Moderation of Assessments in Vocational Education and Training

Prepared by: Graham S. Maxwell
University of Queensland

Commissioned by: Queensland Government Department of Employment and Training
To a large extent, attempts to ensure the quality of assessment in vocational education and training have focused on the procedures used in the assessment process. Several recent reports, however, including those by Schofield (1999) and Smith (2000) in Queensland, have highlighted the need to address the quality and consistency of the judgments made by assessors, not just the nature of the procedures they follow.

Moderation refers to a process of comparing the assessment judgments in relation to the same learning outcomes of different assessors in a variety of what might be quite different contexts. The purpose of moderation is to ensure comparability of assessment. It is, therefore, a process that can actively address the important issues of consistency of interpretation and consistency of implementation of competency standards in VET.

To date, no formal and systemic process for comparing the assessment judgments of VET assessors has been implemented in this country. This is perhaps understandable given the enormity and complexity of introducing a new approach to training in Australia over the last decade. The time has come, however, to begin thinking of ways in which those implementing the system can ensure the consistency and quality of their assessment judgments.

This report by Dr Maxwell canvasses what might be possible as well as what might be inadvisable in terms of approaches to moderating assessment judgments in VET. It provides a starting point for discussion and debate about how a simple, non-obtrusive, cost-effective yet publicly defensible process of moderation might be implemented in this state.

I would encourage you to consider providing the Department with feedback on the ideas proposed by Dr Maxwell, as well as any other thoughts you might have regarding processes for ensuring the quality and consistency of VET assessment. Feedback should be directed by 14 May 2001 to:

Ms Sharon Mullins, Senior Policy Officer,
Department of Employment and Training
Postal address:  LMB 527,  GPO, Brisbane.  Q 4001
Telephone: (07) 32371380
Email: sharon.mullins@detir.qld.gov.au

Ken Smith
Director-General
Department of Employment and Training
In a study of assessment practices in vocational education and training, *Issues impacting on the quality of assessment in vocational education and training in Queensland*, Smith (2000) identified several matters requiring urgent attention. His main finding was that there is inadequate support for the actual doing of assessment (the how) as distinct from general principles for conducting assessment (the what) and that consequently the quality of much assessment (and the associated training) is of poor quality and doubtful validity. It was found that trainers and assessors are concerned about this situation. This is not the first indication of problems with the quality of assessment. Schofield (1999) also concluded that the quality of assessment needs attention, as too do Clayton and House (2000). It is likely that unless the quality of assessment is improved the integrity of training qualifications will be threatened as the inadequacies and inconsistencies become more pervasive and more noticeable.

Among his recommendations Smith (2000) listed the following (recommendation 6):

- That DET develop and introduce a moderation system for publicly validating assessment processes and outcomes against benchmark standards and practices.
- That the moderation system should be as simple and non-intrusive in operation as possible yet sufficiently rigorous and comprehensive in its processes to allow valid public reporting.
- That the moderation process should be designed so as to facilitate the sharing of ‘good practice’ approaches to assessment across the system.

This recommendation was the starting point for this report. That is, the initial question was how might such a moderation system be implemented. However, a consideration of research and practice in vocational education and training in Australia and other countries indicates that the question could be more broadly stated as how might an assessment quality control process be implemented. This reframing can be seen as either defining what moderation means (‘an assessment quality control process’) or situating it within a more extended framework for validating the quality of assessments of vocational competencies (which might be termed more broadly ‘quality management’).

**TERMINOLOGY**

Terms can have different meanings in different contexts and this often creates confusion. It is necessary to clarify how terms are being used in this discussion.

‘Quality assurance’ refers to procedures for providing assurance that appropriate procedures are being adopted, directed at guaranteeing that particular standards will be met. Usually this is by self report that the procedures, and the resources to support those procedures, are in place. ‘Under the Australian Vocational Education and Training system, quality assurance is the responsibility of State and Territory training authorities, and quality assurance requirements are included in the criteria which each training provider must meet’ (Robinson, 2000, p.28). This includes quality assurance for
assessment. The assumption in the case of assessment is that if assessors possess appropriate knowledge and competencies, and follow defined procedures, they will be able to conduct valid and reliable assessments without further monitoring. This assumption is now being called in question. At issue is whether ‘in advance’ quality assurance of this kind is sufficient to guarantee that there is consistency of interpretation of the competency standards and also in the resulting judgements of competence. Quality assurance is a ‘feed-forward’ mechanism in the sense that the application of well-defined procedures is expected to deliver the desired outcomes. However, assessment procedures are not well-defined since they must be tailored to particular situations and contexts. Also, the judgement of competence is itself not capable of being explicitly defined so that consistency is delivered automatically (Cresswell, 2000).

‘Quality control’ (or ‘quality management’) refers to procedures adopted to monitor and endorse, and where necessary to adjust or correct, the actual implementation of some activity while it is occurring and before it is completed. In the case of VET assessments, this would still require properly trained and professionally committed assessors. But it assumes that the actual assessment procedures and judgements need to be checked in some way to ensure that the assessment procedures are in fact appropriate and there is consistency in the resulting judgements of competence. Research on assessment shows that different people do not interpret written standards as having similar meanings, often not even closely, and that the enactment of competency statements is by no means unproblematic. It can be expected that assessments of competence will differ widely, not through wilfulness or incompetence but because understandings of competence are not consistent. Quality control is a ‘feed-back’ mechanism that sees ‘monitoring while in progress’ as important for keeping procedures ‘on track’ and being confident that desired outcomes are being delivered.1

The Training Package for Assessment and Workplace Training (BSZ98) defines quality assurance as: ‘a planned and systematic process of ensuring that the requirements of the assessment system, competency standards and any other criteria are applied in a consistent manner’ (p.136). However, no guidelines are provided for undertaking such quality assurance. Also, such a process of ‘ensurance’ rather than ‘assurance’ is properly referred to as ‘quality control’ in the sense previously defined. Quality assurance focuses on whether the proper procedures are in place for an expected outcome whereas quality control focuses on whether the outcomes themselves are satisfactory.

‘Moderation’ refers to a particular process of quality control involving the monitoring and approval of assessment procedures and judgements to ensure there is consistency in the interpretation and application of the performance standards. Typically, this involves a single person (a moderator) or a group of people (a moderation panel) looking at samples of evidence of student performance and determining whether they agree with the

1 Houston (1999) also uses these terms as here. However, the terminology is contentious. Some would see ‘quality control’ as a centralised bureaucratic process and ‘quality assurance’ as a more encompassing term covering both feed-forward and feed-back processes. The terminology is less important than the concepts to which they refer. Here, quality control is taken to involve direct management of outcomes, particularly the process of assessor judgement at the point of judgement.
assessments of the assessor. Feedback to the assessor may include advice on improving their assessment procedures or evidence gathering and adjusting their assessment standards, that is, modifying their judgement of what constitutes competence. Moderation can have a collegial orientation of support rather than a managerial orientation of control, that is, moderation can be seen more broadly as a form of quality management. Every moderation system needs to determine where the ultimate authority lies for approval, what is an appropriate balance of rights and powers of all participants, and how differences of opinion can be resolved.

‘Verification’ has similar connotations to ‘moderation’. The Training Package for Assessment and Workplace Training (BSZ98) defines verification as: ‘the means of ensuring that the assessment decision is consistent and reliable [such as] having another assessor(s) confirm the assessment decisions’ (p.137). However, no further explanation is offered about how this might be managed and no systems for doing this have been put in place. It seems to inhabit the realm of ‘a good idea’ but to be entirely optional. In England, verification is a mandatory process through the use of internal and external ‘verifiers’. Recently, concerns have been raised about the formalistic character of this process and suggestions made for modification. There is a strong tradition in England of external control which has not been part of the Australian tradition.

In Australia, the term ‘moderation’ carries strong connotations of procedures adopted within schools, particularly in the senior secondary school and particularly in Queensland where there is a 30 year tradition of ‘moderated school-based assessment’, and it may be confusing to use the same term to apply to somewhat different quality control procedures in VET. The moderation procedures used in schools are unlikely to be directly applicable within VET. ‘Verification’ is an alternative term. Here, it does not carry the same connotations as it does in England and perhaps could be invested with the particular meanings of a quality control system appropriate for VET in Australia. It has the advantage of directly indicating the purpose of the process to which it refers, whereas ‘moderation’ does not. Another alternative is ‘authentication’. However, ‘moderation’ is the more general and more common term.

The Training Package for Assessment and Workplace Training (BSZ98), also refers to ‘monitoring and review’ where review is defined as ‘planned and systematic analysis of the assessment system’ (p.137), directed at improving that system in its future operation. This definition focuses on the system procedures rather than the actual assessment judgements, as the competency standards for the unit titled ‘Review Assessment’ demonstrate (and where the notion of monitoring is dropped altogether). A properly functioning process of verification might form an important part of a broader process of monitoring and review or might largely replace it. In fact, ‘monitoring and review’ is represented as verification in another part of the training package where, in part explanation of ‘reliability’, it says that: ‘assessment practices … need to be monitored and reviewed to ensure consistency of judgement’ (p.18, italics added) but this is not further developed, so it remains a rhetorical statement rather than an implementable expectation.
It is possible to see all of these processes as part of an overall quality management system covering the periods before, during and after the assessment judgements. Quality assurance occurs before, quality control occurs during, and review occurs after the assessment judgements are made. Quality assurance is concerned with establishing appropriate circumstances for assessment to take place. Quality control is concerned with verifying that assessment procedures and judgements are appropriate (and ‘approving’ them), and review is concerned with retrospective analysis to see whether improvements in assessment procedures and judgements should be made in the further.

THE PRESENT SITUATION

Under the Australian Qualification Framework (AQF), competency standards for the award of recognised qualifications are defined through training packages developed by the appropriate Industry Training Advisory Body (ITAB) (see AQF, 2000). Training packages include endorsed assessment guidelines along with the competency standards.

In addition, a set of twelve National Assessment Principles has been adopted within the Australian Recognition Framework (ARF) (see ANTA, 2000) as the basis of development of assessment systems and practices. The following principles are relevant to the current discussion:

- **Principle 4:** Assessment should be undertaken by or quality endorsed by a Registered Training Organisation (RTO).
- **Principle 5:** Assessment for national recognition purposes shall be conducted within a quality assurance framework.
- **Principle 6:** Responsibility for assessment resides with the body that issues the qualification under the Australian Qualification Framework.
- **Principle 7:** The assessment process shall be valid, reliable, flexible and fair.
- **Principle 11:** Assessment systems should incorporate ongoing monitoring and review processes.

These principles place the responsibility for assuring the quality of assessment processes and judgements with the RTOs. This would seem to impose some constraints on the development of any system of statewide moderation, as noted by Clayton and House (2000).

The requirement that assessment be conducted within a quality assurance framework clearly emphasises the formal requirements for assessment, such as trained personnel, appropriate resources, appropriate assessment practices, keeping accurate records and ongoing review. Adoption of appropriate assessment practices means implementation of the endorsed assessment guidelines, which is assumed to be an unproblematic process.
Training for assessors follows the same approach as for other qualifications within the AQF. Competency standards are defined within the training package for Assessment and Workplace Training (Certificate IV and Diploma levels). As with all competencies within the AQF, what is assumed (or believed) is that:

- the defined competencies are sufficient for the demands of the job;
- the assessment of these competencies is to a consistent standard;
- the competencies are transferable to new situations;
- the competencies will continue to be displayed.

These assumptions (or beliefs) may be warranted. However, in the case of assessment, the following difficulties are apparent:

- training packages provide very little procedural assistance;
- competencies are not (and cannot be) stated in unambiguous terms;
- judgements of competency standards require personal interpretation;
- different assessors can apply competency standards differently;
- new situations can produce unusual cases and puzzling situations;
- understandings of competency standards need to be continually refreshed and updated.

Research has shown that a written statement of competence is unlikely to be interpreted and applied identically by different assessors. Written statements are necessarily ‘fuzzy’ in their meaning, can mean different things to people with different backgrounds and different knowledge, require different interpretations for different situations, and must be applied to performances which are necessarily themselves different in different settings. Common interpretations of written competency standards require a process of sharing among assessors focusing on discussion and agreement on performances that do and do not exhibit the competency standards. Such discussions must include consideration of actual examples not just discussion about the standards. This process of sharing is necessary for the development of common understandings (Sadler, 1987; Eraut, 1994). Quality assurance procedures can support such a process but do not fully compensate for its absence.

**ASSESSMENT MODERATION**

Moderation usually means ‘social moderation’ in the sense used by Linn (1996). That is, it involves comparisons of the assessment judgements of different assessors in different settings but all relating to the same learning outcomes (or competency standards), with the purpose of ensuring that the judgements are comparable. Comparability of assessment judgements means that there is agreement that the assessed performances are appropriately classified in terms of the standard they demonstrate. This involves both similar interpretation of the standards and similar recognition of performances that demonstrate those standards. A moderation process is therefore one involving approval of assessor judgements, with the implication that there may need to be some adjustment of
those judgements to conform to the common standard. It is not a passive process that simply checks how much agreement there is but an active process in which assessment judgements are aligned with each other to create consistency of interpretation and implementation of standards across the whole system. Differences of opinion therefore must be resolved rather than simply noted and accepted.

Moderation systems can differ in style and complexity. Four basic approaches can be identified, under two general headings, as shown in table 1.

**Table 1: Types of moderation systems**

<table>
<thead>
<tr>
<th>Strong control; high stakes</th>
<th>Weak control; low stakes</th>
</tr>
</thead>
<tbody>
<tr>
<td>External moderator</td>
<td>Assessor meetings</td>
</tr>
<tr>
<td>External moderation panel(s)</td>
<td>Assessor partnerships</td>
</tr>
</tbody>
</table>

Strong control is necessary where the assessments involve high stakes. Weak control is more appropriate where the assessments are low stakes. ‘High stakes’ refers to situations where the consequences of obtaining (or not obtaining) the qualification (or relevant competencies) can be considered serious (for the trainee, the RTO or the employer). ‘Low stakes’ refers to situations where the consequences are less serious. Clearly, the distinction is not objectively categorical; rather there is a gradation from higher to lower stakes and the distinction is somewhat subjective. Decisions about whether the stakes should be considered high or low requires judgement of the seriousness of the consequences in particular circumstances. For VET qualifications, serious consequences are those where a high degree of risk is involved in the exercise of the qualification (through a ‘licence to practice’). The determination of high stakes situations is therefore related to the matter of ‘risk assessment’ (see later).²

High stakes assessments typically demand a high degree of consistency in assessment judgements. Consistency is often seen as simply a matter of relative comparison, that is, whether there is common interpretation and implementation of the competency standards. However, for VET assessments, high stakes assessments also demand rigorous attention to the competency standards themselves. That is, there is an issue of absolute judgement concerning the standards not just an issue of relative comparability. It would not be satisfactory for all assessors to be in agreement but for the performance standard on which they agree to fall short of the standard necessary for the demands of the job, especially where issues of public safety are involved.

² Typical definitions of ‘high stakes’ refer to school or school system consequences, particularly where funding decisions are involved. Sometimes, they refer to personal consequences for the students, such as where life chances are affected, for example, where there is competitive selection for employment, promotion or further training. The argument here is concerned with different actions relating to high stakes and low stakes rather than different ways of determining high stakes and low stakes situations.
Verification of Assessments in VET

Clearly, VET assessments differ in terms of whether they can be deemed high stakes or low stakes, depending on the industry and the qualification. Therefore, a single approach to moderation would seem neither appropriate nor necessary. Different procedures may be desirable for different qualifications. Stronger control is needed where the stakes are higher.

The following discussion provides further explanation of the forms of strong and weak control indicated in table 1. However, as will be seen later these are not exhaustive of the possibilities.

**External moderators**

External moderators offer strong external control. ‘External’ here means external to the school or training organisation. The external authority resides in a certifying agency which is therefore responsible for confirming or approving the assessments. The moderator is trained for the role and will assume responsibility for moderation of an appropriate number of assessors. This can be defined as a geographical area. There could be a number of external moderators for the state. Moderators’ activities typically involve review of student folios as well as site visitation for discussion with assessors and observation of assessment situations. Clearly they need to be experienced in training and assessment in the relevant qualifications.

The advantages of external moderators include:

- they can offer authoritative interpretations of competency standards;
- they can carry the standards from site to site and assessor to assessor;
- they can offer advice on assessment approaches and procedures;
- they can observe actual assessments not just view folios;
- they can be a trouble-shooting resource for assessors to draw on;
- they can induct novice assessors quickly into high quality assessment.

The disadvantages of external moderators include:

- there are substantial costs involved (salary, base office, travel accommodation, communications, training, moderator conferences, etc.);
- there are logistic problems to be overcome in covering all assessors adequately;
- ‘authoritative’ interpretations are not always right or appropriate;
- ‘external authority’ can be stultifying rather than liberating, encouraging conformity rather than innovation.

In the UK, external moderators have been used extensively throughout the education system, though not necessarily under that name. External moderators have been used for verifying ‘coursework assessment’ in schools; external examiners are used in universities to set and mark assessments; external verifiers are used in the VET sector to check the validity of vocational assessments. Although there are important differences among these three forms of external surveillance, they are all concerned with the quality of assessment processes and judgements. However, it must be noted that the UK has a stronger tradition
of such external surveillance in all sectors of education than has been the case in Australia. There have been no such uses of external moderators in Australia. Further, there is a different structure for certification in the VET sector in the UK, where awarding bodies are separate from the training organisations. In that situation, awarding bodies have a responsibility to verify the assessments they are certifying.

A new approach to the process of verification in the VET sector has been proposed and partially introduced in the UK. This new approach transforms all assessment situations into ‘professional conversations’. The essence of a professional conversation is more equal participation in deciding whether competency standards are being met, and this ‘professional conversation’ may be between teacher and student or between moderator and assessor (Devereux, 1997; 1999). This concept is discussed in greater detail later.³

**External moderation panels**

External moderation panels offer another form of external confirmation of assessment decisions by a certification agency. This is the approach taken in Queensland senior secondary schools where the certification agency is the Queensland Board of Senior Secondary School Studies (QBSSSS) and the student assessment is wholly school-based. Although external moderation panels can function in a variety of ways, what will be described here is the way that QBSSSS panels currently operate.

QBSSSS moderation panels, called review panels, consist of selected teachers, chosen for their experience and excellence in teaching and assessment. Their participation is voluntary and supported by their school. Most would see participation as a valuable professional development opportunity. It does not involve additional remuneration and the cost of involvement is borne by the schools.⁴ Panel membership is reconstituted every six years, involving replacement of at least one-third of the panel.

The state is divided into administrative districts. In each subject or subject area, there is a district review panel in each district plus a state review panel; some subjects with small enrolments have different arrangements. State review panels moderate the operation of the district review panels, advise on any unresolved issues within the district review panels, and resolve disagreements between schools and district review panels.

The moderation process is seen as consisting of the following components:

- accreditation of school plans for implementing the subject syllabus (a form of quality assurance);
- review of each school’s assessments through monitoring, verification and approval (see later);
- random sampling (a post-hoc mechanism for evaluating the effectiveness of the moderation process).

³ In the earlier book, Devereux (1997) uses the term ‘professional discussion’ but replaces this with ‘professional conversation’ in the later book (Devereux, 1999).
⁴ There are perceived benefits which may offset the costs for both the teachers and the schools. These include status, networking and professional development.
Verification of Assessments in VET

The review process forms the heart of these moderation procedures. Monitoring advises schools on the appropriateness of their assessment processes and standards at the end of year 11 (half way through the course). Verification advises schools towards the end of year 12 on the appropriateness of their judgements of standards of performance of their students. Approval is the final process of validation of the results to appear on the Senior Certificate.

The part of this process which is of most relevance here is the operation of the review panels during monitoring and verification. A sample of student folios is sent from each school in each subject to the relevant review panel. Each folio includes the relevant pieces of student work or other records of their performances as well as the school’s judgements of the standard reached by the student. Teachers’ judgements concerning how close each student is to the grade boundaries are also provided. Adjustment for any student in the sample can have repercussions for other students in the group.5

Members of the review panel meet and review each of the sampled folios, considering whether they agree with the teacher judgements of standards demonstrated by each student. The panel seeks agreement within the panel before offering its advice to the school. Where adjustments are recommended, the panel chair may enter into discussion and negotiation with the school. A process exists for appeals. Ultimately, the Board has the power to reject a school’s opinion. However, disagreements are negotiated at length and in good faith and almost always satisfactorily resolved.

There are, of course, many other features to the QBSSSS moderation process but these are the details most relevant to the issues being addressed here. Further details are found in the Moderation Handbook (QBSSSS, 1999).

It is important to notice the following features of QBSSSS panel-based moderation:

- the initial process involves a quality assurance approach (accreditation);
- the assessment judgement is a global judgement, that is, a synthesis of all the performance evidence in the student’s folio;
- advice is offered to schools mid-course not just at the end, which allows schools time to make adjustments to their assessment procedures and standards;
- a sample of student folios from each school is considered, not all students (except where there are very few students);
- review panels consider whether they agree with the school’s judgements of student achievement in full knowledge of those judgements (that is, the focus is on confirming or challenging the school’s assessment judgements, not on making an independent or ‘blind’ judgement which is then compared with the school’s);

5 Sample sizes are five for monitoring and nine for verification (or all students if there are fewer than this). Each sample covers the demonstrated range of achievement levels.
• advice is offered to schools on the quality of their assessments and the appropriateness of their judgements of standards;
• disagreements between a school and a panel are discussed and negotiated;
• the Board has ultimate power of approval and some things may be non-negotiable.

Modifications to this process would be needed if it were to be applied to VET assessments. The timing of completion of competencies and qualifications in VET is more staggered and more individualised. Review panels would have to operate on a more or less continuous basis to cover all groups of trainees. Alternatively, they would have to focus on monitoring rather than verification and approval. However, this would be more appropriate for low stakes situations (second column of table 1) since the role would be advisory rather than certifying.

Advantages and disadvantages of external moderation panels depend on how they are implemented. However, there are some general characteristics, especially when contrasted with external moderators.

The advantages of external moderation panels include:
- panels are likely to make more consistent decisions than individuals;
- panels can cover a larger group of assessors than a single moderator can;
- panels can offer more comprehensive advice on assessment approaches and procedures (by being able to draw on a wider cross-section of examples);
- panels represent collective authority rather than single person authority;
- they provide powerful professional development for those involved.

The disadvantages of external moderator panels include:
- there are substantial costs involved (travel, accommodation, communications, training, moderator conferences, etc.);
- there would need to be some fulltime officers to organise the process (so the costs would not be less and may be more than for external moderators);
- there are logistic problems to be overcome in covering all assessors adequately;
- panel activities take members away from their other work (and may require payment).

**Assessor meetings**

Assessor meetings require the participation of all assessors. These may occur within an institution or across several institutions. In Queensland senior secondary schools, it is assumed that there will be meetings of all teachers of a subject within a school at various stages of teaching and assessing to establish a process of within-school moderation. These stages might be before beginning teaching (to align their expectations of students), before undertaking an assessment (to discuss the criteria and standards for assessing performance), after undertaking an assessment (to discuss whether they have applied
similar standards and to make adjustments where necessary) and at the end of the course (to compare and confirm their final judgements of the exit standards reached by students).

At times also there have been meetings of teachers from several primary schools to compare their assessments of students and moderate their judgements against common benchmarks of learning outcomes. Clearly, under these conditions, the incentive to participate and to modify one’s own practice depends on professional motivation not on any sanctions. Even if such meetings are mandated, unless there is a certification function to which they contribute, personal engagement is dependent on teachers seeing that the meetings are helpful for improving their own practice rather than simply verifying whether they are adopting common standards. Again, though, this can be a powerful agency for raising the quality of teacher practice.

One requirement of assessor meetings is that all the participants are assessing the same outcomes. In the case of VET this would mean all assessing the same competency standards within the same qualification. This might be difficult to arrange if such assessors were widely dispersed. However, where such assessors can meet, there might be some mutual benefits. Assessor meetings could be conducted in a similar fashion to moderation panels or in more relaxed fashion focussing on sharing rather than approval.

The advantages of assessor meetings include:
- the opportunity for direct comparison and sharing among assessors;
- a less judgemental atmosphere than for external moderators or panels;
- personal ownership of any new insights and understandings and ideas;
- opportunity to develop networks of support for resolving new problems;
- powerful professional development for those involved;
- mandatory participation and public scrutiny of one’s own practices and judgements could encourage serious attention to the issue of quality;
- being a useful supplement to other quality control procedures.

The disadvantages of assessor meetings include:
- meetings would need to be organised and facilitated (with attendant costs);
- would assessors come to meetings if they were voluntary? (or alternatively, what sanctions could be used to encourage participation?);
- there would need to be some fulltime officers to organise the process (though the costs for this would probably be less than for moderators or panels);
- there are substantial logistic obstacles to covering all assessors;
- without a formal approval process, there would be no guarantee of quality outcomes;
- they would not satisfy the need for quality control in a high stakes situation.

Assessor partnerships

Assessor partnerships simply involve sharing of assessments within a small group of assessors, maybe between just two assessors. This is a reduced form of assessor meeting,
one where the public scrutiny is much reduced and the focus is on mutual assistance and confirmation.

Partnerships clearly have the potential to be of some assistance and are preferable to no action at all. However, external management is problematic. They must be self managed. But this means that some will not treat such partnerships seriously and they may well have little effect. Also, some will treat the process as an imposition and resist or subvert it. Requiring sign-off means that many would treat it as a bureaucratic procedure rather than something of real benefit.

Assessor partnerships are discussed more fully later under a framework of possible verification procedures and will not be further elaborated here.

SOME TECHNICAL ISSUES

It is important to note that any system of assessment cannot be perfect. It will always be necessary to accept a certain degree of tolerance, that is, an allowable margin of variation or inconsistency in assessor judgements. This is a principle that is well known in industrial practice and can be easily transferred to assessment practice. There is no simple measure of tolerance in assessment judgements of competency but it is evident that the tolerance level has been exceeded when there is widespread disenchantment and loss of confidence in those judgements. Smith (2000) has reported growing concern for the quality of VET assessments. In general, it would not seem that the tolerance level has yet been exceeded though it may be close in some cases. Even so, it is important to take action to protect the integrity of VET qualifications before public confidence is undermined.

Public confidence can be undermined, that is, the level of tolerance exceeded, if the margin of variation in assessor judgements comes to produce serious inequities or to endorse incompetence. First, quality control needs to ensure that there is consistency among assessors in the evidence considered appropriate and the standard of performance considered adequate to satisfy the competency standards. Second, quality control must ensure that people are not judged as competent if they would generally be judged as incompetent by other assessors. The first of these concerns prevention of ‘false negatives’, the second the prevention of ‘false positives’. The degree of seriousness attached to each of these types of ‘mis-judgements’ depends on the type of qualification and the area of employment.

There are five questions to address in verifying competency judgements:

• is there a sufficiency of evidence to make a judgement?
• has the competency standard been demonstrated?

---

6 ‘False negatives’ here are people judged ‘not yet competent’ when they should be judged ‘competent’. ‘False positives’ here are people judged ‘competent’ when they should be judged ‘not yet competent’.

7 The introduction of graded competencies would not change the primary importance of the competent versus not-yet-competent boundary. However, it would add to pressures for moderation at the other grade boundaries, especially if, as expected, employers would use such grades in selection decisions.
Verification of Assessments in VET

- will the competence demonstrated in this setting transfer to other settings?
- are the judgements of each assessor consistent with (that is, comparable to) their judgements at different times and in different contexts?
- are the judgements of one assessor consistent with (that is, comparable to) the judgements of other assessors (at different times and in different contexts)\(^8\)

Moderation/verification procedures need to address all five questions and to do so in a way that promotes confidence in the actual competency judgements.

Successful quality control (or quality management) that addresses these questions has other benefits besides the maintenance of confidence in the assessment judgements. It will also:
- develop and strengthen the assessment skills of assessors;
- reduce confusion and strengthen confidence among assessors;
- develop a shared ‘community of assessors’ (Eraut, 1994);
- provide mutual encouragement and support for assessors.

This clearly requires a participatory process, one which values and supports the judgements of assessors. It must also be flexible and interactive, not formulaic and prescriptive. It is important to reach consensus, not to impose it. That is, moderation/verification should be based on notions of partnership, not bureaucratic control. This is consistent with principles of successful institutional change (Fullan, 1996).

**ASSESSMENT AS PROFESSIONAL CONVERSATION**

Devereux (1997; 1999) has promoted the concept of assessment through professional conversation. This is in reaction to the tendency for assessment in the UK to focus on the formalistic aspects of documentation, especially through checklists and portfolios. Devereux proposed that:
- assessment should be seen as making a judgement on the evidence offered;
- the student should be responsible for offering the evidence and arguing the case for competence;
- the assessor and the student should make joint judgements about what has been achieved;
- there should be a focus on the why of doing not just on the what;
- assessment should be holistic rather than segmented into checklists;
- both the student and the assessor should adopt a critical and analytical stance.

Devereux (1997) proposes an assessment framework to enact these intentions. This framework involves three professional conversations (plus a pilot conversation) over a

---

\(^8\) These questions can be read as a mix of validity and reliability issues but the classical concepts of validity and reliability appear less and less relevant and helpful for assessment of this kind. They can be redefined, as Jessup (1991) attempted to do, or be abandoned in favour of more direct attention to the issues behind the concepts. This report adopts the latter approach.
period of 12 weeks, each one becoming progressively more focussed, with tape recording of the conversations for future reference. Although not mentioned, it is easy to see that students could present test information, if that was part of the assessment evidence, along with their on-the-job case material. There is an assumption that a single assessor will be responsible for the overall judgement of competence.

The justification for this approach to assessment is that it is more natural and direct (avoiding the need to construct portfolios that students are not used to constructing, a skill that is not related to the competencies themselves). Devereux (1999) claims that it has also ‘yielded rapid and effective results in terms of employee confidence, competence, productivity and morale’ and that it does so by creating ‘staff who think and act in a thoughtful and confident manner’ (p.ii). It achieves this by being collaborative but rigorous. Clearly, both assessor and student are participating in a conversation about competency standards and thereby creating in both cases a clearer understanding about what those standards mean and how they can be achieved. The progressive assessments also act formatively on the student’s developing competence. ‘Good formative assessment’, says Devereux, 1999, p.v), ‘is more about relationships with individuals than it is about ticking boxes’. This position is supported by the definitive summary of research on formative assessment by Black and William (1998) who conclude that self-assessment is essential to successful formative assessment.

Some important aspects of the professional conversation approach to assessment stressed by Devereux (1997, 1999) include:

- mutual respect for each other’s knowledge and experience;
- an intuitive appreciation of the other person’s needs and feelings;
- good listening skills and sensitivity to subtleties of language;
- a desire to learn from each other for an enriched understanding;
- allowing and encouraging the student to lead the conversation;
- focussing on the student’s performance and the standards, not the individual;
- focussing clearly and rigorously on (‘sticking like glue to’) the standards;
- believing (and communicating) that the student will become competent;
- empowering the student with the reflective skills to assess their own work.

Rephrasing Devereux (1999), a successful professional conversation should leave the student feeling that:

- their understanding (of competencies and standards) has been expanded;
- they can pinpoint what competencies they possess;
- they know what and how further development is needed;
- they have grown in confidence.

Devereux (1999) also, naturally, applies this form of assessment to the training of assessors. This would be one arena in which assessment as professional conversation could be implemented in Australia and that would begin to develop among assessors an understanding of this approach to assessment. It would also have beneficial effects on the assessors’ own developing competence and confidence as assessors, in the ways described by Devereux (1999). However, it seems unlikely in the short-term that such an
approach to assessment would be given widespread acceptance and use throughout the VET system in Australia. Such paradigm shifts take time.

Of more pertinence to the issue of moderation/verification is the possibility that this approach to assessment might offer a way of operating a moderation/verification process. In this case, the student is replaced by the assessor and the assessor is replaced by another assessor (the ‘moderator/verifier’). A system can be conceived in which assessors all belong to moderation panels and each in turn must justify a sample of their assessments to another member (or several members) of the panel. It is not essential that their assessments of students follow this approach but they must bring to the panel the evidence (and their judgements) of the sample of assessments. The advantages include:

- the direct involvement of each assessor in the verification of their own assessments;
- ownership and empowerment of all assessors in their own assessments;
- building of competence and confidence among assessors in their own assessment judgements;
- encouragement of a critical and analytical understanding of the competency standards;
- building of common understandings of assessment procedures and competency standards among panel members (the community of assessors);
- overall strengthening of the quality of competency assessments.

The notion of a moderation/verification process that depends on a professional conversation also sits well with the current devolved approach to assessment in Australian VET. The professional conversation is a shared but rigorous process, one which aims to strengthen self-monitoring. This could be a key component of any moderation/verification process.

**RISK ASSESSMENT: MOST EFFORT WHERE MOST NEEDED**

Risk assessment and risk management have become key factors in modern management. Risk involves unexpected or unfortunate consequences. Risk assessment is concerned with gauging the likelihood of such consequences. Risk management encompasses risk assessment but also includes mechanisms for reducing the assessed risks.

There are two levels at which risk assessment and risk management could be used in VET assessment. The first is at the level of the RTO and will be referred to here as internal risk; the second is at the level of the (state) system as a whole and will be referred to here as external risk.

Internal risk assessment and risk management are directed at identifying factors that may contribute inappropriate or undesirable assessment outcomes, assessing the potential for those factors to produce such outcomes (both false positives and false negatives) and implementing strategies to strengthen the characteristics of those factors to reduce the likelihood of those outcomes. The intention is to focus on intervention where it is most needed, that is, on those factors that are most troublesome in terms of their potential to
produce unsatisfactory assessment outcomes. Internal risk assessment and risk management are therefore tools that can be used to strengthen quality assurance approaches. They can, of course, be extended beyond a focus on assessment to include the whole of the delivery of training by an RTO. They can provide a way of looking in greater detail and greater depth at the RTO’s own capabilities for delivery and assessment than is provided by current registration requirements.

In the UK, the Qualifications and Curriculum Authority has produced a set of proposals for a risk assessment strategy of this kind and these are currently being piloted (QCA, 2000). It is notable that these UK proposals focus on assessment capabilities, procedures and perceptions (such as the number and kinds of student complaints), in other words, a combination of quality assurance and review. The moderation of assessor judgements is not covered. That is, there is no direct focus on actual assessor judgements, only on the managerial conditions that support assessor judgements. Further, since the process is conducted within the one organisation, there is no opportunity for comparison of assessor judgements and competency standards across organisations (the essence of moderation).

While risk assessment and risk management of this kind may be a useful component of quality assurance, and might be something that the VET system in Australia should consider implementing, it is not particularly relevant to quality control and moderation. Consequently, internal risk assessment and risk management will not be considered further in this report. Rather, attention will be devoted to external risk assessment.

External risk assessment as defined here is a whole state strategy. It is proposed that risk assessment in this case be seen as attending to the potential consequences of inappropriate assessment decisions in order to set priorities for implementing moderation/verification procedures. The assessment of potential consequences would need to be done separately for each qualification and might be the responsibility of each ITAB once a set of procedures is devised. In addition, a central representative committee could make an assessment of the relative risks for each industry. This could provide a basis on which to determine the relative importance of allocating resources to moderation/verification for each industry and qualification.

Table 2 provides a tentative list of external risk factors and risk indicators. These can provide the basis for undertaking an external risk assessment. Note that for safety and scope the risk indicators are associated with the consequences of an incompetent performance. Hence, also, these two risk factors are concerned only with false positives.

---

9 Key risks are identified as expertise of assessors and verifiers, management systems, assessment systems and organisational pressures. For each key risk there is a set of risk indicators and attached scales. The success of this approach depends critically on whether all the relevant risk factors have been identified, which is almost certainly unlikely. However, that does not mean it is worthless, only that is may be unable to prevent unexpected consequences, which is true of any risk strategy.

10 Throughout this report, issues of between state moderation have been ignored.
Table 2: Some possible external risk factors and risk indicators

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Risk indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety</td>
<td>Degree of employee independence in carrying out their work</td>
</tr>
<tr>
<td></td>
<td>Potential losses to employer from inadequate performance</td>
</tr>
<tr>
<td></td>
<td>Potential dangers to clients from inadequate performance</td>
</tr>
<tr>
<td></td>
<td>Potential dangers to public from inadequate performance</td>
</tr>
<tr>
<td></td>
<td>Potential dangers to self and co-workers from inadequate performance</td>
</tr>
<tr>
<td></td>
<td>Legal ramifications of incompetent performance</td>
</tr>
<tr>
<td>Scope</td>
<td>Amount of realisable expense from inadequate performance</td>
</tr>
<tr>
<td></td>
<td>Number of clients or public who may be affected</td>
</tr>
<tr>
<td></td>
<td>Sensitivity to complaint (by clients or public)</td>
</tr>
<tr>
<td>Equity</td>
<td>Extent of obvious variation in application of competency standards</td>
</tr>
<tr>
<td></td>
<td>Extent of student complaints about different standards being applied</td>
</tr>
<tr>
<td></td>
<td>Extent of employer complaint about variability in certified capabilities</td>
</tr>
<tr>
<td></td>
<td>Degree of seriousness attached to variability of certified capabilities</td>
</tr>
<tr>
<td></td>
<td>Extent to which qualification affects future prospects and earnings</td>
</tr>
<tr>
<td>Reputation</td>
<td>Effects on public image of vocational education and training</td>
</tr>
<tr>
<td></td>
<td>Effects on employer confidence in VET qualifications</td>
</tr>
<tr>
<td></td>
<td>Effects on employer confidence in particular RTOs</td>
</tr>
<tr>
<td></td>
<td>Effects on student confidence in the worth of particular qualifications</td>
</tr>
<tr>
<td>Improvability</td>
<td>Degree of assessor confidence in their own and others’ assessments</td>
</tr>
<tr>
<td></td>
<td>Capacity for consistency in assessment judgements to be improved</td>
</tr>
</tbody>
</table>

A FRAMEWORK OF POSSIBLE MODERATION/VERIFICATION PROCEDURES

The following framework of possible moderation/verification procedures includes both voluntary and mandatory procedures. Voluntary procedures would be easier to implement and involve less cost. Mandatory procedures would be harder to implement and involve more cost. Within each category, when considering each possibility by itself, there is a rough progression from less complex, less costly but less effective possibilities to more complex, more costly but more effective possibilities. A mix of possibilities would be more desirable than any one of them by itself.

Table 3: A framework of possible moderation/verification procedures

<table>
<thead>
<tr>
<th>Voluntary Procedures</th>
<th>Mandatory Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessor Partnerships</td>
<td>Internal Moderation/verification</td>
</tr>
<tr>
<td>Assessor Networks</td>
<td>External Moderation/verification</td>
</tr>
<tr>
<td>Assessor Workshops</td>
<td>Moderation/verification Meetings</td>
</tr>
<tr>
<td>Assessor Resources</td>
<td>Auditing</td>
</tr>
</tbody>
</table>
Voluntary Procedures

Voluntary procedures are less intrusive than mandatory procedures but are therefore likely to have limited effect. They may be more effective if combined with one or more mandatory procedures. The purpose of voluntary procedures is to provide various kinds of support to assessors, especially in the planning and implementation of assessment procedures and the making of judgements of competence. The assumption is that assessment practices can be and ought to be improved over time, that this is an on-going process, that assessors want to improve their practice, and that they will improve their practice through consultation with other assessors and through the sharing of assessment resources.

Assessor Partnerships

Assessor partnerships are small groups of assessors who:

- share and discuss each others’ assessment approaches and materials;
- provide mutual support for evaluating their assessment practices;
- assist each other in resolving any puzzles or problems;
- validate their assessment judgements in terms of the competency standards.

Partnerships could consist of only two assessors. The number of people in a partnership is limited only by the ability of the group to be self managing and fully inclusive but would probably not involve more than five members.

Partnerships could become a required part of quality assurance procedures, though this would change their character. If this is done, partnerships should be required to engage in a professional conversation not simply to approve each other’s assessments. However, as already noted, the conduct of professional conversations will need to be developed and supported through the provision of training and resources. Since it is desirable for the professional conversation approach to infuse the whole assessment system, this would not be an unreasonable expectation.

Activities for assessor partnerships include:

- sharing ideas;
- comparing notes;
- resolving problems;
- observing/commenting;
- critiquing assessment materials;
- reviewing competency judgements;
- professional conversation.

Advantages of assessor partnerships include:

- they can be locally organised and do not need bureaucratic support;
- there are few external costs apart from any promotional material;
- partnerships can be promoted as providing mutual benefit to partners;
• participation is personally empowering, reducing uncertainties and enhancing assessment capabilities.

Difficulties with assessor partnerships include:
• it is possible that nothing may happen if it is voluntary;
• participation depends on individual initiative and intrinsic motivation;
• level of involvement depends on personal commitment;
• successful partnerships require personal compatibility;
• there may need to be some professional support and resources;
• partners may simply reinforce each other’s errors and misconceptions.

Assessor partnerships offer some important benefits but probably not by themselves. They can support other mechanisms of verification but do not, and are not intended to, provide a mechanism for verification themselves. The introduction of a requirement that they be part of the quality assurance requirements of registration would force them to be treated more seriously but would undercut their value as an agency for boot-strapping the quality of assessment practices through enlightened self-interest and mutual support.

There is also a likely restriction to local partnerships. Encouragement might be given to seek partners in other RTOs, thus expanding the range of alternative ideas and practices encountered. However, competition among RTOs may restrict altruistic sharing. This might only be overcome by the situation of assessor partnerships within a broader verification system which makes assessment practices more transparent and accessible.

Assessor partnerships would also need to be re-formed on a regular basis to ensure the ‘cross-pollination’ of assessment understandings and practices around the state. Alternatively, it might be possible to encourage different partnerships for different courses so that assessors come into contact with different partners for different assessment programs. If it was possible to create new alliances over time or different alliances for different courses, it might be possible for standards to become aligned without the need for external intervention. However, this is rather optimistic without some mechanism for ensuring that assessor partnerships develop overlapping linkages around the state. Again, assessor partnerships can play a supporting role but in themselves could not constitute a system of quality control.

The central contribution that can be made by assessor partnerships is the opportunity created for discussion of competency standards and their implementation. Only through discussion and a focus on actual assessments is it possible to clarify the meaning of the competency standards and reach towards common understandings and judgements of competence.

Assessor Networks

The idea of a network is to connect people with similar interests so that issues and problems can be shared, discussed and possibly resolved. This can be achieved in a variety of ways. One way is to publish a list of contacts with profiles of their
specialisations and areas of expertise and encourage interested people to contact other people on the list. Another way is to organise regular meetings or conferences on particular themes so that personal contacts with other assessors are generated. Another way is through internet bulletin boards and discussion groups, allowing participation in questions and comments by members of the group according to need and interest (see Tower, Bloch & Harvey, 1999). The important thing in all networks is the opportunity created for people to maintain contact with other people and to use those contacts to solve problems and share ideas.

Activities for assessor networks include:
- sharing ideas and practices;
- seeking answers to questions;
- identifying similar problems;
- offering solutions to problems;
- offering comments on issues raised;
- directing to available resources;
- maintaining contact with colleagues;
- organising local visits and meetings;
- sharing materials and resources.

Advantages of assessor networks include:
- allow flexibility of membership, focus and coverage;
- can be tailored to meet any particular need or interest;
- self-sustaining provided they serve their purpose;
- networks can be promoted as providing mutual benefit to participants;
- participation is personally empowering, reducing uncertainties and enhancing assessment capabilities;
- personality plays very little role in connecting with others;
- contacts can be widespread and unrestricted by location;
- strong possibilities for engaging with new ideas and alternative approaches.

Difficulties with assessor partnerships include:
- networks have to be created, they do not simply happen;
- this requires central management and maintenance;
- there are costs involved in central management and maintenance;
- participation depends on personal interest and intrinsic motivation;
- whether any changes in practice occur depends on personal commitment;
- unless members venture an opinion, their own practices remain unchallenged;
- it is difficult to share cases and contexts, particularly on-the-job assessment.

Assessor networks do not constitute a quality control mechanism but they may make a contribution to assessment quality by fostering discussion of assessment practices and judgements among assessors. Such networks can support more deliberate forms of quality control.
There would be establishment and maintenance costs for networks. Some internet-based networks already exist and these might simply be extended. Costs would depend on how many networks were to be established, what kind of networks, whether they involve initial or periodic meetings or conferences, how many people are involved in each network and how much maintenance is necessary. Any move in this direction would need to be piloted first to assess the best way to establish them, the most effective way to organise them, the extent to which they require maintenance and the costs of establishment and maintenance.

It should be noted that as videostreaming becomes more widespread in future it will become feasible to share videotapes of real performances via the internet for direct comparison of the application of competency standards in work situations. Alternatively, visual portfolios can be created of work samples and assessment records.

**Assessor Workshops**

Assessor workshops are proactive. They provide opportunities for assessors to present and discuss ideas and practices. They can be organised around aspects of assessment that are known to be problematic (perhaps identified through other quality control mechanisms) or that are identified by assessors as of concern to them. Assessor workshops can be organised on a regular basis or as the need arises. They are an extension of training, recognising that situations will arise in practice that were not treated in training. An important aspect of such workshops would be sharing examples of assessment situations and assessment tasks and the judgements of competency associated with these (whether ungraded or graded).

Activities for assessor workshops include:

- sharing ideas and practices;
- providing answers to questions;
- offering solutions to problems;
- offering comments on issues raised;
- demonstrating assessment practices;
- illustrating assessment judgements;
- extending beyond basic training;
- directing to available resources;
- sharing materials and resources.

Advantages of assessor workshops include:

- they can be tailored to meet any particular need or interest;
- they can reduce personal uncertainties and enhance assessment capabilities;
- they offer a chance for direct discussion with other assessors;
- it is possible to run simulations to develop common understandings;
- they can serve as a basis for establishing assessor networks.

Difficulties with assessor workshops include:

- they require planning and organisation, probably at state level;
• they require central monitoring if not central organisation;
• there are substantial costs involved running workshops;
• there is a question of who should meet these costs;
• there are problems of location, timing and accessibility;
• participation depends on personal interest and intrinsic motivation;
• whether any changes in practice occur depends on personal commitment.

Assessor workshops do not in themselves function as quality control mechanisms though they can contribute to improvements in assessment practice. Periodic attendance at workshops could be made a requirement of keeping assessment qualifications current.

Workshops are only effective if they are relevant to assessor needs and of high quality. This requires careful and expert planning and design. Examples need to be drawn from real assessment situations. This means that workshops need to be combined with other forms of quality assurance and quality control.

Assessor Resources

Assessor resources are also proactive. They are a weak support for common interpretations of competency standards but can be a useful adjunct to other procedures. They can provide strong guidelines and prototypes for assessment without resorting to prescribed or standardised requirements as suggested by some people.

A recent study conducted by VETASSESS, National Strategic Evaluation of Consistency in Assessment: Final Report, has recommended the further development of assessment support materials. Some resources do already exist. For example, Infochannel Australia has produced a commercially available resource pack of three videos and handbook entitled Assessing Competence On and Off the Job. This claims to present detailed practical information, examples and guidelines for assessment. Necessarily, this is general advice and not specific to particular industries and qualifications. More resources are needed dealing with the specifics.

Such resources also need to be readily accessible and freely available to assessors as a part of the system support for assessment. The provision of such resources should be seen as a necessary cost for maintenance of high quality assessment in vocational education and training.

Mandatory procedures

Mandatory procedures offer the possibility of stronger control. However, they require some form of central monitoring or control together with possible sanctions to ensure compliance. In a decentralised system, sanctions necessarily relate to registration of the training organisation. There are also substantial costs in managing such procedures.
Internal moderation/verification

Internal moderation/verification can be required as part of the quality management procedures of each RTO. The UK model of internal verification identifies particular people as internal verifiers and requires them to sign off the adequacy of the assessments undertaken and the validity of the judgements of competency standards. This is not entirely successful because it can be diminished to a routine bureaucratic procedure and does not get to the heart of the issue, which is supporting the actual doing of assessment. An alternative is to require, not simply encourage, pairs or groups of assessors to act as internal verifiers for each other. Whether this would also deteriorate into formalistic procedures (ticking boxes and signing forms) is not clear but it seems likely without sufficient incentive and support. Also, since, the internal verification is ‘in-house’ (within the RTO), it does not address the consistency of assessment across RTOs. Linkages across RTOs will also be needed (see external verification).

Mandatory internal verification would require some process for checking compliance. This could be through the establishment of formal procedures of agreement between assessors, monitored within the RTO by someone designated as the internal moderator. The internal moderator could be required to keep records of the internal verification transactions, subject to audit on a regular basis or at the time of registration review.

The difficulty is that superficial compliance is not the aim. Rather, the aim is for assessors to engage in substantive discussion about their assessment procedures and assessment judgements. Whatever method of internal verification might be adopted, it would be necessary to provide training in how to do it. This would add to the requirements for training assessors, or might require review of the current competencies.

Preferably, internal verification would be through a professional conversation. This would require serious attention to the question of training for carrying out internal verification. It is difficult to see assessors implementing professional conversations without some modelling of how to do it.

Whatever the approach taken, the role of the person acting as internal verifier in any pairing of assessors is to confirm the procedures and judgements of the other assessor. The internal verifier does not undertake additional or parallel assessment and does not make independent judgements. Also, this is not an exercise in counting and reporting how often the assessors agree or disagree. Rather, it is aimed at being supportive and reaching consensus. Lack of consensus needs to be reported to a third party for resolution.

This raises the question of timing. If consensus is to be reached, and have some effect, it is necessary for the process to occur before final decisions are made about the checking-off of competencies and the issuing of an award. This can not be delayed until the award is about to be granted but must occur throughout the training program. So, internal verification should be continuous, involving periodic meetings between assessors on a regular basis. This allows attention to assessment approaches and processes as well as to
judgements in relation to the competency standards. It therefore has the capacity to enable corrective action during the training program if this is discovered to be necessary.

There is also a question of coverage: must the assessment for all students be verified? This would be unrealistic. Typically in moderation and verification procedures the overall program of assessment is considered – including set tests and tasks, other methods of evidence collection, assessment criteria, recording methods, and methods of synthesising evidence and deciding whether competency standards have been met – as well as the appropriateness of the judgements themselves. For the latter, a sample of judgements is sufficient. Sampling procedures may be specified, for example, 10 percent or √n of cases, or left to the discretion of the assessors. The purpose is to reach a point where both parties are confident of the appropriateness of the assessment procedures and judgements and are prepared to verify that this is so.

In some cases, the internal verifier may need to observe an assessment situation concurrently with the assessor. An alternative to direct observation is videotaping or audiotaping. An internal verifier might alternatively conduct a professional conversation with a student or listen to an audiotaping of a professional conversation between the assessor and the student.

Activities for internal verifiers include:
- reviewing assessment plans;
- sampling and reviewing assessment evidence;
- sampling and reviewing competency judgements;
- conducting professional conversations;
- referring disputed decisions to third party (moderator).

Advantages of internal verification include:
- it is located within the RTO (where official responsibility currently lies);
- there are few external costs apart from any organisational support unit;
- it invests assessor partnerships with more serious intent (than if merely voluntary);
- if properly organised can be personally empowering, reducing uncertainties and enhancing assessment capabilities;
- it can be linked to registration review to ensure that it occurs.

Difficulties with internal verification include:
- unless there is training and support, it may become formalistic;
- there are cost implications for RTOs (extra staff time and effort);
- level of involvement depends on personal commitment;
- successful verifier-assessor partnerships require personal compatibility;
- partners may simply reinforce each other’s errors and misconceptions;
- requires some procedure for checking compliance;
- it does not address the issue of consistency across RTOs.
External moderation/verification

External verification addresses the issue of consistency across RTOs. There are three basic approaches that can be taken. Two of these are external verifiers and external verification panels, along the lines discussed earlier under ‘assessment moderation’ (under the headings ‘external moderators’ and ‘external moderator panels’). It is not clear whether either of these approaches is workable within the decentralised Australian system. There are also substantial logistical and financial implications.

A third possibility, as proposed in New Zealand, offers a compromise between internal and external verification. This involves the pairings discussed in the previous section being drawn from two different RTOs (called Industry Training Organisations, ITOs, in NZ). In New Zealand this is referred to as interprovider moderation.

In New Zealand, interprovider moderation involves each provider establishing links with one other provider (link provider) to undertake external moderation (within the same domain and level of training). The link provider examines samples of student work sent by the other provider. Where the link provider cannot approve (verify) these materials, the materials are forwarded to a Moderation Coordinator for ‘check-moderation’. Moderation Coordinators are appointed and trained by NZQA Moderation Services. There is also, within each area of study, a National Moderation Coordinator who is responsible for training the Moderation Coordinators and check-moderating a sample of Moderation Coordinators decisions. Moderation Services is a business unit of NZQA responsible for managing the administration, training and contracts of national moderation coordinators and moderation coordinators and for keeping the moderation system under review. Providers are required to participate in the external moderation process or lose their accreditation status.

The New Zealand system of interprovider moderation would not seem to be entirely appropriate for Australia, being overly centralised, hierarchical and costly. Further, it would seem likely to be in danger of experiencing similar problems to the UK in the possibilities for formalistic acquiescence to the bureaucratic procedures. There is a lack of real support for improvements in assessor competence and confidence. The focus is on assessment materials and documents, that is, it is paper-based. It would be preferable for interagency pairing of assessors to involve more attention to reaching consensus and more attention to supporting improvement in the assessment procedures and judgements of competency standards. Even so, mandatory pairings of assessors across RTOs offers some advantages over voluntary pairings (but within the superstructure of moderation coordinators).

Moderation/verification meetings

A different approach which has not yet been tried would be to require attendance of assessors at moderation/verification meetings, attendance being a requirement of continued accreditation as an assessor. Meetings would need to be arranged on an
industry and regional basis. The purpose would be moderation of assessment procedures and judgements based on samples of materials brought to the meeting. Necessarily, decisions on the recognition of competencies and the award of qualifications could not be held in abeyance between meetings. Hence, the effects would be prospective not retrospective. That is, inappropriate assessment procedures and judgements could not be undone. The need for change in procedures and judgements could only be indicative for the future. However, assessors who are consistently and seriously at variance with other assessors might be targeted for additional assistance and training, or de-accreditation.

A few points can be made about the organisation of moderation/verification meetings:

- meetings need to be industry and program specific;
- they need to be organised on a regional basis;
- they need to be centrally arranged and managed;
- meetings should be held yearly or half-yearly;
- each assessor brings a sample of case portfolios;
- focus on difficult, unusual and borderline cases;
- attention can be given to incomplete assessments;
- each case is reviewed by two other assessors;
- combinations of assessors are rotated throughout the meeting;
- attention is given to assessment evidence and competency judgements;
- differences are discussed with the aim of reaching consensus;
- lack of consensus is arbitrated by involving more assessors;
- a coordinator must be designated and trained to run the meeting.

Advantages of moderation/verification meetings include:

- involvement of all assessors in an external moderation process;
- democratic involvement – all assessors act as verifiers;
- cross referencing of standards across assessors and RTOs;
- exposure of assessment details to only two other assessors;
- opportunity for discussion about differences of opinion;
- possibility of discussing cases where assessment is still incomplete;
- involvement is agency for professional development;
- generation of informal networks among assessors (see Foyster, 1995);
- direct effects on assessors through discussion and example;
- avoidance of overemphasis on paper based communication;
- formal validation before award is traded for stronger cross-referencing, assessor involvement, professional development and networking.

Disadvantages of moderation/verification meetings include:

- there are substantial costs involved (organisation of venues, communication with assessors, registration, accommodation, coordinator training, etc.);
- a central moderation services unit is needed to organise the process;
- there are logistic problems to be overcome concerning times and venues;
- records will need to be kept concerning attendance and outcomes;
• it is infeasible for all assessments to await validation at a moderation meeting;
• the aim is long-term improvement rather than short-term validation.

Auditing

Auditing is an ‘after the event’ form of quality control, where the purpose is improvement in future assessments. It can also be linked to requirements for continued registration.

Audits can be conducted to randomly sample or to systematically cover the whole system within a nominated time frame. Another possibility is to focus on selected assessors, units or RTOs based on a risk assessment. In the latter case, audits are best used to supplement other forms of quality control.

For purposes of quality control, audits need to focus on actual assessment processes and judgements. Therefore, they involve site visitation or materials review or both. Site visitation can make use of observation, materials analysis and discussion, focusing on assessment procedures and judgements. Preferably, site visitation could involve a professional conversation with assessors. In other words, the auditor would function very similarly to an external verifier. This carries implications for staffing and training of auditors similar to those of external verifiers or moderation coordinators.

Advantages of auditing include:
• encourages RTOs and assessors to take voluntary moderation seriously;
• lower cost than complete verification of all assessments;
• less intrusive and demanding than other forms of mandatory verification;
• can be linked to registration requirements for RTOs.

Disadvantages of auditing include:
• relatively costly (needs central management, appointment and training of auditors, travel and accommodation for auditors, communications, etc.);
• not a stand-alone procedure (with additional costs for other procedures);
• has most of the disadvantages of external moderators.

SOME POSSIBLE SANCTIONS (GRADUATED)

These possible sanctions are offered without further comment since they already appear in the earlier discussion:

• professional conversation (not really a sanction; more a support between equals);
• negotiated agreement on processes/judgements (verifier holds trump card);
• assessor retraining required (to maintain certification or update competency);
• withdrawal of assessor accreditation (only if persistent and irremediable deficiencies);
• withdrawal of RTO accreditation (only if no support for improvement).
SOME SUPPORTING OPTIONS

These possibilities are offered here only in summary form:

- development of a Code of Practice or Modus Operandi (see *The UK-NVQ Code of Practice* – 19 pages plus 5 appendices);
- development of an Industry Component for assessor training, dealing with the specifics of assessment within that industry;
- requirement of regular updating to maintain assessor certification;
- more focus in all assessment materials and guidance on assessment criteria and judgements.

A SUGGESTED WAY FORWARD

This report has charted options for moderation in the VET system in Queensland. Decisions about which of these options to implement depend on policy considerations relating to structures, authority and financing. What is suggested here is a way in which further consideration of the issues raised in this report might be advanced and a way in which an initial pilot of some of the options might be undertaken. It is necessary to think in terms of an initial pilot rather than full-scaled implementation since the exact form of any moderation system will need to be refined in the crucible of industry practice and acceptance. The final shape of a moderation/verification system is likely involve a mix of various procedures.

1. A Moderation Services Unit should be established within DET to oversee development of a moderation/verification system for VET in Queensland.

In the first instance this should be a fairly small unit since its initial remit would be piloting of moderation/verification procedures. However, if the intent to implement moderation/verification procedures is serious, this will need to develop into a more substantial unit. Decisions on how to proceed beyond the pilot stage will depend on the success or otherwise of the pilot program.

For the pilot program, the Moderation Services Unit will need at least one senior officer and one project officer plus administrative/secretarial support. Funding must also cover office, communications, travel and moderation meeting costs. A modest one-year pilot program might be manageable for $150 000–$210 000 depending on the time commitment of the senior officer.

2. The first task of the Moderation Services Unit would be to conduct an external risk assessment along the lines indicated in this report.

The purpose of the external risk assessment is to determine where the need is greatest for development of moderation/verification procedures. For purposes of the pilot program, this could be fairly low-key and in-house. It is a question of starting points rather than final structures. Willingness of trainer/assessors to participate may be a critical factor.
3. **Emphasis should be placed on piloting of moderation/verification meetings**

Moderation/verification meetings offer the greatest advantages for developing a community of assessors who share common understandings of assessment procedures and competency standards. They also offer the most powerful professional development of assessors and the greatest potential for dissemination of good assessment practice. It is difficult to see the other mandatory procedures discussed in this report being appropriate within the Australian context (apart from auditing which is supplementary rather than stand-alone). Voluntary procedures should be seen as fall-back options if moderation/verification meetings cannot be implemented successfully.

For purposes of a pilot program, participation in moderation/meetings might be voluntary. For full implementation, it will be necessary later to devise some way of extending such meetings to cover all assessors, at least in those industries where moderation of this type is implemented. As suggested earlier in this report, the final system of moderation might involve different procedures for different industries and programs, depending on the assessed risk. In other words, moderation/verification meetings might not be implemented in all cases. The mix of moderation approaches does not need to be a standardised one. Furthermore, it is preferable for voluntary procedures to be adopted rather than to have no procedures at all.

Since assessment in VET does not neatly align with particular blocks of time as in secondary schools, a different approach is needed to the relationship between moderation/verification meetings and certification. Issuing of student awards cannot wait for such meetings. In any case, the issuing of student awards is a responsibility vested in the RTO. Where the meeting endorses the assessor’s assessment procedures and judgements, the implications are retrospective – to confirm the decisions that already have been made. Where the meeting does not endorse the assessor’s assessment procedures and judgements, the implications are prospective – to suggest ways in which their assessment procedures and judgements need to be adjusted in the future.

Sufficient numbers of accessible meetings would need to be organised to enable every assessor to attend at least one meeting each year. Yearly attendance could be a requirement of continued accreditation as an assessor. Failure to move towards full endorsement over a reasonable period of time, say three years, might be reason for loss of registration or participation in voluntary procedures for strengthening their assessment approaches and interpretations of competency standards.

4. **All moderation approaches should make use of professional conversation as the method for arriving at consensus on the quality of the assessment procedures and the validity of the judgements of competency standards.**

Professional conversations recognise the equal participation of both parties (moderator and assessor) in the moderation process. They also emphasise the development of the capacity of the assessor to think clearly and deeply about their assessment procedures and standards and to internalise common understandings shared among all assessors.
References


