



Teaching, learning and assessment in TVET: the case for an ecology of assessment

DAVID D CURTIS

NATIONAL CENTRE FOR
VOCATIONAL EDUCATION RESEARCH

Paper presented at the South-East Asian Ministers of Education Organisation (SEAMEO) VOCTECH Conference 'Exploring Best VET Policies and Practices through Research and Networking Initiatives', Brunei Darussalam, 23 June 2010

The views and opinions expressed in this document are those of the author and do not necessarily reflect the views of the Australian Government or state and territory governments.

© National Centre for Vocational Education Research, 2010

With the exception of cover design, artwork, photos, all logos, and any other material where copyright is owned by a third party, all material presented in this document is provided under a Creative Commons Attribution 3.0 Australia <<http://creativecommons.org/licenses/by/3.0/au>>.



This document should be attributed as Curtis D D 2010, *Teaching, learning and assessment in TVET: the case for an ecology of assessment*.

The National Centre for Vocational Education Research (NCVER) is an independent body responsible for collecting, managing and analysing, evaluating and communicating research and statistics about vocational education and training (VET).

NCVER's inhouse research and evaluation program undertakes projects which are strategic to the VET sector. These projects are developed and conducted by the NCVER's research staff and are funded by NCVER. This research aims to improve policy and practice in the VET sector.

ISBN 978 1 921809 52 1 web edition

TD/TNC 102.34

Published by NCVER

ABN 87 007 967 311

Level 11, 33 King William Street, Adelaide, SA 5000

PO Box 8288 Station Arcade, Adelaide SA 5000, Australia

ph +61 8 8230 8400 fax +61 8 8212 3436

email ncver@ncver.edu.au

<<http://www.ncver.edu.au>>

<<http://www.ncver.edu.au/publications/2328.html>>

About the research



Teaching, learning and assessment in TVET: the case for an ecology of assessment

David D Curtis, NCVER

Rather than being an event that occurs after instruction to check on individual learning, assessment can be central in driving high-quality learning and instruction. As such, assessment should be a process that is designed into teaching and learning.

It is helpful to think about teaching, learning and assessment in vocational education and training (VET) as components of an 'ecosystem' of skills, including their development and deployment; of the agents who operate within the system, the teachers, learners and administrators of the system; and of the social and industrial contexts in which skills are developed, certified and deployed.

The paper argues that:

Selecting an appropriate mix of assessment strategies can satisfy the multiple goals of key stakeholders in VET.

Forms of assessment must be consistent with the goals of learning and with a broad range of valued learning outcomes.

Many forms of assessment are available, and a key task is to match these forms with goals for assessment and with desired learning outcomes. This matching needs to occur within courses and at a systems level in order to achieve the diversity of assessment goals.

Tom Karmel
Managing Director, NCVER

Contents

Teaching, learning and assessment in TVET: the case for an ecology of assessment	6
The multifaceted environment of TVET	6
Purposes served by assessment systems	8
Features of the Australian VET system that influence teaching, learning and assessment practices	10
A typology of assessment methods	13
A case study of formative assessment and learning	16
Implications for TVET policy and practice	18
References	20

Teaching, learning and assessment in TVET: the case for an ecology of assessment

Teaching, learning and assessment in technical and vocational education and training (TVET) is a broad topic, and it seems sensible to focus on a key element—assessment. Rather than being an event that occurs after instruction to check on individual learning, assessment can be central in driving high-quality learning and instruction. Assessment should be a process that is designed into teaching and learning, and being a design feature of TVET it must meet the needs of the many TVET stakeholders. These include governments, industry bodies and employers, TVET practitioners and of course learners.

It is helpful to think about teaching, learning and assessment in TVET as components of an ecosystem of skills, including their development and deployment; of the agents who operate within the system, the teachers, learners and administrators of the system; and of the social and industrial contexts in which skills are developed, certified and deployed. Some of these components and relationships among them are illustrated in figure 1.

The multifaceted environment of TVET

TVET systems serve several important functions, most notably that of building the pool of skills within a country's economy. TVET systems make other direct and indirect contributions. For example, adult and community education (ACE) is linked to the TVET systems of many countries and provides foundational skills development for those who have not acquired these skills through the school systems. Such contributions can enhance the participation of adults in social and community as well as in work contexts and build the social capital of communities.

In most countries, TVET systems are primarily responsive to the requirements of industry, although the balance of influence on TVET policy-making bodies between government, industry and employer organisations, education and training providers, and employees varies.

The tightness of the coupling between qualifications and the labour market also varies. In many Western European countries, the coupling is tight. For an individual to practise at a given level in a professional or vocational field, they must have a qualification at a specified level within that field (Allen & van der Velden 2005; van der Velden & Wolbers 2008). In other countries, this linkage between the labour market and qualifications is much weaker and differs by occupation and level. Certain professions and occupations, for example, licensed trades, have controlled entry requirements, while others have rather weak links (Karmel, Mlotkowski & Awodeyi 2008). Strong linkages provide some level of quality assurance for employers, industries and possibly consumers, while the weaker linkages provide a degree of flexibility in the labour market. The degree of coupling influences the level of control exercised by licensing and accrediting bodies in course content and assessment requirements.

Government policies exert very strong influences over the operation of TVET. Where TVET is dependent upon governments for funding or where TVET providers are highly regulated, their operations are constrained. Strict regulation seeks to place a floor level under the quality of provision, whereas more open systems depend on training markets to control the quality and relevance of training provision. Other policy objectives, for example, those associated with increasing labour market participation and equity targets place other pressures on TVET providers.

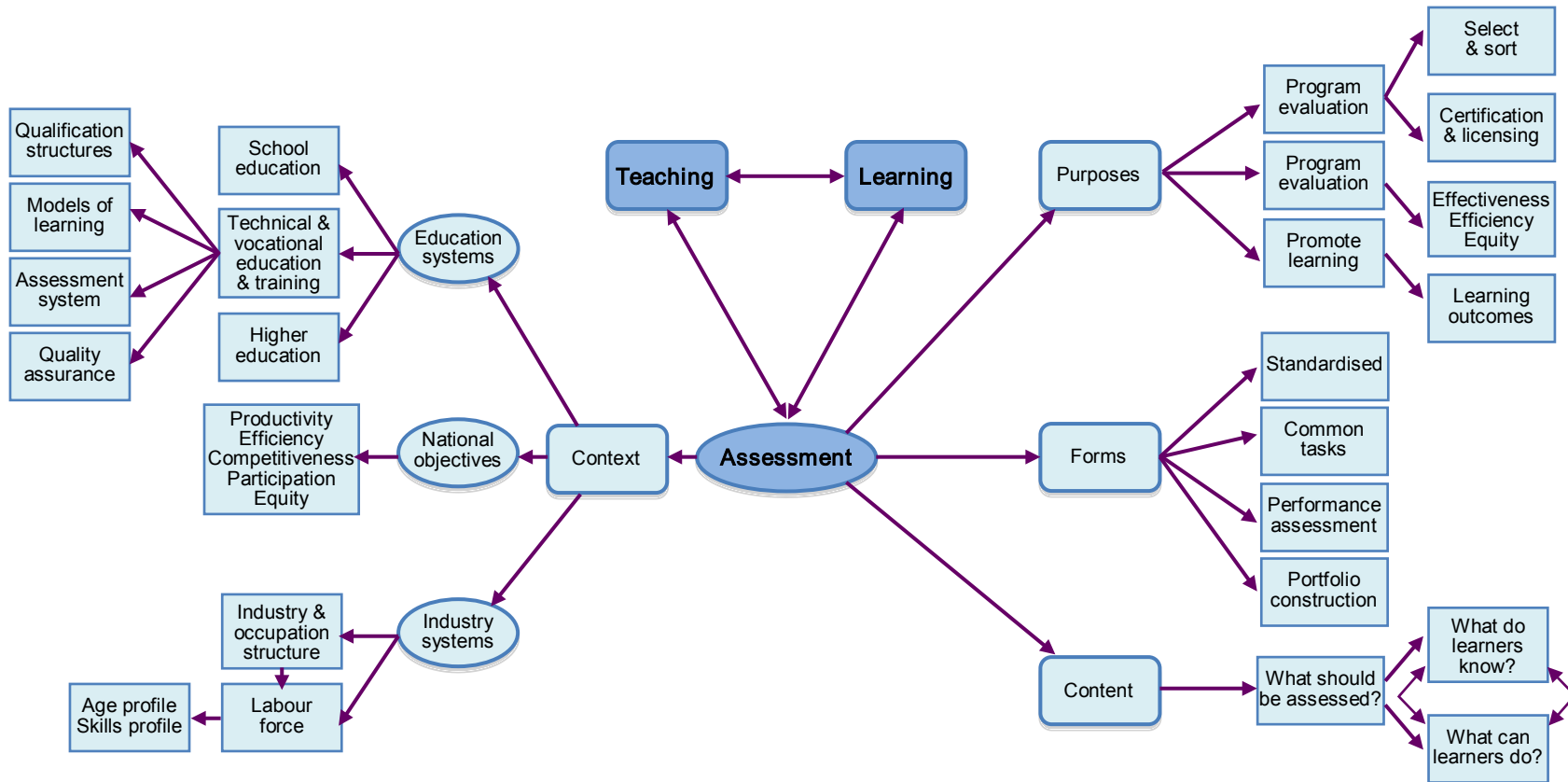


Figure 1 Environmental context of teaching, learning and assessment in TVET

These are broad contextual factors that influence the operation of TVET systems and providers. They are not discussed further in this paper. However, after considering the purposes of and approaches to assessment, some structural and regulatory features of TVET systems that influence teaching, learning and assessment methods are reviewed.

Purposes served by assessment systems

I focus on assessment in this paper. Teaching, learning and assessment form key elements of a larger structure, and I allude to some of these elements above. A key point about assessment is that it is, whether by design or not, a driver for teaching and learning. If we establish the right assessment processes, effective teaching and learning will follow. Boud (1995, p.40), quoted Eisner (1993) as saying ‘... the act of assessment signals the importance of what is being assessed, so assessment is a driver for learning’.

Assessment is a process of gathering evidence, making judgments and drawing inferences about student achievement and performance. Airasian (1994) and Pellegrino, Chudowsky and Glaser (2001) summarised the purposes of assessment as:

promoting learning

measuring individual achievement

evaluating programs.

Measuring individual achievement

We seek to measure student achievement for competitive selection into institutions and courses, into jobs, and for certification. Such assessment is said to be summative—it provides a summary of students’ accumulated learning during a course of study. These purposes are important, well understood and widely practised. Methods normally used for summative purposes can generate information that can be used for individual diagnostic purposes and aggregated to the classroom (teacher) level, the school level and the system level. An example of this approach can be found in the National Assessment Program—Literacy and Numeracy (NAPLAN) testing program that operates at Years 3, 5, 7 and 9 in all Australian states and territories.

Evaluating programs and providers

Less well understood and practised is the use of individual student achievement, measured through summative assessment activities such as examinations and tests, to provide information about the quality of instruction or the effectiveness of the educational systems that have given rise to the observed performances. There are, however, major international programs that aggregate individual learner achievement data in order to evaluate programs and systems, and do it well. Examples include the International Association for the Evaluation of Educational Achievement (IEA) series of tests in reading, mathematics, science and civics and the Programme for International Student Assessment (PISA) conducted under the auspices of the Organisation for Economic Co-operation and Development (OECD). These examples (and the NAPLAN one cited above) are drawn from the school sector. However, similar tests are conducted on adult populations in programs such as the International Adult Literacy Survey (IALS) of the mid-1990s, the Adult Literacy and Life Skills (ALLS) survey conducted between 2003 and 2006 and the forthcoming International Programme for the Assessment of Adult Competencies (PIAAC). A new round of international comparative testing, the Assessment of Higher Education Learning Outcomes (AHELO) is being trialled by the OECD. These studies are valuable because they enable countries to compare the performances of their students at a given age or grade level with similar students from other countries. Differences between national means, which no doubt cause a degree of political heartburn, can be informative and countries whose achievement is consistently high become exemplars for others. Note that these examples are drawn from the school and higher education sectors and adult populations, but

that the TVET sector is missing. The relative absence of this sector in research and debates about assessment is concerning.

The testing programs listed above operate mostly in advanced economies and are undertaken as human capital development is seen in those countries as an element of international competitiveness.

Promoting learning

Let us return to the first of the three purposes listed above—promoting learning. How can we use assessment to promote learning? The answer may lie in formative assessment ‘... the frequent assessment of learner understanding and progress to identify needs and shape teaching’ (Looney 2008, p.21). However, there is no archetypal formative ‘method of assessment’. Several aspects of assessment make it formative.

First, assessment is embedded into learning routines. This may be as simple as prompting learners with questions to focus their attention on aspects of a problem or task.

Second, such assessment occurs frequently. It can only occur frequently if it is embedded in routine learning and teaching practices. If assessment tasks do not contribute directly to learning, the frequency of their use would almost certainly detract from learning, as students would be spending less time engaged in knowledge building, and we know from a very large body of research that time spent on a task is a key determinant of learning. This, then poses a challenge. How can assessment tasks be designed so that they engage students in activities that do lead to learning?

This brings us to the third and a central characteristic of formative assessment—it must be accompanied by informative feedback to learners about their learning progress. Informative feedback depends upon teachers being able to evaluate learners’ responses to tasks and to diagnose deficiencies in learners’ understandings and to provide instruction and suggest actions that will remediate those deficiencies. That is, the job of the teacher is to identify the gap between what students know and can do now and the goals for what students should ‘know and be able to do’. Feedback, then must be ‘... information about the gap between the actual and the reference level of a system parameter which is used to alter the gap in some way’ (Ramprasad 1983). The provision of informative feedback, therefore, requires constructive interactions between teachers and learners.

A fourth and critical element of formative assessment is that learners must engage in the assessment process. The point has been made by many researchers (for example, Boud 2002) that, for most learners, assessment is something that is done to them and they are passive subjects of it. Several approaches have been taken to encourage students to engage in assessment. One of the ways this has been done is through self-assessment, although peer-assessment is also used (Black & Wiliam 1998; Boud & Hawke 2003; Sadler 1989; Wiggins 1998). Boud (2002, p.43) argues that:

By deliberately keeping assessment out of the hands of learners, we are denying them one of the essential tools—perhaps the essential tool—which enables them to become lifelong learners.

Wiggins (1998, p.46) provides an excellent example of feedback combined with self-assessment that is particularly relevant to technical and vocational learning. He cites the case of a welding teacher who provided annotated work samples. Students were required to undertake a task—welding a tee-joint—and to present their finished product to the teacher. But, as they left their work benches to take their work to the teacher, they walked past a set of work samples, each with notes attached pointing out the good and poor points of each job. Students invariably compared their work with the work samples and took note of the comments. Many students returned to their benches and re-attempted the task. They had looked at the work samples and had internalised the assessment standards that were explicit in the comments attached to them. They had also self-diagnosed the ‘gap’ between their own efforts and the desired level of performance. This example also draws attention to a feature of formative assessment that reveals its potential importance in lifelong

learning. In workplaces, competent practitioners do not have an assessor telling them whether their work is adequate or not; each practitioner, whether a professional or tradesperson, is expected to make judgments about the suitability of each element of work they undertake. Those who cannot do this may be restricted to low-value work in highly structured and supervised contexts.

Formative assessment can lead to very considerable learning gains, as Black and Wiliam (1998) demonstrate by assembling a substantial body of evidence. They conclude that the learning gains associated with formative assessment are ‘among the largest ever reported for educational interventions’ (p.34). However, formative assessment does not always yield these gains and does not invariably lead to the outcomes that Wiggins (1998) alluded to for welding learners.

Various aspects of the environment in which formative assessment is used moderate its effectiveness. Davies and Ecclestone (2008) present two case studies in which similar formative practices are adopted, but from which quite different learning outcomes arise. In one, a productive environment, there is a coherence in the relationship between learning goals and practices, which leads to greater learner engagement with the content, whereas in the other, there is a dissonance between the goals and practices, which leads to the formative practices becoming ‘instrumental’. That is, learners (and teachers) go through the motions of formative practices, but the feedback relates mainly to achievement goals and to directing students to satisfy the immediate demands of summative assessment requirements, but without focusing on enduring learning outcomes. Davies and Ecclestone attribute part of the difference between the two cases to differences in teacher beliefs and attitudes about what students could do and about what goals are important. They also draw attention to the learning environment of the colleges they compare—they show that institutions have cultures of learning that are more or less conducive to sustainable learning. They highlight features of learning systems beyond the institutions in which teachers and learners operate. They refer, for example, to the ‘syllabus or course specification, the assessment and qualification specifications’ and to the ‘funding and body procedures and regulations, and government policy’ (p.75). This leads us to an examination of some of the features of the Australian education and training systems that influence assessment practices and which may have wider applicability.

Features of the Australian VET system that influence teaching, learning and assessment practices

These comments evoke four aspects of TVET and other education systems. Four aspects of the Australian VET system, namely the Australian Qualifications Framework (AQF), the use of training packages, the use of competency-based training and assessment, and the quality assurance processes of the Australian Quality Training Framework (AQTF) are used to illustrate the relationship between leaning and assessment and system characteristics.

The Australian Qualifications Framework

The Australian Qualifications Framework describes a system of nationally recognised qualifications that spans the school, vocational education and training and the higher education sectors and includes qualifications from senior secondary levels through to doctoral degrees. While I examine the AQF, similar frameworks either exist or are under development in and between many countries (see, for example, European Commission 2008). Such frameworks have two broad purposes. One is to enable qualifications to be compared (perhaps between countries) on a common basis to facilitate labour mobility, and this is a major thrust of the European Qualifications Framework. Such frameworks also facilitate the movement of learners between qualifications. In particular, they provide a basis for crediting learning at one level towards a qualification at another level and facilitate occupational mobility within industries. For example, an initial laboratory technician qualification can be upgraded through further study to a professional, say industrial chemist, level. Thus, qualifications frameworks that include recognition of learning achievements, and which therefore depend on compatible assessment models, can facilitate mobility between TVET and

higher education. This is a relatively underutilised feature of these structures, although there are some examples of good practice in Australia (Curtis 2009).

Such opportunities are valuable for individuals, but they are also important in economies in which the projected level of skills in demand is increasing (Karmel 2007; Shah & Burke 2006; Skills Australia 2010). In order to support mobility between qualifications, it is necessary to have a basis for comparing learning outcomes, and the qualification frameworks need to go beyond mere categories of qualifications and identify the learning that underlies each qualification. At a systems level, this learning needs to be certified and this requires reliable summative assessment and reporting. However, if we consider learner mobility, the assessment standards that underlie the framework must enable learners to understand where they are located on a learning continuum that spans related qualifications. It is in this respect that the specification of learning outcomes standards becomes part of the environmental context that either facilitates or impedes learner mobility.

Training packages

Training packages are not curriculum documents in that they do not specify any particular teaching or learning approach. These are matters for training providers. Training packages are developed by industry skills councils and specify the qualifications that may be offered in that industry area and the content and assessment requirements of the units of competency constituting each qualification. Within each unit, standards of competency and assessment guidelines are specified. Despite their limited intention, training packages are taken as prescriptive and treated as curriculum documents and they constrain the pedagogical practices of TVET teachers, notwithstanding recommendations that they be used in more flexible ways (Schofield & McDonald 2004).

Competency-based training and assessment

A core feature of training packages and indeed of the delivery of almost all TVET programs in Australia is the use of competency-based training and assessment. The use of competency-based training and assessment continues to be contentious. Of course, one would not want to argue against the notion that learners, upon graduation, should be competent practitioners. However, how competence is conceived, what constitutes competence, and how it can be developed are all extensively debated. Sadly, the discussion of these issues in the literature of cognitive psychology seems to find little purchase in the debates of TVET literature (see, for example, Eraut [1994] but the literature is extensive).

Considerable debate has taken place over the construction of qualifications as discrete units of competency, and the critique that the atomisation of competent performance into discrete units is inconsistent with broader notions of competence is a common theme (Eraut 1998, pp.130–5). A second criticism is the separation of doing from knowing. In recent discussions around competence in the school and higher education sectors, the phrase what a person ‘knows and can do’ is common. It is applied, for example, in the development of professional standards for teachers (Teaching Australia 2010).

Wheelahan (2009) argues that competency-based training and assessment, as it is practised in Australia, disenfranchises learners because it separates doing from knowing and localises doing to narrow situation-dependent contexts. Her emphasis is on the impact of competency-based training and assessment on individuals, but a wider impact is also of concern. If training is restricted by narrow situations (perhaps firm-specific ones), learners may not develop the capacity to transfer and adapt as new situations arise. In Australia, we have seen very substantial restructuring of industry that has left many workers behind. In some cases, this arises through the almost complete loss of industries (for example, footwear manufacturing), while in other cases, change that occurs through innovation or the deployment of new technologies can lead to redundancies. At the individual level, we may regard redundancy as an indicator of the lack of relevant human capital. Human capital is acquired through formal education and training (and through informal means), and may be enhanced through its application in workplaces or through ongoing development, or it

may depreciate through the lack of application or because of a changing context. One of the themes of the paper is the ecological relationship between individuals and their environments and the various agents of change that influence both individuals and their environments. Individuals may be very well adapted to a situation, but at the same time may not be adaptable. Training systems must address both characteristics. We want workers to be well adapted and therefore efficient in the situations that apply now, but we must also be conscious of the certainty of change and therefore of the need for adaptability. A system that focuses upon a demonstrated ability to perform specified actions in specific situations without attending to the knowledge and understanding that underpins that action renders learners poorly equipped to adapt.

Competency-based assessment in Australian VET results in one of two grades being awarded at the level of a unit of competency, namely, competent or not-yet-competent. But competency can be reported using much finer distinctions. Considerable work has been undertaken on graded assessment in Australian VET (see, for example, Griffin, Gillis & Calvitto 2007). It is worth noting that grades are assigned, beginning with competent and recognising levels of performance above that benchmark. A complementary approach has been developed in speech pathology. McAllister (2005), working on a project sponsored by Speech Pathology Australia, developed a set of performance levels that defined the generic professional skills required by speech therapists. However, in this case, competent performance for commencing professional practice is the highest, rather than the lowest, level in the sequence. The sequence begins with the performance expected of novices and scaffolds their development to the level of competence required for commencing professional practice. Learners and their teachers have a graded set of performance standards against which to monitor performance on the journey from novice to competent practitioner. Such standards-referenced levels may usefully inform the application of formative assessment methods with the standards being the reference points that both teachers and learners can use to identify the gap between current and desired performance and also become the basis for teacher-generated informative feedback directed towards bridging that gap.

There is more to competence than standards of competency. In the discussion above, I refer to competence and perceive it as operating on a continuum. The model of competence evident in competency-based training and assessment being used now focuses the attention of teachers and learners on a single benchmark level of competence. It portrays competence as a binary entity—either you have it or you do not. Traditional conceptions of competence are much richer than this, and we have lost some of this richness. Ray (2001) shows that in the traditional UK model of apprenticeships, five stages of emerging capability were recognised; namely, novice, apprentice, journeyman, tradesman and master. Dreyfus and Dreyfus (1984) similarly recognise five stages: novice, advanced beginner, competent, proficient and expert. The so-called novice–expert continuum has been studied extensively by many educational psychologists. Much of that work has focused on identifying differences between novices and experts and then trying to bridge the gap; that is, the two states were treated as separate conditions and not as a continuum (Alexander 2003). Lajoie (2003) refers to the value of ‘dynamic assessment’ and says ‘... assessment can reduce the time it takes to become competent if coupled with effective feedback and practice opportunities for learner engagement in realistic contexts’ (p.23). Clearly, Lajoie’s reference to dynamic assessment evokes the attributes that Black and Wiliam (1998) ascribe to formative assessment. Lajoie’s reference to ‘realistic contexts’ makes her proposal particularly relevant to TVET and this issue is discussed below.

The Australian Quality Training Framework

The Australian Quality Training Framework is a set of standards that describe the requirements that organisations must meet in order to become recognised providers of nationally accredited training. The current version of those standards (Australian Quality Training Framework 2007) includes a set of performance indicators used in evaluating the quality of provision by registered training organisations. These indicators include measures of competency completions, measures of learner

satisfaction and engagement, and measures of employer satisfaction with the training system. These standards meet the information needs of various stakeholders, who include:

industry, government and regulatory agencies who are interested in the effectiveness of the training system

employers who are interested in the quality of provision by particular providers from whom they may purchase training services

potential learners who are interested in the relative merits of alternative providers

providers themselves who may wish to use information derived from these indicators in their internal quality-improvement processes.

The National Centre for Vocational Education Research (NCVER) is currently undertaking an exploration of performance indicators that may provide information for each of these stakeholder groups. Among the indicators being examined are competency completion measures, an overall learner satisfaction scale and satisfaction with teaching, assessment and learning outcomes scales. The publication of these indicators may drive providers to evaluate their approaches to delivery, and if the indicators are the right ones, they should lead to improvements in the TVET system. In particular, the satisfaction with teaching, assessment and learning should focus the attention of providers onto these issues and should lead them to evaluate these aspects of their programs. We are finding significant differences between providers on these indicators. Here, I argue that engaging learners with assessment standards and practices through explicit attention to formative assessment methods should lead directly to improvements in students' perceptions of the quality of assessment and should indirectly result in improvements in learning outcomes and satisfaction with teaching. It is here, however, that we move out of the realm of evidence and into that of speculation, as there are no systematic data in TVET on changes in learner perceptions following the adoption of formative approaches to assessment. Nonetheless, the imposition of quality frameworks has the potential to influence the way in which teaching, learning and assessment are conducted in TVET.

System-level characteristics may either promote or inhibit assessment practices that are conducive to effective and sustainable learning.

A typology of assessment methods

I identify three main purposes for assessment, namely, promoting learning, measuring individual achievement, and evaluating programs. It is useful to explore assessment options in TVET that can satisfy each of these purposes. Table 1 shows four broad categories of assessment types against criteria conventionally used to judge the suitability of forms of assessment to identified purposes. The criteria conventionally used are: validity, reliability, objectivity, and feasibility. I take authenticity, a particularly important aspect of teaching, learning and assessment in TVET, to be a facet of validity. In addition, any influence of assessment methods on teaching and learning—backwash effects—need to be considered. Backwash effects may be favourable and enhance learning or may detract from the quality of learning. Since I argue that assessment is a driver of teaching and learning, positive backwash effects must be sought. This is the basis of much of the argument for formative assessment, whereby learner engagement in assessment processes and standards through self-assessment and the provision of explicit and informative feedback are expected to lead to sustainable learning. The destructive effects of some forms of assessment are also identified. For example, a focus on low-level knowledge recall leads students to adopt surface approaches to learning (Resnick 2010).

The four general categories of assessment, standardised assessment, the use of common assessment tasks, performance assessment, and portfolio construction are described briefly.

Table 1 Summary of the application of evaluation criteria to prospective assessment methods

Criteria for evaluating assessment methods	Assessment methods			
	Standardised assessment	Common assessment tasks	Performance assessment	Portfolio construction
Validity	<p>The validity of standardised assessment is questioned. Tasks are largely pencil-and-paper ones that may not enable all aspects of the construct to be assessed.</p> <p>This assessment format does not permit contextual assessment of problem-solving, as tasks are common to all students, irrespective of their courses.</p>	<p>The careful selection of tasks so that they provide opportunities for students to apply and develop their skills can enhance the validity of this assessment method.</p> <p>Tasks that are common across a range of courses would compromise the authenticity.</p>	<p>The method needs to have high validity, provided tasks are selected or constructed to provide opportunities to develop and apply the target skill.</p> <p>The method gives rise to authentic assessment, as the target construct is being developed in the context of the students' disciplinary or vocational domain.</p>	<p>The validity is called into question because of the likely attention paid to the portfolio as a product rather than directed to the target construct.</p>
Reliability	<p>This assessment method normally has the highest level of reliability. Careful attention to item design followed by piloting normally leads to highly reliable assessment.</p>	<p>Assessment is shown to have low reliability at the individual level unless many tasks are provided. However, as the number of tasks is increased, the cost of administering and rating student performance rises sharply.</p>	<p>The method is unlikely to have the reliability of the numerous small-scale tasks typically administered in standardised tests. There is likely to be some variation in the affordances of the tasks for the target construct (problem-solving) and this reduces the reliability.</p>	<p>Reliability is likely to be quite low. This problem can be addressed by developing rubrics for aspects of the portfolio.</p>
Objectivity	<p>Such tests are highly objective. No student is either advantaged or disadvantaged by the test, and objectivity can be verified through pilot testing, in which any systematic bias in items can be detected and rectified.</p>	<p>Well-developed scoring rubrics accompanied by the use of multiple raters, at least for samples of scripts, can reduce any lack of objectivity. Methods are available to control for systematic differences between raters.</p>	<p>Objectivity is likely to be medium to high, provided well-developed standards are established.</p>	<p>This assessment model is unlikely to yield objective assessment as each student produces a unique assemblage of evidence to be judged.</p>
Feasibility	<p>On a large scale, this assessment approach is feasible.</p> <p>Establishment costs are high, as the infrastructure for the testing needs to be established.</p> <p>The design and development costs of assessments are high, but once these costs have been met, large-scale assessment can be conducted at a low unit cost.</p>	<p>The method is feasible. If the intention is to sample population performance, this method is feasible. If the number of tasks has to be increased to provide more reliable estimates of individuals' performances, the costs rise sharply and the method becomes more difficult logistically.</p>	<p>Review of existing tasks is necessary. If they do not provide opportunities for the application of target knowledge and skills, they do not enable their development. New tasks may need to be developed.</p>	<p>Simply developing a portfolio as a vehicle to assemble and present evidence imposes a low load on schools. If the portfolios are to be judged, the assessment load is likely to be quite high.</p>

Criteria for evaluating assessment methods	Assessment methods			
	Standardised assessment	Common assessment tasks	Performance assessment	Portfolio construction
Other considerations	<p>It is 'objective' and provides a basis for comparison.</p> <p>This assessment method does not provide opportunities for feedback, nor for self-assessment. A low level of learner engagement in these tasks is anticipated.</p> <p>No backwash effect is expected to arise from this method of assessment.</p>	<p>The method can provide feedback to learners as scripts are returned. The feedback is unlikely to be immediate.</p> <p>Self-assessment can be used in that students can be asked to assess their work using the same criteria as raters prior to submitting their work for assessment.</p> <p>There is some doubt that any backwash effect can influence the development of the target skill in the students' primary learning domain.</p>	<p>The use of established tasks enhances the authenticity of the assessment. They are amenable to both self-assessment and frequent feedback. Backwash effects may be substantial.</p>	<p>Self-assessment is implicit in this model because students need to make judgments about what to include and how to relate what is presented to target constructs.</p> <p>The method is likely to have very little impact on programs offered to enhance learning of the target constructs.</p>

Standardised assessment

Standardised assessment is exemplified in large-scale assessment programs such as Programme for International Student Assessment (PISA). Multiple-choice response formats are most common as they enable automated marking. Most large-scale assessments are very carefully constructed and have very high reliability. They are often criticised for lacking validity and authenticity. They are particularly useful for testing knowledge recall, but when carefully constructed with appropriately chosen distracters, they can be used to diagnose defective processes (skills) and understanding. Although good multiple-choice tests are expensive to develop, they are very efficient and, therefore, feasible on a large scale. They lack the potential for feedback and immediate backwash effects are likely to be negative, although in systems in which learning is poorly coupled with expected outcomes, such assessment may force providers to focus on what is tested.

This form of assessment is well suited to evaluating student learning. This applies at the level of individual learners, for example, for licensing purposes, and the data can be aggregated readily to provider and system levels (as they are in large-scale national and international testing) and therefore contribute to program and provider evaluation for quality assurance purposes.

Common assessment tasks

Common assessment tasks are used in systems where there is a desire to enhance the authenticity of assessment, but to preserve sufficient commonality across tasks to enable comparisons to be made between individuals and providers. Well-chosen tasks can meet the authenticity objective, but their development, along with assessment criteria and guidelines, can be expensive. The reliability of this method depends on the tasks, the performance standards that are established and the training of teachers to provide consistent ratings of performance. This is assured through moderation and cross-marking arrangements. A backwash effect can be expected, depending upon the way in which the tasks are administered. There is a danger that teaching could be directed to those specific tasks and therefore ignore other aspects of the intended curriculum.

This method is suited for promoting learning, but less so for certifying individual achievement or program and provider evaluation.

Performance assessment

Performance assessment has a long history in some areas of the arts, for example, music and dance, but it also has considerable potential in TVET contexts. Performance assessments depend on the selection of tasks. Validity and authenticity of chosen tasks are expected to be very high, but reliability depends upon the clear specification of performance standards and on training teachers to rate performances consistently. Because of its immediacy, performance assessment has the potential for substantial positive backwash effects.

This technique is well suited for promoting learning, but not for program and provider evaluation.

Portfolio construction

Portfolio construction is a problematic form of assessment. It is a traditional way of assembling evidence of performance in certain fields, such as art, architecture and design and has become more common in other fields. I distinguish the construction of a portfolio from the activities represented by the assembled artefacts. Research on portfolios (see, for example, Troper & Smith 1997) shows they have low reliability and that their assessment is very time-intensive and this compromises the feasibility of their use. There is little evidence that they contribute to learning and they cannot be used on a large scale. However, in those areas where they have been used extensively, they have high-stakes consequences, for example, for selection into courses and employment.

A case study of formative assessment and learning

In this section, I briefly outline a project that was undertaken to test some of the propositions surrounding formative assessment and learning gains. This work was undertaken in two stages. The first stage was conducted in an electronics technicians' course and a business studies course in a TVET context (Curtis & Denton 2003). The second phase was conducted in a university in an electrical engineering course (Curtis 2010).

The overall objective of the study was to use key aspects of formative assessment, namely, self-assessment with extensive feedback, to evaluate their influence on students' acquisition of a generic skill—problem-solving. We developed a structured assessment tool that was based on a particular process model of problem-solving (after Bransford & Stein 1993). For each of the five main problem-solving processes, a set of indicators was identified and, for each indicator, from two to four performance levels were described. These were incorporated into an assessment instrument—the Problem Solving Assessment (PSA). Students undertook the tasks that were routinely required in their course and when they submitted their work, they also submitted a completed PSA, in which they indicated the level at which they believed they had exercised their problem-solving skills. They also recorded evidence for that judgment. When their lecturers graded the substantive assessment task, they also looked at the self-assessed problem-solving performance and at the evidence cited by learners, made judgments about the performance based on that evidence, and provided feedback about the learner's performance. We referred to this method as self-assessment followed by lecturer validation. The lecturer's grades were recoded in the student records system. The study was designed so that students undertook several assessment tasks using this method.

From this phase of the study, we found that the assessment tool was a useful one, although we made minor revisions to it for the second study. We intended to use the instrument to generate scores of sufficient reliability to warrant reporting in a nationally recognised qualification. The reliability of the scale is sufficiently robust to warrant such reporting. (The Cronbach alpha for the scale is 0.80.)

In a qualitative evaluation we found that students gained from explicit instruction and involvement in the assessment process. Sample comments included:

Very helpful in breaking down the problem-solving method into identifiable sections. Helped me better understand the process.

I think the key competencies is a good way of being recognised for things that you do but are not necessarily recognised in any other way.

K C [key competencies] not only make future job employers aware of your skills but it makes you aware of your skills and gives you the extra confidence in a job interview situation.

The second phase of the study used the same approach as the first and operated over two semesters of an academic year. Data on students' performances on three sequential assessments tasks are available for analysis. Students' self-assessed scores and teachers' scores were recorded, along with background information on the students and a set of standardised test scores. Multilevel regression modelling was used to evaluate change in performance over time while controlling for student characteristics.

Several characteristics of the study can be noted.

The target construct is articulated in terms of a set of processes and these are informed by a very substantial body of theory.

The assessment of a target skill is embedded within routine content-based assessment tasks.

The Structure of the Observed Learning Outcome (SOLO) taxonomy (Biggs & Collis 1982) is used as a framework for specifying performance standards.

A strong measurement approach is taken to evaluating the validity and reliability of the assessment tool.

The assessment tool is used in a way that requires learner engagement. It directs their attention to salient features of their work.

Teacher feedback is required, although we have no guarantee that learners attend to this feedback.

Teachers and learners use the same instrument to grade the work, although in our second study we focused on learners' own grades.

The assessment tool is not a rubric; that is, it is a generic tool that can be used across a variety of tasks rather than being specific to a particular task. Considerable work has been done on rubrics and compromise is required between the specificity of rubrics to particular tasks and their generality for promoting sustainable learning. Rubrics that are highly specific are effective summative assessment tools and may be used formatively, but they run the risk of generating instrumentalist approaches to assessment by both teachers and learners and subverting the formative learning intentions behind their use (Davies & Ecclestone 2008; Torrance et al. 2005). We used the SOLO taxonomy, a cognitive developmental framework, to describe increasing levels of complexity in learners' responses to assessment tasks. The generality of this tool means that feedback based on one assignment can be applied to a later, different assessment task.

The issue of assigning grades is raised by Black and Wiliam in their 1998 review. They present evidence to show that grading may subvert the learning intentions of formative approaches as it leads students to take an achieving (or instrumental), rather than a deep, approach to their learning (Biggs & Moore 1993). This finding indicates that a single assessment task cannot meet both learning and certification assessment objectives. We did not test this possibility as marks were given for all pieces of student work.

The main finding is that repeated assessment of problem-solving, which incorporates self-assessment and which is accompanied by teacher feedback, leads to significant growth in performance over time. The features of the assessment model that we believe are important in this improvement are the use of an iterative (repeated) model of assessment, in which students undertook assigned tasks, assessed

their own work, received feedback on that work and, we expected, used that feedback in subsequent tasks. Thus there is an assessment tool that is embedded in a formative assessment process, and we believe that both are important in promoting learning.

This model of assessment makes certain assumptions about learning. Foremost, it is consistent with a cognitive view of learning, in which learners build complex knowledge structures and whereby, as expertise emerges, those structures become extensively interconnected and conditioned by experience. It is not consistent with behaviourist theories of learning, and it is worth noting that behaviourist models are consistent with the simple notion of competency embedded in competency-based training and assessment.

Implications for TVET policy and practice

A considerable body of research is being assembled on assessment methods and, in particular, on formative assessment. There is strong evidence to indicate that the use of formative assessment can lead to substantial learning gains. Much of this evidence originates in the school sector, some in higher education, but little in TVET. What little research has been conducted in TVET is consistent with the findings from other sectors.

Two questions arise from the foregoing discussion.

What are the implications of emerging views on learning and assessment for teaching practices in TVET systems?

What are the implications for TVET policy development?

Implications for TVET practice

A diversity of assessment methods is required to achieve the multiple purposes and to satisfy the requirements of the many stakeholders of TVET systems.

Certification, licensing and selection require highly reliable assessment methods. In general, these purposes operate on a large scale, so efficiency is also a core requirement. Other criteria for the evaluation of assessment methods tend to be less important. For these purposes, standardised assessment methods tend to be used.

Program and provider evaluation require data that are reliable at the individual level and which can be aggregated readily to program, provider and system levels. Thus, standardised assessment methods are used.

Formative practices can lead to improved learning, but those gains are not automatic. It is possible to implement features of formative assessment practices—for example, self-assessment and feedback—but if the self-assessment uses simple rubrics that do not require deep engagement in evaluation or if teacher feedback is not diagnostic and informative, the promise of formative practices may not be realised.

The use of formative practices may require the redevelopment of teaching resources to accommodate a focus on frequent assessment that is integrated with learning and which has a form likely to be unfamiliar to many TVET teachers and learners. Building a body of experienced practitioners will take time.

Implications for policy

The lack of high-quality research on formative assessment in TVET systems must be addressed. It would not be sensible to rush the implementation of formative assessment methods unless there is a body of evidence that shows that formative practices do lead to learning gains and that they are feasible in the variety of contexts in which TVET programs are delivered. These contexts include

institutional settings, workplace settings involving apprentices, and workplace-based learning models for experienced workers.

Some of the structural and regulatory frameworks in which TVET practice is embedded—for example, commitments to prescriptive training packages and a commitment to a limited conception of competence—work against the types of learning and therefore the types of teaching that are implicated in formative assessment methods. Quality assurance and regulatory arrangements may need to be adjusted to be consistent with the forms of learning and teaching that are implied in formative assessment methods. This may not require a substantial adjustment. If formative assessment does lead to the suggested learning gains, indicators that reflect quality teaching, quality learning and quality assessment will reflect those gains.

References

- Airasian, PW 1994, *Classroom assessment*, 2nd edn, McGraw Hill, New York.
- Alexander, PA 2003, 'The development of expertise: the journey from acclimation to proficiency', *Educational Researcher*, vol.32, no.8, pp.10–14.
- Allen, J & van der Velden, R 2005, *The flexible professional in the knowledge society: conceptual framework of the REFLEX project*, REFLEX working paper no.1, University of Maastricht, The Netherlands.
- Biggs, JB & Collis, KF 1982, *Evaluating the quality of learning: the SOLO taxonomy (Structure of the Observed Learning Outcome)*, Academic Press, New York.
- Biggs, JB & Moore, PJ 1993, *The process of learning*, 3rd edn, Prentice Hall, New York.
- Black, P & Wiliam, D 1998, 'Assessment and classroom learning', *Assessment in Education: Principles, Policy and Practice*, vol.5, no.1, pp.7–74.
- Boud, D 1995, 'Assessment and learning: contradictory or complementary?', in *Assessment for learning in higher education*, ed. P. Knight, Kogan Page, London, pp.35–48.
- 2002, 'Assessment: have we forgotten something?', in *Fresh thinking about learning and learners: a blue sky project*, ed. ANTA, Melbourne, pp.42–5.
- Boud, D & Hawke, G 2003, *Changing pedagogy: vocational learning and assessment*, viewed 2 June 2004, <http://www.oval.uts.edu.au/working_papers/2003WP/2003wp0317boud.pdf>.
- Bransford, JD & Stein, BS 1993, *The ideal problem solver*, WH Freeman, New York.
- Curtis, DD 2009, *Student transfer: At a glance*, NCVER, Adelaide.
- 2010, 'Defining, assessing and measuring generic competencies', PhD thesis, School of Education, Flinders University, Adelaide.
- Curtis, DD & Denton, R 2003, *The authentic performance-based assessment of problem-solving*, NCVER, Adelaide.
- Davies, J & Ecclestone, K 2008, "'Straightjacket" or "springboard for sustainable learning"? The implications of formative assessment practices in vocational learning cultures', *The Curriculum Journal*, vol.19, no.2, pp.71–86.
- Dreyfus, HL & Dreyfus, SE 1984, 'Putting computers in their proper place: analysis versus intuition in the classroom', *Teachers College Record*, vol.85, no.4, pp.578–601.
- Eraut, M 1994, *Developing professional knowledge and competence*, The Falmer Press, London.
- 1998, 'Concepts of competence', *Journal of Interprofessional Care*, vol.12, no.3, pp.127–39.
- European Commission 2008, *The European qualifications framework for lifelong learning*, European Commission, Luxembourg.
- Griffin, P, Gillis, S & Calvitto, L 2007, 'Standards-referenced assessment for vocational education and training in schools', *Australian Journal of Education*, vol.51, no.1, pp.19–38.
- Karmel, T 2007, 'The demand for tertiary education in Australia', paper presented at the 16th National VET Research Conference, viewed April 2010, <<http://www.ncver.edu.au/publications/1825.html>>.
- Karmel, T, Mlotkowski, P & Awodeyi, T 2008, *Is VET vocational? The relevance of training to the occupations of vocational education and training graduates*, NCVER, Adelaide.
- Lajoie, SP, 2003, 'Transitions and trajectories for studies of expertise', *Educational Researcher*, vol.32, no.8, pp.21–5.
- Looney, J 2008, *Teaching, learning and assessment for adults: improving foundation skills*, CERI/OECD, Paris.
- McAllister, SM 2005, 'Competency based assessment of speech pathology students' performance in the workplace', unpublished thesis, University of Sydney, Sydney.
- Pellegrino, J, Chudowsky, N & Glaser, R (eds) 2001, *Knowing what students know: the science and design of educational assessment*, a report of the National Research Council, National Academy Press, Washington, DC.
- Ramprasad, A 1983, 'On the definition of feedback', *Behavioural Science*, vol.28, pp.4–13.
- Ray, J 2001, *Apprenticeship in Australia: an historical snapshot*, NCVER, Adelaide.
- Resnick, LB 2010, 'Nested learning systems for the thinking curriculum', *Educational Researcher*, vol.39, no.3, pp.183–97.
- Sadler, DR 1989, 'Formative assessment and the design of instructional systems', *Instructional Science*, vol.18, no.2, pp.119–44.
- Schofield, K & McDonald, R 2004, *Moving on ... report of the high level review of training packages*, ANTA, Melbourne.

- Shah, C & Burke, G 2006, *Qualifications and the future labour market in Australia*, report commissioned by the National Training Reform Taskforce, Centre for the Economics of Education and Training, Melbourne.
- Skills Australia 2010, *Australian workforce futures: a national workforce development strategy*, Skills Australia, Melbourne.
- Teaching Australia 2010, *National professional standards for teachers* (draft), Teaching Australia, Sydney.
- Torrance, H, Colley, H, Garratt, D, Jarvis, J, Piper, H & Ecclestone, K 2005, *The impact of different modes of assessment on achievement and progress in the learning and skills sector*, Learning and Skills Research Centre, London.
- Troper, J & Smith, C 1997, 'Workplace readiness portfolios', in *Workforce readiness: competencies and assessment*, ed. HF O'Neil, Lawrence Erlbaum, Mahwah, NJ, pp.357–82.
- van der Velden, R & Wolbers, MH 2008, *A framework for monitoring transition systems*, OECD education working papers no.20, OECD Publishing, Paris.
- Wheelahan, L 2009, 'The problem with CBT (and why constructivism makes things worse)', *Journal of Education and Work*, vol.22, no.3, pp.227–42.
- Wiggins, GP 1998, *Educative assessment: designing assessment to inform and improve student performance*, Jossey-Bass, San Francisco.