

### **Internal summative assessment in National Vocational Qualifications (NVQs): the concerns of employers**

Kathryn Ecclestone, University of Exeter

#### **INTRODUCTION**

This paper outlines the role of NVQs in the UK's education and training system, describes its main features in terms of reliability and validity (also presented in the working paper for this seminar) and raises some questions about how employers might regard the assessment regime and, in particular, the role of internal assessment by college teachers. Questions in the paper about employers' perceptions do not arise from a robust body of empirical evidence or from a review of independent evaluations of NVQs over their 15 year history. This evidence needs to be collected. Instead, questions here arise from academic analyses of how NVQs were designed and implemented and a recent, in-depth qualitative study for the Learning and Skills Development Agency of post-16 assessment regimes, including NVQs in social care, sports and recreation and motor vehicle maintenance, with which I have been involved (Torrance et al, forthcoming).

It is therefore likely that the paper has missed data and insights about NVQs from bodies like City & Guilds and from independent academic evaluations. However, it is intended as a starting point for discussion about employers' perceptions of 'teacher' (internal) summative assessment in vocational qualifications and the extent to which concerns about internal assessment in NVQs mirror those in other assessment regimes. Discussion of the paper at the Assessment Systems of the Future seminar could provide a basis for a systematic research-based evaluation of data from awarding bodies about take-up of NVQs and independent studies of NVQs over the past 15 years. Such a review should explore how different constituencies in the system, including large and small employers, regard NVQs and their assessment regime.

The paper outlines key features and then raises questions about possible concerns amongst employers in relation to:

- aims of the NVQ assessment regime
- features of assessment
- properties of validity and reliability
- practicability

#### **AIMS OF NVQs**

Introduced in 1989, NVQs were a radical and controversial over-haul of work-based qualifications on offer at the time. They were supposed to update and improve the quality of work-place training by reflecting the demands of employers for key competences in specific occupational roles. However, a controversial and radical goal

was to take the design and implementation of assessment for work-based competence away from 'over-theoretical' programmes and inappropriate methods, assessed in education institutions. Instead, designers wanted assessment to be carried out by internal work-place assessors rather than 'teachers': the term 'teacher assessment' does not therefore resonate with the NVQ system, unless college teachers are also assessing candidates as part of a work-based programme (see Jessup, 1991).

Specifications of competence and criteria for assessing and accrediting them were created initially by industry lead bodies representing employers' interests in different occupational sectors: these were reorganised as sector skills councils in the 2001 Learning and Skills Act. NVQs can be taken whilst working in a job or apprenticeship scheme, on a specific course in a college, a qualification and training programme offered by a training provider or a mixture of these formats.

A coherent system of NVQs was supposed to streamline and rationalise the design and methods of assessment for all work-based qualifications as part of a coherent framework of progression to competence at different levels in a national qualifications framework. NVQs were designed to be available from entry level to employment to degree and professional equivalent. However, the biggest growth areas for NVQs have been at level 2 in the qualifications framework for public sector workers in health and social care, retail and leisure industries and in the use of NVQs to accredit modern apprenticeships taken by school leavers.

In addition to problems of coverage, the goal of rationalising qualifications has been undermined by a huge proliferation of non-NVQ qualifications through customised qualifications designed for individual employers who find NVQs unsuitable for their training needs. These are also accredited by awarding bodies. Some of these customised qualifications have similar assessment regimes to NVQs while others do not.

The extent of employer demand has also been undermined by the way that the vast majority of funding for NVQs has come from taxpayers: the goal of employers paying for NVQs after their initial introduction has not been realised. As Field points out, without compulsion for public sector organisations such as Royal Mail and the NHS to offer NVQs, and significant amounts of public subsidy, NVQs would not have got off the ground (Field, 1994). This begs the question of how far NVQs continue to be used because they are paid for largely by the Learning and Skills Council.

One effect has been the creation over the past 15 years of the most complicated and confusing system of work-based qualifications in the world...probably....(to parody a famous advertising slogan!) The UK has thousands of vocational qualifications, hundreds of awarding bodies and thousands of providers including FE colleges, employers and private training organisations. In addition to long-established, well-known and respected vocational bodies such as City & Guilds, NVQs can be accredited by unions, professional organisations, sector-specific bodies and small awarding bodies. The largest awarding body for NVQs is City & Guilds which works in diverse occupational sectors and with thousands of providers and candidates. NVQs are also 'exported' to other countries through the British Council.

## ASSESSMENT FEATURES

### Specifications and their design

The *competence-based assessment* regime of NVQs is a particularly strong form of criterion-referencing which emphasises the authenticity and validity of work-place competences. Detailed specifications of competence in different roles, the range of contexts in which it must be demonstrated and the indicators of performance that show it has been achieved are used by assessors and candidates to assess and record achievement, or to set further targets. The goal of its designers was that accurate specification would enable assessors to assess validly and thereby reliably.

Sector skills councils are responsible for creating and updating the specifications to reflect employers' needs in different organisations within a particular sector, such as retail. An NVQ comprises units of competence that can be taken and accredited separately and as part of a whole qualification.

### Summative assessment

Summative assessment is based on the internal assessment of practical competence by workplace supervisors in assessment roles, college teachers and other assessors, supplemented by externally designed unit tests of underpinning knowledge; some of these are accessed on-line as and when candidates are ready to take them. There are requirements for training in competence-based assessment for anyone assessing NVQs.

Assessment demands centre on what the learner (often also the employee) can do, and can be seen to do, in relation to the tasks required of them for competent practice. Detailed specifications of outcomes and assessment criteria promote and demand 'mastery' (i.e. coverage of all demands) as opposed to compensation and grading (ie where assessors can off-set poorer performance in some areas by better performance in others).

Candidates are required to show evidence of workplace competence in diverse forms, relevant to demonstrating mastery. These include: observation by supervisors and/or external assessors; written testimony by colleagues or managers; written assignments; practical tasks; oral feedback and testimony. There is a strong emphasis on assessment tasks being 'fit for purpose' and the validity of assessment as opposed to reliability. Candidates can repeat assessment tasks until they are deemed to be competent, producing assessment decisions of 'not yet competent' (working towards..) or 'competent'.

NVQs require learners to demonstrate achievement when they are ready ('readiness'), preparation and detailed help in the form of formative guidance and feedback and, in some cases, repeated assessments until the candidate achieves the outcomes. NVQs are also supposed to be rooted in authentic, work-place contexts and assessed by people inside those contexts. Thus 'achievement' is defined in terms of demonstrated competence, while 'fairness' involves transparency of criteria and procedure, comparability/similarity of assessment tasks and contexts, and multiple opportunities to demonstrate the required competence(s).

In addition to externally-set tests of ‘underpinning knowledge’, some NVQ candidates, such as those on college programmes or modern apprenticeships have to take ‘key skills’ tests of literacy, numeracy and ICT and these appear, from the LSDA project cited above, to be widely disliked. Apprentices with relatively poor prior academic qualifications dislike key skills tests because they remind them of school failure; but relatively well qualified apprentices (e.g. with 4, 5 or more GCSEs at A\*-C) regard them as an ‘insult to their intelligence’. Non-completion of Key Skills tests seems to be a major factor in non-completion of Modern Apprenticeships (even though the young workers in question attain their NVQs).

In NVQs, candidates can repeat tasks until they demonstrate competence, with as much guidance as necessary. In general A-levels, teachers are accustomed to the one-off pressure of a summative examination as part of an attempt to create reliable differentiation in grades. They therefore see formative guidance as an ‘unfair’ advantage because they believe that assessment should reflect the students’ ‘uncontaminated’ ability and performance on a particular occasion.

### **Formative assessment**

Formative and summative assessment are synonymous in NVQs, where diagnosis, target-setting, review, recording competence becomes the learning (assessment) programme. Assessors gear their coaching and teaching to meet the next competence or target, according to how well trainees performed summatively in the previous one. A number of assessors might be involved in one trainee’s programme: a supervisor at work, a colleague, a teacher/trainer in the college or training provider, the employer.

Some tasks cover a number of competences. Evidence gathered during learning and work activities are interpreted in terms of progress towards units of competence and the smaller tasks that comprise those units. The involvement of trainees in this process is extremely variable from one setting to another – in some, it seems that assessors compile the portfolio for the trainees so that they are almost unaware of how their tasks at work lead to the NVQ! In others, the self-assessment, review and close engagement by trainees with designing their programmes of work and assessment, envisaged by the architects of NVQs, are more visible.

Assessment is also used to accredit prior learning (APL), where review of past performance and activities, and evidence from these, can be used to award units of NVQs without new learning being needed.

## **PROPERTIES**

### **Construct validity**

This is central to the design and specification of the units, competences and performance criteria and the NVQ was designed with this property as its central goal in order to ensure that assessment was fit for purpose, did not require unnecessary theory, was ‘employer-led’ and based on authentic, real-life tasks for specific occupations. However, validity rooted in the authenticity of work-based tasks is compromised by simulation of activities, for example in colleges, and by lack of access to the full range of occupational tasks needed to make up a unit of competence.

There are questions in NVQs about how far the diverse range of methods used as 'evidence of competence' reflects the learning being assessed. Specifically-designed tasks and tests are designed for learning and assessment as synonymous activities defined in the criteria and specifications. These form the 'syllabus' and set the goals of learning: how far these are wide-ranging, covering skills, attitudes and creative and critical thinking skills is questionable and varies among NVQs.

### **Consequential validity**

NVQ outcomes are used to judge the institution's or organisations overall achievement rates, and where NVQs are funded by the LSC, these outcomes are monitored as part of national targets for the achievement of NVQs at different levels. Teaching and training focus on the summative outcomes but this as much a product of the strong criterion-referenced format and the emphasis on teacher (assessor) assessment as the way that outcomes are used for accountability.

A tendency to focus on a narrow interpretation of criteria and on performance rather than learning is not solely because of excessive accountability for summative results. It is also because of resource pressures to get trainees through the process as cost effectively as possible and the closely specified and prescriptive competences and criteria. It is not clear in NVQs how some of the positive consequences of greater freedom compared to external tests might apply: the pressure to pass the competences comes from other sources, not least the fusion of formative and summative processes and activities.

### **Reliability**

It is important to note that 'standards' in NVQs mean 'occupational standards of competence' rather than its more familiar meaning of 'comparable standards' of achievement within qualifications or between centres and cohorts. NVQs trade national reliability and standardisation of assessors' judgements between centres and providers in order to privilege validity and authenticity. Awarding bodies therefore place more emphasis on verifying that procedures and guidance have been adhered because of the sheer variety of practices, providers and programmes. However, awarding bodies do moderate and sample assessment decisions and so although reliability is a feature of NVQs, it perhaps takes a different form than in other qualifications.

The chief source of low reliability in NVQs is the local, individual nature of interpretation of evidence by assessors. Institutions are required to carry out internal verification and moderation of procedures and outcomes, with some sampling of portfolios. Awarding bodies follow this up with annual visits to providers by a subject specialist and issue annual reports about the national standards of work in each area. QCA does not collect national data to compare centres and providers in terms of NVQ outcomes. Comparability of outcome and provider is therefore arrived at through tight specifications of tasks and procedures rather than by moderating assessment decisions.

## **Practicability**

The running costs of NVQs are high because of the intensity of the assessment process and its individual focus. There is wide variation in quality of training underpinning the assessment process, in terms of time spent on training, individual reviews and portfolio building. Assessors also report spending a great deal of time translating the criteria, tracking the evidence in the portfolios and generally auditing and managing assessment and quality assurance processes for the awarding bodies in a complex and prescriptive assessment regime.

In addition, where providers offer different NVQs from different awarding bodies, the administrative and quality assurance costs can be very high: a large college, for example, might pay fees to 20 different awarding bodies in order to offer NVQs and other vocational qualifications. This also requires creating links between different quality assurance procedures for verification and moderation.

## **CONCERNS ABOUT ASSESSMENT**

The concerns discussed here arise from assessors, candidates and awarding body officials interviewed for the LSDA project and further research is needed to verify how far they also reflect employers' concerns. However, some of the concerns discussed below were also raised by Alison Wolf from her work with the design of NVQs in the late 1980s and her analysis of competence-based assessment (Wolf, 1995).

### **Understanding the assessment demands**

One element in characteristics of validity and reliability is the enduring opaqueness of the outcomes-based, competence-based language of NVQs. While improvements have been made to many of the original specification documents, successive generations of learners and assessors have to learn the language anew, despite the efforts of awarding bodies to simplify the language and format of the standards. Assessors in the LSDA project reported the need to repeat information, to translate the 'gibberish' and 'wordiness' of the specifications and to wait for understanding to emerge after candidates have done one or two units.

### **Getting candidates through the requirements**

The LSDA project shows very high levels of support and guidance in the process of identifying evidence. There are many examples in the data of tutors and assessors interpreting awarding body specifications and criteria and providing simple translations of what they "really mean". Sometimes, these examples tread a very fine line between eliciting what the candidate 'knows' and leading them, sometimes word for word, to articulate the desired answer. For example, there were numerous examples of workplace settings where such support was observed through 'leading questions' by assessors of a 'good lad' to help him through observations of his workshop practice and compile his portfolio evidence. Similarly in Social Care, assessors were observed asking leading questions to help candidates articulate what they (supposedly) already know and can do.

The practices of translation, support and direction revealed by the LSDA project have implications for the desirable interpretation and mediation of assessment procedures and practices at local level. There are also questions about the nature of the knowledge candidates have and how they acquired it. The project raises questions about the extent to which range statements and performance indicators can be said to 'represent' the reality of workplace competences if they are not recognized as such in the workplace, have to be 'translated' for workplace use, and if observed competences have to be translated back again into acceptable evidence statements.

None of this is necessarily inappropriate or unfair in itself since it might be argued that such practices are at the heart of professional judgements about the interaction between performance and competence that all assessors must make in different regimes. Nevertheless, there are problems about equity if they are not pursued uniformly: questionnaire data in the LSAD project suggest that there can be wide variations in the frequency and length of assessor visits. For example, while the most frequently reported timing of assessor visits amongst the NVQ-takers was 1-2 hours every 4-6 weeks, one reported that they saw their assessor once a week for 2-3 hours, while four reported that they saw their assessors only every 3 months or less *and* for one hour or less (Appendix 3). Some regulation of and minimum recommendations for such visits and levels of support seems to be appropriate.

### **'Fair' assessment**

It is important to note that the emphasis on validity and authenticity, mastery and competence in NVQs led assessors in the LSDA project to have very different notions of what counts as 'fair' assessment in NVQs compared to other post-16 assessment regimes. This affects tutors' and learners' attitudes to what counts as 'achievement' and to the respective roles of formative and summative assessment. In contrast to NVQs, for example, formative assessment in Advanced Vocational Certificates of Education (AVCE) helps students improve their grades and this is integral to the educational ethos of the qualification. This takes the form of guidance on draft assignments and close attention to the criteria which students are encouraged to use in detail.

### **Opportunities for valid assessment**

The LSDA study showed that what one might call 'opportunities to verify' vary greatly across work-based and college-based settings. In turn, this has implications for reliability of the standards of competence achieved. For example, small garages may not provide NVQ level 3 opportunities to conduct diagnostic work with the latest computer technology. Equally however, and somewhat ironically, well-resourced main dealers for leading carmakers do not always provide NVQ level 2 opportunities for basic repair – "clutches don't go wrong on Volvos". In Sport & Recreation small hotel leisure facilities can be very limited in the equipment available, and indeed in client activity, so simulation is often called for with another member of hotel staff acting as a client. Often therefore, 'ways and means' are found to observe and verify competences but this can impede progress and have negative effects on learners and assessors' motivation.

A related issue with respect to ‘opportunities to verify’ is that candidates may not necessarily be in a position to gather relevant evidence. For example candidates already have to be in a supervisory position to demonstrate many level 3 competences, but they wouldn’t be in such a position and cannot secure such a position, if they are not yet considered competent. Similar issues pertain to client safety in Social Care, even at lower levels of the awards. Care workers will not (should not?) be in a position to exhibit evidence of safe practice until they are already competently safe.

Policy-related questions arise about whether the goal of defining current national standards is still appropriate to workplace activities (and if not how they can be updated quickly) and, in turn, whether simulation is acceptable and if so, to what extent. For example, changing a clutch in a college workshop because no such job has occurred in the workplace would seem to be acceptable (if still deemed necessary); ‘pretending’ to do a client fitness appraisal on a colleague that one works with every day is perhaps less appropriate.

## QUESTIONS

- To what extent do employers see work-place assessment as more valuable than assessment by teachers in colleges?
- To what extent does low take-up of NVQs paid for by employers reflect concerns about the reliability of internal assessment as a basis for selecting between candidates for jobs?
- Does low take-up of NVQs and the proliferation of customised qualifications reflect employers’ concerns about the validity of assessment and a demand for local, specific forms of validity in particular workplaces or organisations?
- Do employers value reliability in NVQs? If so, does internal assessment undermine their confidence in it?
- What uses do employers in different sectors make of NVQs for recruitment and promotion?
- How do employers regard NVQs for recruitment and promotion in comparison to other qualifications or mechanisms such as psychometric tests that can be used for these purposes?
- To what extent do employers worry about excessive guidance of candidates as a potential erosion of both validity and reliability?
- Do employers value consequential validity in NVQs?
- What evidence is there of consequential validity in NVQs?
- How do employers deal with the practicability of NVQ assessment?

## REFERENCES

Ecclestone, K. and Hall, I. (2000) *Quality assurance and quality control in post-14 qualifications: a report for the Qualifications and Curriculum Authority* (Newcastle, University of Newcastle)

Field, J. (1994) Reality testing in the workplace: are NVQs employer-led?, in Hodkinson, P. and Issitt, M. (eds) (1994) *The challenge of competence: professionalism through education and training* (London, Cassell)

Jessup, G. (1991) *Outcomes: NVQs and the emerging model of education and training* (Sussex: Falmer Press)

Torrance, H., et al (forthcoming) *The impact of post-16 assessment systems on achievement* (London, LSDA)

Wolf, A. (1995) *Competence-based assessment* (Buckingham: Open University Press)