that past assessment has affirmed and reported to the Record of Achievement? Equally, some learners arrive already capable in some, many or all of the competencies involved despite not having the credit recorded yet on the Record of Achievement. How should we deal with that?

There are significant complexities, both principled and pragmatic. The rules of the prevailing funding regime frequently contribute to them. Moreover, processes to recognise, and perhaps codify, existing knowledge and skill can be disruptive, complex and costly. It may be unrealistic in many contexts to set up and operate a programme of recognition of prior learning (RPL) or recognition of current competence (RCC). Nevertheless, the shift to skill standards encourages a re-think of assessment practice which should encourage assessors to think at a more holistic level. This change should enhance RPL/RCC processes. It may particularly suit portfolios and e-portfolios as assessable evidence of learners' existing skills and knowledge as these can be aligned with graduate outcomes.[15]

Regardless of administrative challenges, the principle remains that learners should not be compelled to repeat the course – or a part of the course – that they have already done, nor to provide evidence of skills and knowledge that they already had, except in the context of a RPL or RCC evaluation.

While giving effect to these goals in the real world is a challenge for all, clear and accessible RPL processes are vital in the context of lifelong learning. More people are changing careers or retraining, reskilling and upskilling. This includes people who are participating in tertiary education for the first time or after significant periods of skill acquisition in the workforce. It includes people needing recognition for mātauranga Māori and indigenous knowledge and therefore needing providers that are sufficiently knowledgeable and skilled to undertake meaningful assessment. All these people are an important resource for regional communities where they may be employed or are themselves employers, or where there is potential for them to be attracted into those communities.

Consistency matters

Everyone values consistency of educational outcomes. Learners have a right to expect that their achievements are recognised equitably. They generally judge that issue by comparison with the experiences of others in their cohort or in their lives beyond the learning context. The public at large wants to be reassured that one graduate (an electrician, for example) has a level of skill and knowledge consistent with all stated requirements and with others in the trade. Industry bodies assume that the effort they expend in helping to develop qualifications is rewarded by the emergence from the system of consistently trained and properly qualified tradespeople. Regulatory bodies such as the Plumbers, Gasfitters & Drainlayers Board want the same certainty. Governments that provide a level of funding to the process want to know that money invested is well spent.

[15] Selena Chan, 'Recognition of Current and Prior Experience in Aotearoa New Zealand and the Role of Eportfolios', in International Handbook on Education Development in Asia-Pacific, ed. Wing On Lee et al. (Singapore: Springer Nature Singapore, 2023), 1–17.

The most reliable assessment outcomes occur when there is a single assessor. Of course this is mostly impracticable, and even individual assessors are not necessarily entirely consistent. The variables involved in humans evaluating the progress of other humans guarantee this. For example, research has shown that court judges' decisions can vary according to time of day and levels of hunger.[16] The more assessors there are, and the more organisations involved, the more room for variation beyond what might be regarded as an acceptable band of tolerance.

Given the importance of consistency, it's tempting to assume that, if every learner is given an identical test, then the outcomes will be fair and consistent. This is not necessarily so. We have already mentioned the challenges to validity when equivalent levels of literacy are assumed (which is a given in the case of any written test).

In some contexts, in Aotearoa/New Zealand it has been assumed that untrained assessors can deliver reliable outcomes as long as clear guidance is provided by way of model answers. The provision of model answers is, incidentally, a form of moderation. This approach is an effort to move some way towards the 'single assessor' model where the writer of the model answer is effectively behind the scenes in the role of remote assessor. It's a blunt instrument though because the assessment is compromised by the move from determining whether the answer given is correct, to whether or how well it complies with, or 'matches' the model answer.

Consistency (or reliability) should always be considered alongside validity. Something can be reliable or consistent (e.g. a test that is administered the same way each time or evidence that is of the same type) but not actually valid (e.g. it tests, or is evidence of, something other than the skill concerned).

A well-functioning system puts substantial resources into managing these challenges. See our later section entitled Consistency measures for a discussion of approaches to moderation.

The system supports learner progress

Principled assessment is a complex undertaking that can challenge the resources of workplaces, institutions, courses and tutors. When evidence of achievement is displayed but not recorded, or recorded but not reported, or when learners are restricted by the timing or nature of assessment opportunities, it is frequently because 'the system' is effectively withholding the opportunity for its own (possibly valid[17]) administrative purposes.

A pertinent example of this is provided by recognising the tendency that 'the system' has had to grant the role of recording and reporting learner progress to unit standard and their credit outcomes. This practice has permitted institutions to evade the responsibility of reporting in straightforward, ongoing and meaningful ways how much progress has been made or is being made.

In the world of skill standards, the potential for existing unit standards to be combined and for skill standards to become larger is likely to cast further light on this principle. The inclination to treat standards as curriculum modules to be presented one at a time, assessed at the end and reported accordingly should become less common. Institutions should have systems and processes in place to record and report progress. Tutors/trainers/teachers and assessors should be providing feedback as a matter of course. Providing encouragement or feedback is not the responsibility of assessment standards.

Here we can build on the earlier discussion of formative and summative assessment in relation to the principle that evidence is gathered where it occurs. Learner progress can be fostered by feedback at all stages of learning, including in how assessment of competence is carried out (e.g. transparently, validly, via naturally occurring evidence).

[16] Glöckner A. The irrational hungry judge effect revisited: Simulations reveal that the magnitude of the effect is overestimated. Judgment and Decision Making. 2016;11(6):601-610. doi:10.1017/S1930297500004812.

[17] 'Validity' may include the realities of funding, operational and organisational constraints. However it does restrict learners' progress.

There is an important benefit to learners and industry here. Assessment should support learning design to maximise both performance and long-term learning.[18] It helps attune learners and workers to self-assessment. Self-assessment is the capacity to reflect on, and evaluate, their own work. It builds their autonomy and self-regulation and attunes learners to further or professional development.[19] [20]

Self-assessment also helps develop a healthy respect and appreciation (rather than paralysing fear) for the role of mistakes in learning.[21] It is part of self-regulation in learning, where learners self-monitor and reflect, set their own increasingly challenging goals and persist in the face of motivational setbacks. Research highlights that self-regulated learning is as important in vocational settings as academic ones, though often overlooked because vocational learners are generally seen as less capable in this area.[22] [23] These are hallmarks of the lifelong learning paradigm and crucial to industry having an eye to the future as well as the present.

[18] For a discussion on the productive possibilities between the respective extremes and drawbacks of "discovery learning" and direct instruction see: Manu Kapur, 'Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning', Educational Psychologist 51, no. 2 (2 April 2016): 289–99.

[19] Ernesto Panadero, Daniel Garcia, and Juan Fraile, 'Self-Assessment for Learning in Vocational Education and Training', Handbook of Vocational Education and Training: Developments in the Changing World of Work, 2018, 1–12.

[20] David Boud, 'Sustainable Assessment: Rethinking Assessment for the Learning Society', Studies in Continuing Education 22, no. 2 (November 2000): 151–67, https://doi.org/10.1080/713695728.

[21] Karen Vaughan, Linda Bonne, and Jan Eyre, 'Knowing Practice: Vocational Thresholds for GPs, Carpenters, and Engineering Technicians' (Wellington: New Zealand Council for Educational Research and Ako Aotearoa, December 2015).

[22] Helen Jossberger et al., 'Exploring Students' Self-Regulated Learning in Vocational Education and Training', Vocations and Learning 13, no. 1 (April 2020): 131–58, https://doi.org/10.1007/s12186-019-09232-1.

[23] Karin Smit et al., 'The Self-Regulation of Motivation: Motivational Strategies as Mediator between Motivational Beliefs and Engagement for Learning', International Journal of Educational Research 82 (2017): 124–34.

SKILL STANDARDS-BASED ASSESSMENT

In the environment of Aotearoa/New Zealand's standards-based qualifications system, there are some further considerations that assessors must apply. The development of skill standards as a further type of assessment standard prompts a re-examination of good practice principles in application.

Trusting assessors and the assessment process

The move from unit standards to skill standards is emblematic of a higher-trust model where the degree of specification and prescription should lessen over time. See the Standards and Qualifications Development guide in this toolkit for further discussion. For the assessor, greater trust involves potentially more complex assessment and decision-making processes. For all it certainly means setting aside the 'tick-box' thinking that relegated assessment to an activity involving technicians rather than requiring professionals.

Assessment occurs at the level of the Learning Outcome

This basic rule has always been the case, though frequently not followed. As standards grow in credit values while also reducing their levels of prescription, assessors must pay more attention than before to the relative weight and value of the outcome statements. The following table provides some examples.

Table 1 A comparison of skill standard and unit standard credit values

Standard	Outcomes or Learning Outcomes	Performance or Assessment Criteria	Credit Value
Unit standard 23977 Use and maintain a hydraulic jack and attachments in the motor industry	2	8	1
Unit standard 23229 Use safety harness system when working at height	5	15	4
Skill standard 40000 Manufacture a timber flooring cassette	2	5	20

Making sense of the variability and assigning priorities within a programme of learning and assessment are essential. Ongoing dialogue between SSB and provider, allied with careful documentation of agreed positions is likely to be invaluable to reduce confusion and the potential for later disagreement. A strict application of assessment at the level of the Learning Outcome will enhance RPL/RCC systems for providers.

Using Assessment Criteria

Assessment Criteria, like the Performance Criteria they replace, sit below the Learning Outcomes and give guidance on what constitutes acceptable performance against the standard. An ideal Assessment Criterion explains the behaviour that the assessor should be looking for with an indication of the level it should be at. As with the Learning Outcomes, there's a certain amount of variance in the way that SSBs construct Assessment Criteria. The following table shows samples with our commentary offering guidance.

Table 2 Examples of Assessment Criteria with good practice commentary

Assessment Criteria	Commentary
Best practice to ensure tools, equipment and waste product do not create a hazard is explained.	This provides clear information to the assessor.
Disassemble and reassemble a computer from component parts.	This explains the tasks required but further discussion between provider and SSB will be required to establish mutual understandings about the type and complexity of the computer involved and any quality measures that may be implied, such as the operational status of the assembled device.
Perform the selected task under supervision.	In this case there's little clarity about either the task or the level of performance that's required. Assessor and SSB would need to debate and reach agreement.
Cabling conduit or trunking are installed according to plans and specifications.	This is a clear description of both activity and required level of performance.

Interpreting Assessment Specifications

Assessment Specifications provide both helpful guidance and definitive instructions as well as information. They must be read and interpreted with care. When appropriate, providers should ensure that their reading of the Assessment Specifications complies with the SSB's intentions. The parties may need that concurrence to be recorded in writing. The following table shows samples with our commentary offering guidance.

Table 3 Examples of Assessment Specifications with good practice commentary

Assessment Specification	Commentary
It is required that the practical assessment evidence is obtained in the workplace.	Because of the legal implications of the word 'workplace' this means that assessment in a provider's workshop environment (regardless of how authentic) is excluded. Was that what was meant?
Assessment must be in a manufacturing environment.	This is very clear: a provider wanting to assess must have access to the plant and be involved in the associated commercial processes.
Assessment against this skill standard must be based on evidence from a real or simulated workplace situation.	Knowing how the workplace may be acceptably simulated would be crucial before designing the course.
Tasks can be scenario and/or simulation based, selected and performed under supervision.	Both 'scenarios' and 'simulations' offer a wide range of possibilities, many possibly unacceptable to the SSB. Caution is required. Compare with the two clearer examples: • *Learners may be assessed against this standard in a real-life context using naturally occurring evidence or in a realistic simulation.* • *Skills must be demonstrated in the workplace or in a simulated environment that reflects workplace conditions and contingencies.*
Small mobile plant refers to machinery under 4500kg.	This provides very useful information.
A verifier's checklist is acceptable if accompanied by evidence that includes examples from the learner's performance.	This is an example of an SSB providing very clear expectations.
The plan developed must be to the standard of a commercially competent pump operator.	This is useful as long as the words 'commercially competent' are explained and clarified further, as they are potentially subject to a substantial degree of interpretation.
Evidence presented for assessment against this skill standard must be consistent with safe work practices and be in accordance with applicable industry standards, workplace procedures and legislative requirements.	This final example illustrates a potential risk in the construction of Assessment Specifications. What happens when 'workplace procedures' are at odds with 'industry standards' or 'legislative requirements' (not an uncommon situation)? Which takes precedence?

Repeat performance is a particular focus of many standards and has been an area which has challenged the relationship between the standard-setter and the professional assessor. Sufficiency statements that said, "At least three times…" or, "At least eight types of…" were at least clear, but they were examples of a low-trust environment and moreover were frequently 'finger in the air' guesses as proxies for competence. Thay failed to take into account the aptitude of the assessor and the varying capabilities of learners, managing the fluke factor and, especially, the value and risks of the number itself. Was it three successful performances in a row by a capable learner, or was it three successes from many attempts by a learner needing further development? Put differently, what if three was too many, or too few?

Many SSBs are now using the formula: It is expected that the assessor affirms the learner's ability to repeat their performance against the standard. This is unequivocal. It's not a suggestion; it's a requirement. And it does mean that the assessment process must include methods of recording that assessor affirmation, and that assessors and SSBs are confident that they mean the same thing by that affirmation.

CONSISTENCY MEASURES

Moderation is a second opinion

Moderation is the most common terminology for measures that are designed specifically to help ensure that the outcomes of assessment are consistent, and that assessors continue to grow professionally. Moderation is not just one thing at one time. It comes in a wide range of forms as the following table summarising research into forms of moderation indicates. Note that the earlier stages offer the most scope for professional, collegial activity.[24]

Table 4 Examples of moderation activities at different stages of assessment process

Stage	Focus	Activities
Design	Quality of tasks and overall plan (course and programme)	Peer scrutiny; professional processes
Calibration	Shared understanding of requirements, criteria and standards	Benchmarking workshops; socialisation; comparing and justifying judgements
Judgements	Quality of judgements (adherence to criteria; credibility of evidence; judgement consistency)	Consensus moderation discussions; collaborative or 'double marking'; random checks
External validation or comparison	Comparability of standards	External examining; peer review; professional processes
Monitoring evaluation	Overall quality of assessment and of components of individual stages	Consideration of: learner work samples; learner satisfaction data; examiners' reports; trainer/tutor perceptions; design; criteria and standards; assessment guides

In other words, all forms of reflection are forms of moderation, wherever they occur in an overall process. This would include considering the entire assessment and moderation system itself.

This guide uses the word 'consistency' to emphasise its close relationship to teaching, learning and assessment. It places a discussion about consistency in the section about assessment for a good reason: to summarise the central purposes of moderation, it makes best sense to say that it provides a second opinion for aspects of the learning and assessment process.

This being the case, moderators should 'walk in the shoes' of assessors, understand their contexts and be part of the process that contributes to their growth. Just as good practice assessment is a process rather than an event, so is moderation. In cases where assessment was a mechanistic, 'box-ticking' activity, it was reasonable that moderation would respond accordingly. As assessors develop their professional skills, the moderation process should follow suit, and vice versa.

Moderation doesn't hold the power of veto over assessment decisions. This has perhaps encouraged some to regard moderation as needing to be, effectively, a compliance activity at a higher level, for example at the level of the institution. For example, the term 'quality assurance' is often used as a synonym for moderation (or consistency measures), inadvertently conveying a sense of impersonal oversight that adopts strict criteria rather than the flexible, collaborative process advocated by this guide. The main effects of that attitude are to undermine the capability-building potential of moderation and squander the resources we are able to devote to supporting consistency.

[24] This table was adapted from page 640 in Sue Bloxham, Clair Hughes, and Lenore Adie, 'What's the Point of Moderation? A Discussion of the Purposes Achieved through Contemporary Moderation Practices', Assessment & Evaluation in Higher Education 41, no. 4 (18 May 2016): 638–53.

Moderation is collegial and supportive

If one goal of consistency measures is to encourage professional growth, then institutions which employ many assessors should reasonably expect to invest in collegial activity that brings assessors together with the intention of having several assessors emerge thinking and acting as one.[25]

For internal moderation there is no substitute for the 'community of practice' approach, which is rooted in collegial rather than reporting relationships, and creates a 'craft intimacy' from close interactions around shared problems and knowledge production.[26] This concept places the role of moderator within the group, whether two or twenty, whether working in the same discipline or not. Participants may share experiences and dilemmas; they may choose to 'mark in the group'; they may discuss and debate questions of interpretation, judgment or validation; they may share anecdotes of what worked well and what didn't. They may regard themselves as buddy moderators or peer support. It doesn't really matter.

The principles and practice of internal moderation may well extend beyond the institution where circumstances permit. Broadening the borders of a community of practice to encompass colleagues working at some distance can reduce the potential for localised group-think or shared misinterpretations.[27]

Engaging in the moderation process can be a powerful way for educators and assessors to gain and share not only a deeper understanding of their work but also of the standards that they assess against.[28] Research shows that 'calibration' for the purposes of agreed and deeply internalised standards among colleagues can be effective.[29] While some moderation activities add little more than a public image of systematic checking, 'social moderation' and community building add to both assessors' 'assessment literacy' as well as their knowledge of standards, because it shifts the focus from narrow accountability to enhancing teaching and learner opportunities.[30] [31] External moderation protects the interests of the standard-setter's stakeholders, including the general public. It can (and should) examine closely the systems and administrative practices in place to support good assessment, including internal moderation systems. It can and should be concerned about how well the learning programme is supported by systems, staffing and facilities. It should offer meaningful, supportive and achievable suggestions for improvement where this is appropriate. In other words, a 'systems overview' is an appropriate and realistic way of regarding the external moderation effort.

In this context of mutual support and a focus on improvement, there is really no place for a 'tick-the-box' compliance attitude to consistency measures.

The Skill Standard context

For many participants, the word 'moderation' conjures images of people dealing with paper-based artefacts: terms such as 'pre-assessment moderation' and 'post-assessment moderation' have helped convey the impression that consistency measures are all about the detail or coverage of testing and the accuracy of marking, and not much else. But considerations of consistency apply throughout the process from interpreting the qualifications and standards through programme development, teaching and assessment, and to the systems that support those processes.

[25] For an example of an assessment community of practice for construction industry education, see Vaughan, Gardiner, and Eyre, 'The Transformation of Industry-Led Assessment of On-Job Learning in the Building and Construction Industries'.

[26] Etienne Wenger, Richard McDermott, and William M. Snyder, Cultivating Communities of Practice (Boston: Harvard Business School Press, 2002).

[27] Vaughan, Kear, and MacKenzie, 'Mate, You Should Know This! Re-Negotiating Practice after a Critical Incident in the Assessment of on-Job Learning. In Vocations and Learning.'

[28] Rosemary Hipkins and Sally Robertson, 'Moderation and Teacher Learning: What Can Research Tell Us about Their Interrelationships?' (New Zealand Council for Educational Research, 2011).

[29] D. Royce Sadler, 'Assuring Academic Achievement Standards: From Moderation to Calibration', Assessment in Education: Principles, Policy & Practice 20, no. 1 (1 February 2013): 5–19.

[30] Bloxham, Hughes, and Adie, 'What's the Point of Moderation? A Discussion of the Purposes Achieved through Contemporary Moderation Practices'.

[31] Kim Watty et al., 'Social Moderation, Assessment and Assuring Standards for Accounting Graduates', Assessment & Evaluation in Higher Education 39, no. 4 (19 May 2014): 461–78.

In the environment of the skill standard-based qualification where atomisation and excessive specification are targeted and discouraged, the business of external moderation should shift from 'moderating the unit standards' to evaluating the systems, processes and artefacts that support principled and consistent assessment decisions. As this guide identifies, the best assessment is unobtrusive, ongoing and flexible. To the extent possible, good practice moderation follows similar principles.

It is well understood that the Indicative Content of skill standards is set at the level of guidance rather than instruction. Perhaps for this reason, SSBs have adopted significantly differing approaches to the level of detail that they include in Indicative Content. Moderators need to be collaborating with their SSB development colleagues over their stakeholders' intentions and wishes about interpretations of Indicative Content, as well as other aspects of the skill standards. Likewise, providers should engage preemptively with SSB external moderation teams over these questions before any programme endorsement. The forms that interactions take could be flexible, though a degree of formality in recording agreements would be prudent. What is certain is that identifying uncertainty or disagreement after the programme has been delivered is too late.

It is entirely appropriate that the external moderation process examines the realities of educational delivery as it relates to the programme the SSB has endorsed. Understanding how to evaluate the teaching and learning processes within the provider's specific context is crucial to the relationship. It is not the role of the moderator to tell providers how to deliver their programmes, but to be trusted advisors whose guidance is sought and valued.

If good assessment occurs at the level of the skill standard's Learning Outcomes, then moderation will follow suit. Minutely examining compliance with the particular wording of Performance Criteria has been unhelpful and wasteful of resources. It has also sometimes been unnecessarily confrontational. In the skill standard world, discussions about how Assessment Criteria have assisted in guiding the assessment process provide the opportunity for moderators to genuinely assist by understanding the provider's thinking and potentially recommending alternative approaches which reflect industry's thinking. The following table shows examples of this.

Table 5 A moderation eye on Assessment Criteria

Example Assessment Criteria	Commentary re: moderation
An evaluation is conducted to determine each risk's potential impact on overall ecosystem health.	There's a need for concurrence on the depth of the evaluation, its scope and format, and on what form a successful evaluation takes. This exercise could be done on paper, of course, but benefits from a discussion of principles.
Reflect on own contribution to the group's outcome.	In this case the Assessment Criterion requires a real-life context or a realistic simulation. Moderators want to know (and can assist with) how judgements are made, verified and recorded/ reported. The provider's internal moderation processes will provide important underpinning for consistent outcomes and the moderator has a legitimate interest in matching these components up and making observations and recommendations.
The tools, equipment and products selected are appropriate for the task, and work area is set up to optimise safety and ease of working.	This is a well written AC, though clarity may be required about the words 'appropriate' and 'optimise' with a potential discussion about sufficiency: what if the 'task' needs only one 'tool'? Is that OK?

Equally, moderators and providers must reach common agreements about Assessment Specifications. Frequently enough the level of specification is somewhat loose. The following table provides some examples.

Table 6 A moderation eye on Assessment Specifications

Example Assessment Specifications	Commentary re: moderation
A variety of forms of evidence is accepted, such as photos, signed observation sheets, video recordings, written reports, oral presentations, etc.	This is not strictly speaking specifying anything. In this case it is entirely appropriate that the provider and the moderation team discuss in some detail what is or is not ac-ceptable as an interpretation of 'etc'.
Practical assessment tasks must be verified and signed by a verifier who has relevant current industry expertise.	Here, clarification and agreement may be needed as to how 'current expertise' is established, how 'relevant' it should be and perhaps acceptable forms of sign-off.
Forestry earthworks tasks can be scenario and/or simulation based, selected and performed under supervision.	Discussion is certainly required over how to interpret the words 'scenario' and 'simulation'. The phrase 'under supervision' may also need clarity and agreement.
It is expected that the assessor affirms the learner's ability to repeat their performance against the standard.	This requirement is very clearly stated, but moderator and assessor will certainly need to be 'of a mind' about how the affirmation is arrived at, how that evaluation is applied consistently across learners and how it's recorded.

Positioning moderators as trusted advisors will present a challenge for some partners who have been accustomed to adopting a 'them-and-us' mindset in regard to the relationship between provision and external moderation. It may require adjustments by many. But as assessment starts on day one of the course, the moderation relationship can and should be ongoing and not confined to isolated events.

A less explicit or discussed aspect of consistency is that it is about keeping things as consistent as possible with the way things have been in the past, and usually in order to reduce risk.[32] In this guide, we emphasise consistency measures as an approach to accountability that not only improves the motivation and capacity to make sound judgements but does so without stifling innovation.

[32] E. Louise Hayward and Carolyn Hutchinson, '"Exactly What Do You Mean by Consistency?" Exploring Concepts of Consistency and Standards in Curriculum for Excellence in Scotland', Assessment in Education: Principles, Policy & Practice 20, no. 1 (1 February 2013): 53–68.

REFERENCES

Bandura, Albert. ‘Self-Efficacy: Toward a Unifying Theory of Behavioral Change’. Advances in Behaviour Research and Therapy 1, no. 4 (1978): 139–61.

Bloxham, Sue, Clair Hughes, and Lenore Adie. ‘What’s the Point of Moderation? A Discussion of the Purposes Achieved through Contemporary Moderation Practices’. Assessment & Evaluation in Higher Education 41, no. 4 (18 May 2016): 638–53.

Boud, David. ‘Sustainable Assessment: Rethinking Assessment for the Learning Society’. Studies in Continuing Education 22, no. 2 (November 2000): 151–67. https://doi.org/10.1080/713695728.

Chan, Selena. ‘New Zealand’s Move to Graduate-Profile Framed Qualifications: Implications, Challenges and the Occupational Identity Solution’. International Journal of Training Research 14, no. 1 (2016): 5–18.

———. ‘Recognition of Current and Prior Experience in Aotearoa New Zealand and the Role of Eportfolios’. In International Handbook on Education Development in Asia-Pacific, edited by Wing On Lee, Phillip Brown, A. Lin Goodwin, and Andy Green, 1–17. Singapore: Springer Nature Singapore, 2023.

De Vos, M. E., L. K. J. Baartman, C. P. M. Van Der Vleuten, and E. De Bruijn. ‘Unravelling Workplace Educators’ Judgment Processes When Assessing Students’ Performance at the Workplace’. Journal of Vocational Education & Training 76, no. 3 (26 May 2024): 517–36.

Dweck, Carol S. Mindset. How You Can Fulfill Your Potential. Great Britain: Robinson, 2012.

Gamble, Jeanne. ‘Why Improved Formal Teaching and Learning Are Important in Technical and Vocational Education and Training (TVET)’. Revisiting Global Trends in TVET: Reflections on Theory and Practice 204 (2013).

Gillis, Shelley, and Andrea Bateman. ‘Assessing in VET: Issues of Reliability and Validity’. Adelaide: NCVER, 1999.

Hayward, E. Louise, and Carolyn Hutchinson. ‘“Exactly What Do You Mean by Consistency?” Exploring Concepts of Consistency and Standards in Curriculum for Excellence in Scotland’. Assessment in Education: Principles, Policy & Practice 20, no. 1 (1 February 2013): 53–68.

Hipkins, Rosemary. ‘Assessment of Naturally Occurring Evidence of Literacy’. Assessment Matters 4 (2012): 95–109.

Hipkins, Rosemary, and Sally Robertson. ‘Moderation and Teacher Learning: What Can Research Tell Us about Their Interrelationships?’ New Zealand Council for Educational Research, 2011.

Jossberger, Helen, Saskia Brand-Gruwel, Margje W. J. Van De Wiel, and Henny P. A. Boshuizen. ‘Exploring Students’ Self-Regulated Learning in Vocational Education and Training’. Vocations and Learning 13, no. 1 (April 2020): 131–58. https://doi.org/10.1007/s12186-019-09232-1.

Kapur, Manu. ‘Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning’. Educational Psychologist 51, no. 2 (2 April 2016): 289–99.

New Zealand Qualifications Authority. ‘Aromatawai and the Principles of Assessment Supporting Aromatawai and the Development of Quality Assessment Practices’. Wellington: New Zealand Qualifications Authority, August 2022. https://www2.nzqa.govt.nz/assets/About-us/News/aromatawai-and-the-principles-of-assessment.pdf.

———. ‘Guidelines for Listing Skill Standards on the Directory of Assessment and Skill Standards’. Wellington: New Zealand Qualifications Authority, August 2024. https://www2.nzqa.govt.nz/assets/Tertiary/Approval-accreditation-and-registration/Standards/Skills-standards/Guidelines-for-listing-skill-standards-on-the-DASS.pdf.

Panadero, Ernesto, Daniel Garcia, and Juan Fraile. ‘Self-Assessment for Learning in Vocational Education and Training’. Handbook of Vocational Education and Training: Developments in the Changing World of Work, 2018, 1–12.

Sadler, D. Royce. ‘Assuring Academic Achievement Standards: From Moderation to Calibration’. Assessment in Education: Principles, Policy & Practice 20, no. 1 (1 February 2013): 5–19.

Smit, Karin, Cornelis J. de Brabander, Monique Boekaerts, and Rob L. Martens. ‘The Self-Regulation of Motivation: Motivational Strategies as Mediator between Motivational Beliefs and Engagement for Learning’. International Journal of Educational Research 82 (2017): 124–34.

Vaughan, Karen, Linda Bonne, and Jan Eyre. ‘Knowing Practice: Vocational Thresholds for GPs, Carpenters, and Engineering Technicians’. Wellington: New Zealand Council for Educational Research and Ako Aotearoa, December 2015.

Vaughan, Karen, and Marie Cameron. ‘A Guide to Good Practice in Industry Training Organisation Structures and Systems For On-Job Assessment’. Wellington: Ako Aotearoa, 2010.

Vaughan, Karen, Ben Gardiner, and Jan Eyre. ‘The Transformation of Industry-Led Assessment of On-Job Learning in the Building and Construction Industries’. Wellington: Ako Aotearoa, 2012.

Vaughan, Karen, Andrew Kear, and Heath MacKenzie. ‘Mate, You Should Know This! Re-Negotiating Practice after a Critical Incident in the Assessment of on-Job Learning’. Vocations and Learning 7, no. 3 (2014): 331–44.

Watty, Kim, Mark Freeman, Bryan Howieson, Phil Hancock, Brendan O’Connell, Paul de Lange, and Anne Abraham. ‘Social Moderation, Assessment and Assuring Standards for Accounting Graduates’. Assessment & Evaluation in Higher Education 39, no. 4 (19 May 2014): 461–78.

Wenger, Etienne, Richard McDermott, and William M. Snyder. Cultivating Communities of Practice. Boston: Harvard Business School Press, 2002.