

Raven, J., & Stephenson, J. (Eds.). (2001). *Competence in the Learning Society*. New York: Peter Lang, pp.453-466.

CHAPTER 25

Competence-Based Assessment

Alison Wolf

Version date: 18 September 2000

Editorial Introduction

In this chapter, Wolf illustrates the self-defeating nature of attempts to generate atomistic specifications of vocational competence and then to use these as a basis for decontextualised “objective” assessments. Her chapter stands as a marker for her even more outstanding book with the same title. In that book, Alison documents the way in which the attempt to move away from traditional knowledge-based “educational” programmes and replace them by courses in which students learned how to *do* things they would later need to be able to do--and testify to the outcomes in these competency-based terms--got corrupted back into the very thing that it had been hoped to move away from (i.e., assessments of prescribed cognitive content) precisely because there were no appropriate means of assessing these other outcomes.

The idea of competence has, in the UK, become almost inextricably linked with a particular assessment philosophy promoted by the National Council for Vocational Qualifications (NCVQ, now merged into the Qualifications and Curriculum Authority), and, to a rather lesser extent, by the Scottish Vocational Education Council (SCOTVEC, also now part of a unitary body, the Scottish Qualifications Authority--SQA). These organisations have regarded assessment as an extremely powerful weapon, capable on its own of guaranteeing quality, promoting a truly “competence-based” approach to training and learning, and increasing the skills levels of the population. While they are certainly right about the powerful effect of assessment on practice, their own approach has, unfortunately, had results quite at odds with the ideals of those first responsible for promoting a competence-based approach. Competence-based assessment as it has been understood in the UK is not, in fact, the only possible way of assessing people’s competence or capability; but it is important to understand why it has had not merely disappointing but actively pernicious effects, and how, therefore, they might be avoided.

The ideas behind the type of competence-based assessment practised and preached in the UK are essentially American in origin. The literature on competence-based assessment, which appeared in Britain in the 1980s, is packed with direct echoes of U.S. literature of 10 years before. What is dramatically different between the two countries and periods is that government policy in the UK ensured the general adoption of competence-based approaches by tying them to central government funding. In the United States, localised experimental work had little or no long-term effects, so that American advocates of the approach are now

rediscovering it, in part through the British programme.

The following definition is an American one. Yet it summarises all the major features of competence-based assessment as currently advocated in the UK:

Competence-based assessment is a form of assessment that is derived from a specification of a set of outcomes; that so clearly states both the outcomes--general and specific--that assessors, students and interested third parties can all make reasonably objective judgements with respect to student achievement or nonachievement of these outcomes; and that certifies student progress on the basis of demonstrated achievement of these outcomes. Assessments are not tied to time served in formal educational settings.¹

The three components of competence-based assessment which are especially important, and that the definition above encapsulates are:

1. the emphasis on outcomes; specifically, multiple outcomes, each distinctive and separately considered.
2. the belief that these outcomes can and should be specified to the point where they are clear and “transparent”. Assessors, assessees, and “third parties” should be able to understand what is being assessed and what should be achieved.
3. the decoupling of assessment from particular institutions or learning programmes.

These characteristics define the practice of competence-based assessment. However, the emphasis on outcomes and “transparency” is not peculiar to the competence-based context. It is also a defining characteristic of a rather broader theory of measurement, that of “criterion referencing.” Criterion referencing is similarly concerned with clearly specified outcomes and with assessments that address these outcomes separately rather than dealing with “pass marks” or “norms.” It too has been a very influential approach in recent years (for example within the English National Curriculum), and it too hails conceptually from the United States. Nonetheless, competence-based and criterion-referenced assessment are not synonymous. The former involves an idea of competence that is essentially non-academic. In practice, as noted in the same American text from as that from which our first definition was derived,

It tends . . . to derive from an analysis of a prospective or actual role in modern society and . . . attempts to certify student progress on the basis of demonstrated performance in some or all aspects of that role.

In other words, it is vocational in the broadest sense and is bound up with the idea of “real-life” performance. Indeed, in its early days in the United States, “performance-based” assessment (and education) were the terms used more often than “competence.”

Competence-based assessment became important in England following the 1986 governmental Review of Vocational Qualifications. This led directly to the creation of the National Council for Vocational Qualifications, with a remit to establish a National Vocational Qualifications (NVQ) system of approved vocational awards. The review argued that “assessments carried out by many bodies do not adequately test or record the competences required in employment,” that “assessment methods tend to be biased towards the testing either of knowledge or of skill rather than of competence,” and that there are “many barriers to access arising from attendance and entry requirements.”

These were well-founded concerns, and it is interesting to find in the early documents how broad a conception of “competence” prevailed. Unfortunately, however, a number of influential figures also believed that it was possible to find a simple, all-embracing, way of identifying and translating this conception into practice. This had enormous appeal to politicians and the added advantage of excluding entirely any need for consultation with or input from the despised education sector.

The strategy was to create “lead industry bodies”² that represent a given sector of

industry or employment and give them full responsibility for drawing up detailed standards of occupational competence. These standards, in turn, were to be used, unamended and unchanged, as not merely the *basis* for, but the sole effective *definition* of, vocational awards. No qualification will be recognised as an “NVQ” unless based on the standards issued by the lead industry body concerned, which meant, in effect, that parts of the standards become a qualification, with a certain amount of “topping and tailing” to explain recording and verification procedures.

As national qualifications, NVQs each cover a particular area of work at a specific level of achievement. They are based on the fundamental assumption that, for each industry, there exists a single identifiable model of what “competent” performance entails. The idea that, for each role, there exists such an agreed-upon notion of competence, which can be elicited and command consensus, is fundamental to any assessment system of this type. It is also an heroic--and a questionable--assumption.

The structure of an NVQ is modular or “unit-based.” These units are defined as groups of “elements of competence and associated performance criteria which form a discrete activity or sub-area of competence which has meaning and independent value in the area of employment to which the NVQ relates.”³ An element of competence is a description of something that a person who works in a given occupational area should be able to do. It reflects an action, behaviour, or outcome that has “real meaning” in the occupational sector to which it relates. For example:

create, maintain and enhance effective working relationships is an example of a management competence, while

inform customers about products and services on request is taken from a list of financial services competences.

Both share two compulsory qualities. They involve an active verb and an object--that is, they are performance-based--and they are not tied in any way to particular training programmes.

As expressed, both these examples are obviously very general statements indeed. Each could apply to a huge number of contexts--and to performance of very variable quality. However, the key aspect of NCVQ-approved standards is that they go into far greater detail than this. Lead bodies are expected to define very precisely the nature of what is expected, with the first level of detail corresponding to highly specified *performance criteria*. These are the statements by which an assessor judges whether an individual can perform the workplace activity at the standard required. In effect, the performance criteria state explicit measures of outcomes. Figure 25.1 provides an example of an element of competence with its performance criteria.

To be accredited with a competence, a candidate must demonstrate successfully that he or she has met *every one* of these criteria. This is because competence-based assessment, as interpreted by the National Council for Vocational Qualifications, requires *one-to-one correspondence with outcome-based standards*. This must be comprehensive: Evidence must be collected of a candidate’s having met every single performance criterion. Failure to do this, it is argued, removes an essential characteristic of the system--the fact that we know exactly what someone who has been assessed can do.

Figure 25.1

**Sample Performance Criteria from an NVQ Element:
Financial Services (Building Societies)--Level 2*
Element title: “Set up new customer accounts”**

Performance Criteria:

- Internal/external documents are complete, accurate and legible and delivered to the next stage in the process schedule
- All signatures/authorisations are obtained to schedule and actioned promptly
- Correspondence to customer is accurate and complete--all necessary documents enclosed--and despatched promptly
- Correspondence to other branches of society and other organisations/professional agencies is accurate and complete--all necessary documents enclosed--and despatched promptly
- Cash transactions and financial documents are processed correctly and treated confidentially
- Computer inputs/outputs are accurate and complete
- On completing the setting up, the account is filed in the correct location
- Indicators of contingencies/problems are referred to an appropriate authority

*Provided as an exemplar in the *Guide to National Vocational Qualifications* (NCVQ, 1991).

It is important to emphasise this objective, because it lies at the heart of the particular interpretation of competence-based assessment that we have experienced over the last decade. A competency-based system will, it has been argued, be far superior to traditional forms because it is so transparent, and because it delivers exactly what is described. And it can *be* delivered because performance criteria are so clearly defined that the assessor can describe a candidate as having unambiguously achieved (or “not yet achieved”) them. The requirement is thus for a one-to-one relationship between criteria and competence, and between assessment and criteria.

It is for this reason that the process of NVQ accreditation does not involve any formal discussion of curriculum (except insofar as it is implicit in the standards) or approval of learning programmes. *The assumption is that use of the standards will ensure the latter’s quality.*

We have noted the *assumption* that assessment will be unproblematic because it simply involves comparing behaviour with the transparent “benchmark” of the performance criteria. The reality, unfortunately, is somewhat different. As a result, the short history of NVQs has also been one in which the quest for clarity has produced an ever more complex and complicated “methodology.”

The second part of the paper discusses the technical reasons for this. Here we simply illustrate it by an example. The criteria in Figure 25.2 are intended to apply to a playgroup assistant or registered childminder. Yet, as they stand, they could equally well apply to a child psychiatrist or specialised speech therapist. How does the assessor know what the “standard” actually is?

This lack of clarity become noticeable fairly early on--well before large numbers of NVQs were actually assessed or delivered. The response was to institute a new notion, that of the “range statement.” These quickly became a compulsory addition to all standards. Range statements officially “describe the limits within which performance to the identified standards is expected, if the individual is to be deemed competent.” In other words, they contextualise the performance criteria, and hopefully make clear whether it is a psychiatrist or a childminder who is in question. They impose further assessment requirements because competence must be fully assessed “across the range.” They also greatly increase the length of the documentation--sometimes taking up as much space as the performance criteria

themselves.

Figure 25.2
Performance Criteria from an NVQ Element:
Childcare and Education--Level 2

Element title: “Help children to recognise and deal with their feelings”

Performance Criteria:	
4.3.1	Children are encouraged to express their feelings in words and actions and through play in the safety of a secure and accepting environment.
4.3.2	Methods and activities used to explore feelings are appropriate to children’s level of development and enable them to begin to recognise, name, and deal with their own and others’ feelings in socially acceptable ways.
4.3.3	Emotional outbursts and negative reactions from children are dealt with in a calm and reassuring manner whilst ensuring the safety of the child concerned and minimising the disruption to other children.
4.3.4	Learning opportunities that arise in the daily routine are used to help children develop their understanding of feelings and social relationships.
4.3.5	Opportunities to help children extend their vocabulary of words relating to feelings are developed where possible.
4.3.6	Any concern over the recognition and expression of feeling in individual children is shared with parents, colleagues, or other professionals as appropriate to the situation.
4.3.7	Ways of expressing and dealing with feelings are demonstrated by the candidate in appropriate situations.

The implications for assessment of this burgeoning detail were horrendous. Suppose one takes, as an example, the elements concerned with finance from the Level 2 NVQ concerned with “business administration.” These comprise about one-seventh of the NVQ as a whole, and incorporate, directly and unamended, the underlying standards.

These elements include 63 different performance criteria. Of these, only a few occur naturally together and so allow of integrated assessment. For example, a candidate must demonstrate that he or she can calculate gross pay and voluntary deductions, must complete returns, must identify discrepancies, and must deal with queries. On top of that, all these criteria must be assessed with respect to each component part of the “range statement.”

The “range statement”--a different one for *each* of the 12 competences in this one unit--is equally important. In our example, we find that, in addition to dealing with income tax, National Insurance and pensions, bonuses, overtime, and a whole range of records, candidates must demonstrate that they can deal with attachment of statutory sick pay, maternity pay, and holiday pay. (One wonders what the accountants would think.) Faced with these demands and with the requirement to demonstrate paperwork on all required outcome measures, an atomised, tick-list system is almost bound to result.

In spite of the level of detail created by range statements, they were rapidly followed by another compulsory addition to standards: “specifications of underpinning knowledge and understanding.” These came in response to a growing concern that NVQs were far too narrow--and also backward looking: something that was in fact virtually guaranteed by requiring lead bodies to focus entirely on analysis of current jobs.

The original architects of NVQs assumed that knowledge requirements would be

clearly understood by trainers and assessors on the basis of the criteria for competence and delivered in an integrated fashion. When critics argued--convincingly--that “underpinning knowledge” was in fact being neglected, a new development programme brought together practitioners to make the assumed consensus explicit. In the event, workshops that tried to “extract” or “induce” knowledge requirements from standards demonstrated quite quickly that the knowledge extracted was not at all standard but was subject to very different interpretations.

Nonetheless, formal “knowledge lists” followed, codifying at least an interpretation of what the standards required and creating further detailed assessment requirements. Finally, the “transparency” of assessment requirements came into question in its turn. Just as “range” and “knowledge statements” have been added to standards, so too have assessment requirements. Industry bodies are now expected to add lists of “assessment specifications” to the standards that examining and awarding bodies use.

Yet another level of detail and centralisation was thus added. The resulting standards and qualification have become huge and unwieldy documents. The apparently economical notion of “competence” has become exhaustively defined and constrained. In the process it becomes increasingly undeliverable and increasingly unattractive to employers as a basis for either their own training programmes or as a way of certifying employees. It also becomes increasingly questionable as a suitable approach for a world of rapid technological change and fluid job boundaries.

The early American experiments on which the English programme drew similarly ended with huge volumes of unmanageable paperwork and over-detailed prescription. The contrast between the apparent simplicity and broadness of “competence” as a concept and the restrictive and rigid reality is remarkable. Yet the tendency is inherent in the idea of completely transparent, unambiguous “outcomes” as an operational idea. It becomes inevitable if one attempts to measure competence precisely and use it as a basis for national (or international) certification and accreditation.

What all this detail has failed to do, however, is to realise the claim made for “competence-based assessment”: to make everything clear and unambiguous, so that “consumers” know exactly what an award-holder can do. In spite of the baleful effects of detailed documentation on the type of skills and abilities being assessed, and in spite of the atomised check-lists, assessors’ interpretations--and measurement practice--have continued to differ. The next section explains why this must inevitably be the case if one relies on paper-based outcome definitions.

The Limitations of “Specified Outcomes”

There are general theoretical reasons why attempts to specify outcomes so clearly that anyone can assess them reliably are doomed to failure. Suppose, for example, that one was interested in something highly specific and abstract--far less context-dependent than the average workplace competency, and so, presumably, easier to define. A very specific mathematics skill is a good example--say the ability to multiply whole numbers.

In the United States, where “criterion-referenced” tests have been most developed, the definitions of items to be used to test such skills have become extremely precise--there must be so many items, of such-and-such difficulty, with so many questions involving one digit (e.g., 3×2), so many involving two digits (e.g., 12×20), and so on. And yet, having constructed such a test, can one really say, with absolute confidence, that “these students can multiply double-digit numbers”? How many errors are they allowed? Would they have done as well on a different set of questions? Does 11×11 count as the “same” as 99×99 ?

Many of the performance criteria in competence-based qualifications are almost as narrow as the examples furnished by academic criterion-referenced tests. We have already referred above to the “level II” NVQ in business administration. (These is one of the larger NVQs in terms of entries, since it is well suited to accrediting specific office skills.) It is from this that the example in Figure 25.1 is drawn. The “range statement” for this competence informs one that:

The Competence includes paper-based filing systems covering the retrieval of information from alphabetical and numerical filing systems, involving indexing systems and lateral and vertical filing methods. It requires competence in booking in and out procedures and the tracing of missing or overdue files.

The assessment guidance adds that, if assessed outside the workplace, students must demonstrate competence by dealing consecutively with a minimum of 20 items to be extracted, on a minimum of three separate occasions. A completely different set of documents must be provided for each simulated assessment.

Yet all these additional requirements simply occasion new queries. Suppose there was a slight overlap in the documents used for assessment. Does that invalidate the assessment? Does it matter if the documents are extracted from a system containing 20 files rather than 2,000? How many of them have to come from files for which document movements are actually recorded? What sort of indexing system counts? And so on.

However “precise” one becomes when one goes down this route, there is always a call for yet more definition. This is exactly the UK experience with NVQs. Performance criteria might mean all sorts of things--so we added range. Range can be interpreted in all sorts of ways--we added more lists. At the end of this process, and in all good faith, people can still be ascribing “competence” to very different behaviour.

The original claim was that “individual performance . . . is judged against explicit standards . . . and (therefore) individuals know exactly what they are aiming to achieve.”⁴ Assessment was seen to require far less in the way of complex judgement than the opaque criteria employed by traditional school-based or higher education.

In fact, nothing could be further from the truth. The inherent variability of the contexts in which competence is tested and displayed means that assessors have to make constant, major decisions about how to take account of that context when judging whether an observed piece of evidence “fits” a defined criterion . In other words, they operate with a complex, internalised, and holistic model--not a simple set of descriptors lifted from a printed set of performance indicators.⁵

The limitations of assessor judgement

If written definitions cannot provide the required clarity, the alternative is to rely on a pre-existing consensus and understanding on the part of the assessors. The whole of competence-based assessment starts from an assumption that there exist “standards of competence” for an industry or role and that these standards can be articulated through written documents. The documents do not create the standards: they articulate and clarify them for professionals. The latter understand them because of their prior knowledge and implicit understanding of what “competence” in their own context means. Thus the developers of “standards of competence” will explain that one goes on defining “as long as it is necessary. You stop when everyone understands.”

We have argued that standards cannot begin to provide item specifications so tight that anyone could use them to construct reliable and consistent assessments. But does this, in fact, matter? Or can a shared occupational culture make the requirements unambiguous so

that the endless spiral of specification which, we argued, was always attendant on criterion-referencing is broken by the existence of shared expertise?

It must be said that there has been very little independent evaluation of whether UK standards are implemented in any comparable or consistent way--and, indeed, that it would be rather difficult to do this at all clearly. By nature, those using them in workplaces will be dealing with very different contexts, so it is not clear how one would measure "sameness" precisely. Nonetheless, one must seriously question whether it is likely even in principle that a combination of definitions and prior consensus will produce any very uniform behaviour and also *whether the assumption of pre-existing "standards" and shared understanding is reasonable at all*. One of the exemplars offered by NCVQ in its guidance comes from publishing and states that *costs are minimised through forward purchase of optimum quantities and timing in relation to schedule requirements*. Is it really likely that, industry-wide, there will be consensus on whether this has been achieved, what would be involved, or how one would recognise it?

Certainly such evidence as exists is not terribly encouraging. Black and his colleagues at the Scottish Council for Research in Education (SCRE) studied in detail the way in which a number of colleges were delivering apparently quite specific stock control modules within the Scottish National Certificate, which also embodies competence-based approaches. All the departments were experienced and had close ties with local industry, and the colleges themselves assumed that the stock control modules would be quite easy to deliver to a common standard. In fact, however, both content and standards deviated greatly within the group.⁶

In research at the Institute of Education,⁷ we asked experienced college tutors and workplace supervisors to devise exercises based on very detailed specifications. In spite of the shared occupational culture of the individuals concerned, the assessment items they produced, following these specifications, proved to be very different in content. We also looked at the level of difficulty at which the assessors ascribed mastery by asking them to administer and make judgements using a more standardised "anchor test" at the same time as they used their own. The standard at which they ascribed "competence" on this common exercise turned out to be markedly different, implying that the underlying standard being applied to the different, and therefore not directly comparable, exercises of their own was also highly variable. Comparable results were obtained with tourist guide examiners operating out of different regional offices, even though they had mostly done their own training together and operated an external examiner system that created some cross-region links.⁸ And a large study of NVQ assessors at work, conducted by the University of Sussex, revealed that a very high proportion related their judgements not simply to the NVQ requirements but to other standards as well, including a variety of pre-existing standards in their own workplace or industry. Many also stated that they made assessment decisions that were not strictly in line with the standards.⁹

Discussions of competence-based assessment often imply that assessor judgement is only a minor issue because the assessment criteria are so minutely and clearly specified that one is well down towards the more mechanistic end of the spectrum. Nothing could be further from the truth. Workplaces vary hugely: thus, any assessment process is complex, incremental, and, above all, judgmental. *It has to be because the actual performance which one observes--directly, or in the form of artifacts--is intrinsically variable*: One person's playing of a piano piece, one person's operations plan, is by definition not exactly the same as another's, and cannot be fitted mechanistically to either a written list of criteria or to an exemplar.

The current approach to competence-based assessment has led us down a cul-de-sac.

This is partly the result of over-ambitiousness regarding what can, or should, be achieved in the way of national uniformity and central control over content and standards; partly because of a failure to understand either the nature of human judgement; and partly because “occupational standards” have been concerned, in practice, not with competence and capability but with the precise task analysis of current jobs. However, it is important to realise that this was one, not “the,” definition of what assessing competence can involve and that a more enlightened model may have a benign rather than a malign effect on practice.

Notes

1. Adapted from Grant et al. (1979).
2. The “lead industry bodies” have now been reconstituted, with additional responsibilities, as “National Training Organisations.”
3. NCVQ (1991).
4. Fletcher (1991).
5. See e.g. Christie & Forrest (1981); Cresswell (1987); Wolf (1995).
6. Black et al. (1989).
7. Wolf & Silver (1986).
8. Wolf (1995).
9. Eraut, Steadman, Trill, & Porkes (1966).

References

- Black, H. et al. (1989). *The Quality of Assessments*. Edinburgh: Scottish Council for Research in Education.
- Christie, T., & Forrest, G. (1981). *Defining Public Examination Standards*. London: Macmillan.
- Cresswell, M. (1987). Describing Examination Performance. *Educational Studies*, 13 (3) 247–265.
- Eraut, M., Steadman, S., Trill, J., & Porkes, J. (1966). *The Assessment of NVQs*. Research Report No. 4. Sussex: University of Sussex Institute of Education.
- Fletcher, S. (1991). *NVQs, Standards and Competence*. London: Kogan Page.
- Grant, G., Elbow, P., Ewens, T., Gamson, Z., Kohli, W., Neumann, W., Olesen, V., & Riesman, D. (1979). *On Competence: A Critical Analysis of Competence-Based Reforms in Higher Education*. San Francisco: Jossey-Bass.
- NCVQ (National Council for Vocational Qualifications). (1991). *Guide to National Vocational Qualifications*. London: NCVQ.
- Wolf, A. (1995). *Competence-Based Assessment*. Buckingham: Open University Press.
- Wolf, A., & Silver, R. (1986). *Work-Based Learning*. Sheffield: Employment Department.