ESRC CENTRE ON
SKILLS, KNOWLEDGE
AND ORGANISATIONAL
PERFORMANCE

The Purposes and Validity of Vocational Qualifications

SKOPE Research Paper No. 105 November 2011

Cathy Stasz

RAND Corporation

Editor's Foreword

SKOPE Publications

This series publishes the work of the members and associates of SKOPE. A formal editorial process ensures that standards of quality and objectivity are maintained.

**Orders for publications should be addressed to the SKOPE Secretary,
School of Social Sciences, Cardiff University, Glamorgan Building,
King Edward VII Avenue, Cardiff CF10 3WT
Research papers can be downloaded from the website:
www.skope.ox.ac.uk**

ISSN 1466-1535

Abstract

UK policy makers continue to focus on increasing skills as a means to economic and social prosperity and social mobility. Qualifications – the certificates and diplomas awarded following education, training or learning – stand as a proxy for skill and many policies aim to raise levels of qualifications held. Policy makers have focused much attention on vocational qualifications and history is replete with efforts to reform the vocational education and training (VET) system. Many studies have identified both strengths and weaknesses in the current system, most recently the Wolf Review (2011). This paper does not attempt to cover old ground, but to look at more basic questions: What are the purposes of vocational qualifications? Are they fit for those purposes? While the first question has been addressed in policy and scholarly circles, less attention has been paid to the second question. This paper draws on literature related to the validity of assessment, because an award of a vocational qualification rests on a candidate's successful performance on particular types of assessment tasks. It examines conceptions of validity and their implications for the interpretation of assessment results. The review shows that judging the validity of vocational qualifications is much more complicated than the architects of National Vocational Qualifications envisioned. The purposes of vocational qualifications have expanded and also vary for different stakeholders. The paper argues that the extent to which vocational qualifications support valid inferences for different purposes remains largely unexplored.

Introduction

In its 2009 report, *Ambition 2020: World Class Skills and Jobs for the UK*, the UK Commission for Employment and Skills (UKCES) set out the aim for the UK to become one of the top countries in the world for jobs, productivity and skills, re-affirming goals set in the Leitch review of skills (2006). The analysis of the UK's current status and the progress needed to attain higher rankings on the international stage rests partly on attainment of qualifications. Qualifications – the certificates and diplomas awarded following education, training or learning – stand as a proxy for skill. And, according to UKCES, the picture looks bleak: 'one in eight adults of working age have no qualifications; more than a quarter are not qualified to Level 2; and just shy of half are not qualified above Level 2' (UKCES 2009: 7). Furthermore, UKCES's projections suggest that the qualifications of UK adults will not improve enough by 2020 to close the international skills gap between the UK and the top countries. Foreign workers may be needed to fill the gap, and as the global market place for human capital expands, qualifications have become central for enabling labour mobility (Johnson and Wolf 2009).

The UKCES analysis is but one of a high stack of government papers that bemoan lack of qualifications as a main stumbling block toward economic prosperity and social equity. In citing poor qualifications as the problem, it is not surprising that most analyses recommend raising qualifications as the solution. To this end the Labour government directed many policies and schemes including: Train to Gain, raising the school leaving age to 18, reform of general and vocational qualifications and accreditation of prior learning (APL).

Policy makers have focused much attention on vocational qualifications, and recent history is replete with efforts to reform the vocational education and training (VET) system. Some have argued that vocational qualifications have been subject to intense government oversight and regulation in recent decades, and even more so than academic qualifications (Sykes 2009). Yet despite these efforts, the UK's relative position in relation to other countries (measured by counting qualifications achieved) has not improved.

Currently, policy makers continue to express concerns about qualifications. Michael Gove, the Education Secretary, commissioned a review of vocational qualifications (the *Wolf Review*, DfE 2011a) that highlighted some successful

programmes but also noted a number of shortcomings. Its overall conclusion is that the vocational education system is still failing many young people. This latest review only adds to the numerous prior studies of vocational qualifications that have identified both strengths and weaknesses in the system of National Vocational Qualifications (NVQs) and in their related learning programmes (e.g. Wolf 1995, Grugulis 2002, Cox 2007, Young 2010).

This paper does not attempt to cover old ground, but aims to address more basic questions about vocational qualifications. It focuses first on the purposes of vocational qualifications. What are vocational qualifications for? For what purposes and functions are they used? Are they meant to certify a level of skill in an occupation or have they become merely a key performance indicator for assessing the government's skills policy? Second, are vocational qualifications fit for those purposes? To address this question, the paper draws on the literature related to validity of assessment, because an award of a vocational qualification rests on a candidate's successful performance on particular assessment tasks. Thus, the paper examines conceptions of validity and their implications for interpretation of assessments, a topic that has received little policy or research attention in relation to vocational qualifications. The paper begins with some background on the rationale for qualifications and on the processes for developing and accrediting qualifications.

Background

The current reasons for promoting attainment of qualifications align with successive governments' aims for increasing skills, as qualifications are seen as a proxy for skill. These are, broadly speaking, both economic and social. For example:

> In developing skills of the current and future workforce, the department's policies are helping to create both a more prosperous society and a more equal one. (Labour government, DFES *et al* 2003:8)

> Skills are vital to our future and improving skills is essential to building sustainable growth and stronger communities. A skilled workforce is necessary to stimulate the private-sector growth that will bring new jobs and new prosperity for people all over this country.
>
> And a strong further education and skills system is fundamental to social mobility, re-opening routes for people from wherever they begin to succeed in work, become confident through becoming accomplished and play a full part in civil society. (Coalition government, DBIS 2010: 3)

> Vocational education is immensely valuable for two, crucial, reasons. First, it is an essential part of a broad curriculum… Second, vocational education is a vital underpinning for our economy. (DfE 2011a)

Like many other nations the UK's skills policy rests on the view that skills, learning and knowledge are the key drivers of competitive advantage in an era of globalisation. From this it follows that policies to increase the supply of certified skills will help improve economic performance.

Qualifications frameworks[1]

A straightforward way to demonstrate that policy is instrumental in increasing the supply of skills is to show that the stock of certified skills is increasing for the targeted individuals. In the current system, certified skills equate to those qualifications recognised in the National Qualifications Framework. The NQF sets out levels against which a qualification can be recognised in England, Wales and Northern Ireland.[2] There are eight levels awarded; levels 4-8 broadly compare to the Framework for Higher Education Qualifications (FHEQ), which covers those qualifications awarded by universities and other HE institutions (see Table A.1, Appendix A).[3] According to this scheme, qualifications at the same level as another are judged to be broadly similar in terms of the demands they place on the learner, however can still be very different in terms of content and duration.

From 2006-8, the regulatory authorities trialled arrangements for a unit and qualifications system underpinned by credit, called the Qualifications and Credit Framework (QCF). The QCF contains new vocational (or work-related) qualifications available in England, Wales and Northern Ireland. In this framework, qualifications are made up of units that are worth credits. Units may be studied at an individual's own pace, are transferable, and can be built up to full qualifications of different sizes over time. Qualifications can be built from different units (required

[1] The information in this background section reflects the regulatory situation prior to May 2010, and is updated where possible according to plans or policies set out by the new Coalition government. Although the Coalition government may eventually propose different policies or programmes than Labour did, the main economic and social purposes for increasing skills remain the same.

[2] The *Wolf Review* calls for all vocational qualifications to be recognised, whether registered in the NQF or not. The DfE has adopted this recommendation and reinstated some qualifications that will be available for teaching in September 2011 (DfE 2011b).

[3] Higher level NVQs and related qualifications will continue to be awarded against the previous NQF levels (QCA/QCAAW/CCEA March 2006. http://www.qcda.gov.uk/libraryAssets/media/qca-06-2298-nqf-web.pdf; Accessed 15 August 2011.

units plus learner choices), provided that combination rules are followed. Some in-house training can also yield QCF units.

Units and qualifications also range in difficulty, from entry level to level 8 (similar to the levels in the NQF). A credit amounts to about ten hours and relates to the overall size of the qualification (award = 1-12 credits; certificate = 13-36 credits; diploma = 37 credits or more). Existing qualifications are being migrated into the QCF, a process that began in September 2010 and is ongoing. As of February 2011, the QCF included over 7500 qualifications (Ofqual 2011). In addition to permitting comparability, the frameworks set the design parameters for qualifications. Policy makers use the frameworks to judge the impact of policy on the types and levels of qualifications attained.

Regulation of qualifications

The development and awarding of qualifications in Britain is a complicated system involving multiple government departments, public and private organisations and elaborate and detailed rules and specifications. The current arrangement is in transition, consisting of a mix of policies set forth in the Labour government and new Coalition government directives. Even though revisions are in process, the overall regulatory framework is likely to change only in the details.

Qualifications are regulated through Section 96 of the Learning and Skills Act 2000, which covers students in maintained schools and young people under 19 years of age in further and work-based education and training. Qualifications approved by the Secretary of State are eligible for funding. This section of the Act is currently under revision.

From 20 July 2011 a new set of criteria for approval came into effect and apply to the consideration of any qualifications for which awarding bodies are seeking approval under Section 96.[4]

- No new qualifications will be approved for use pre-16, except for newly accredited General Certificates of Secondary Education (GCSEs) and equivalent academic qualifications (e.g. International GCSEs). Pre-16 qualifications do not need to be compliant with the QCF.
- Newly accredited qualifications for use post-16 will in general be approved.

[4] Section 96 Interim Arrangements – updated 20 July 2011. See http://www.education.gov.uk/section96/info/index.shtml (Accessed 15 August 2011).

- Only qualifications that meet high-quality standards shall be eligible for consideration for public funding. The Office of Qualifications and Examinations Regulation (Ofqual) sets the standards and it, or a regulatory body recognised by it, will be expected to provide these assurances. Suitability for the age group (in terms of maturity, social well-being and health and safety requirements) will be taken into account.

Approved qualifications for individuals over 19 years of age are outlined in Section 97 in the Learning and Skills Act, which is also currently under revision. Like Section 96 regulations, these regulations determine what can receive available funding. The Skills Funding Agency (SFA) confirms qualifications for public funding that are in line with the Skills Investment Strategy and such parameters as a qualification's size (credit hours) and purpose. The SFA takes advice from Sector Skills Councils, Standards Setting Bodies and Sector Bodies (SSCs).[5] Those confirmed for funding are published on the SFA Web site. While vocational qualifications can be achieved from age 14 through adulthood, the current legislative and funding responsibilities fall to two government departments: The Department for Business, Innovation and Skills has responsibility for post-19 qualifications, while the Department for Education (DfE) oversees pre-19. As the Wolf Review was commissioned by the DfE for pre-19 qualifications, its implications for post-19 qualifications reform is still under consideration.

Ofqual regulates vocational qualifications in England and Northern Ireland (the Department for Education and Skills has this responsibility in Wales). Awarding bodies must be recognised to offer qualifications through a review process to ensure that awarding bodies have the necessary elements in place (e.g. governance, resource, expertise, systems and processes) to develop, assess and award qualifications in a valid, reliable and consistent way. Once recognised, the awarding bodies submit vocational qualifications to the regulators. Ofqual reviews the proposed qualifications to determine that they meet regulatory requirements and have the support of SSCs before accrediting the qualification and entering it into the Register of Regulated Qualifications. Ofqual monitors awarding bodies to ensure that they have systems and procedures in place to deliver qualifications.[6]

[5] Qualifications and Credit Framework—Frequently Asked Questions (FAQ), updated July 2011. Available at http://readingroom.lsc.gov.uk/SFA/QCF_-_July_2011_FAQs.pdf

[6] http://www.ofqual.gov.uk/qualifications-assessments/89-articles/517-vocational-qualifications; updated 10 December 2010 (Accessed 15 August 2011)

The *Criteria for National Vocational Qualifications*, supplemented by the *NVQ Code of Practice 2006*, are used as the basis for the development and accreditation of NVQs.[7] Briefly, the main criteria for NVQs include the following:

- NVQ content must consist of relevant units taken from the National Occupational Standards, developed by the appropriate SSC or sector body and approved by the National Occupational Standards Board.
- Assessment must implement the assessment strategies specified by the SSC/sector body and approved by regulatory authorities.
- Assessment strategies must incorporate:
    - application of specified skills, knowledge and understanding to standards required in the workplace;
    - type and amount of evidence to be collected;
    - identification of aspects that must be assessed through performance in the workplace;
    - clarification of extent of simulated working conditions that can be used in assessment;
    - specification of the occupational expertise of assessors and verifiers. (Their role is discussed further in the next section).
- Awarding bodies must maintain a register of external verifiers.
- Awarding bodies must report in a way that allows for the recognition of NVQ units awarded by other awarding organisations, where the units form part of the NVQs.

It is useful to note that the 'criteria' are really a mix of instructions for how awarding bodies should structure their proposals (e.g. the specific order in which the title of a qualification is to be set out); rules for how awarding bodies are to operate (e.g. they must apply certain assessment strategies); and guidance about what a qualification should include (e.g. content related to National Occupational Standards).

Assessment of National Vocational Qualifications

NVQs were initially developed as competence-based assessments closely aligned with occupational standards. Gilbert Jessup, the recognised architect of the British competence-based approach to qualifications, argued that assessments aligned with

[7] The *Criteria for National Vocational Qualifications*, Available at: http://www.ofqual.gov.uk/downloads/category/90-nvq-criteria); the *NVQ Code of Practice 2006*, plus new amendments, Available at: http://www.ofqual.gov.uk/downloads/category/93-nvq-code-of-practice (Accessed 16 August 2011).

standards would not only convey exactly what an assessor should look for but would also be transparent and understandable to stakeholders. In particular, employers would see vocational qualifications as more valid and more credible, especially if competences were assessed in the workplace (Wolf 1998).

Implementation of this approach, however, requires an assessment methodology involving large numbers of detailed and specific performance criteria. The complexity arises partly because the occupational standards are broken down into units. Units are further divided into elements or groups of performance criteria. Additional knowledge requirements may also be specified. Each of the units that make up the total qualification must be assessed.

To achieve a vocational qualification, candidates must produce evidence to prove that they have developed the necessary level of competence to meet NVQ standards. Specially trained assessors assess the elements of competence, organised into units that make up the award, when the candidates are ready (see example of a level 3 unit in Appendix B). Assessment of competence for each unit is normally through 'performance tasks' consisting of observation of performance on-the-job (or in simulated job-related situations) and questioning of the candidate to assess underpinning knowledge and understanding (Stasz *et al* 2004, Young and Allais 2009). Most NVQs also include a portfolio-type assessment, where candidates assemble or document evidence of competence that assessors evaluate against the standards (Johnson 2008, Wolf 1998). Portfolios might contain the candidate's work, witness testimony, assessment forms, artwork, videos or other types of evidence. If an assessor determines that the evidence meets required standards, the candidate receives a pass grade for the unit.

Ofqual provides guidance to awarding bodies for their assessors and verifiers, and awarding bodies are responsible for ensuring the quality and consistency of assessment for qualifications. Assessors must be trained to apply criteria to diverse sources of evidence, and verifiers must be trained to check comparability of assessor judgments. Several studies indicate that the assessors make judgments, inferences and interpretations of evidence even when comparing it to standard criteria (Eraut, *et al* 1996, Wolf 1998, Konrad 2000, Greatorex 2005, Young 2008, 2010). Verification does not always mean that assessors' *judgments* are verified, but that the accumulated paperwork shows that each individual objective has been assessed. There is also a

moral hazard problem for providers who also act as assessors where the NVQ is the unit of funding and where funds are dependent on its achievement.

Although Ofqual approves and regulates the awarding bodies, the largest organisations are mainly private, for profit organisations (although some are registered charities). The processes adopted to develop and verify assessments linked to awarding of qualifications have not often been subjected to independent research (Wolf 1998, Greatorex 2005).

The decision to adopt a competence-based approach to awarding vocational qualifications adds to the complexity of the system and has set the UK in a particular direction. The emphasis on demonstrating competence through performance specifically downplays the learning process, whereas in other countries the emphasis is on learning as defined by a curriculum or programme of instruction (Brockmann *et al* 2008).

Purposes and Functions of Vocational Qualifications

The literature indicates that vocational qualifications can serve a number of purposes. Government legislation provides its view as to *why* qualifications are important.

> The key purpose of qualifications is to show clearly and publicly the knowledge, skills and attributes that an individual has gained, especially to inform prospective employers and future providers of education and training. The value of qualifications is therefore in very large part defined by how well they are understood by those to whom they may be presented as evidence.[8]

This statement relays three key messages: 1) that a qualification stands for knowledge skills and attributes that an individual has gained, and 2) that a qualification provides that information to consumers, 3) provided that consumers understand what the qualification stands for. This understanding is what makes qualifications transferable to different contexts.

A recent international review of qualification systems conducted for the European Centre for the Development of Vocational Training (Cedefop 2010) noted three broad purposes of qualifications systems:

1. Social reproduction: supporting demarcations in knowledge and skills, promoting particular explicit/implicit values;

[8] http://www.dcsf.gov.uk/section96/info/index.shtml (Accessed 8 February 2010). Although section 96 is under revision, the current government espouses similar views.

2. Structuring pathways to employment and further learning, formalising progression routes and thus providing patterns of incentives for participation in education and training;
3. Shaping learning through affecting the nature, structure and content of learning programmes (Cedefop 2010: 37)

Elaborating on these broad purposes, the study further identified the following eight functions that qualifications serve:[9]

- Promoting learning;
- Responding to social and economic policies;
- Measuring and promoting human capital;
- Equating supply and demand for skills;
- Quality assurance;
- Regulating the national education priorities;
- Creating the conditions for enhanced personal status/identity formation;
- International benchmarking

The review identified further functions within these three broad groupings and led to a taxonomy that lists *40* separate purposes (Cedefop 2010, Table 7). Evidence that vocational qualifications in the UK serve these eight broad purposes and functions can be found in both policy and research literatures (e.g. Eraut 2001, Unwin, *et al* 2004, Young 2010).

Ofqual recently carried out a trial of its own classification of purposes for qualifications (Ofqual 2009). The *Regulatory Arrangements for the Qualifications and Credit Framework* specify that qualification proposals submitted for accreditation must identify the purposes from the list shown in Table 1.[10] Each qualification proposed for accreditation within the QCF must identify one of the main purposes from the list in column one and at least one of the sub-purposes related to the main purpose (column two). The purposes are stated in the form of outputs, which is in line with the outcomes-based design of the QCF. The intent of the trial is to determine

[9] Other comparative studies note a similar list of purposes for qualification frameworks, with different emphases between countries which aim to improve transparency for individual qualifications and those more focused on improving how qualifications are used and understood (Allais 2011).

[10] Ofqual (2009) states that the purposes were to be trialled through summer 2009, and following review of the trial the regulators would decide in what way qualifications within the QCF should identify their purposes in the future. Whether or not the requirement still holds, the list is suggestive of how regulators define purposes of qualifications.

whether the information about purposes helps learners and other users understand what a qualification has been designed to do, but not what it may be used for in the wider policy context.

Table 1: Purposes of Qualifications

A. Recognise personal growth and engagement in learning	A1. Recognise development of skills for life
	A2. Recognise development of knowledge and/or skills to operate independently and effectively in life, learning and work
	A3. Recognise development of personal skills and/or knowledge
	A4. Recognise development of employability skills and knowledge
B. Prepare for further learning or training and/or develop knowledge / skills in a subject area	B1. Prepare for further learning and/or training
	B2. Develop knowledge and/or skills in a subject area
C. Prepare for employment	C1. Prepare for employment in a broad occupational area
	C2. Prepare for employment in a specific occupational area
D. Confirm occupational competence and/or License to Practice	D1. Confirm competence in an occupational role to the standards required
	D2. Confirm the ability to meet a 'licence to practice' or other legal requirements made by the relevant sector, professional or industry body
E. Updating and Continuing Professional Development (CPD)	E1. Update knowledge and/or skills relating to legal, technical, process or best practice changes/requirements
	E2. Develop knowledge and/or skills in order to gain recognition at a higher level or in a different role
	E3. Develop knowledge and/or skills relevant to a particular specialisation within an occupation or set of occupations

Source: Identifying purposes for qualifications in the Qualifications and Credit Framework: Testing and trialling, Ofqual, February 2009.

With the exception of the first purpose – recognise personal growth and engagement in learning – the remaining purposes converge on the notion of

qualifications as 'currency'; a qualification as leading to further education, training, employment or professional standing. Thus, this scheme mainly aligns with Cedefop's second main purpose and with the legislative language quoted at the top of this section. In separating purposes from uses, this scheme does not relate well to eight of the functions of qualifications that Cedefop identified.

It is not altogether clear why Ofqual specifically separates purposes and use of qualifications, as they are clearly intertwined. A qualification's use, as 'currency' in the labour market for example, is also important to individuals and is relevant to whether to pursue a qualification or not (Keep and James 2010). In addition, qualifications have become a 'currency' in the education and training system in that they are now a proxy for individual entitlement to learning, and even more importantly, for funding for providers and their main key performance indicator. These are purposes for which vocational qualifications were not necessarily designed.

The research literature on vocational qualifications in the UK similarly emphasises their purpose as signifiers of an individual's knowledge and skill which have meaning in the education and training market. This is exactly as the initial NVQ designers intended. As summarised by Young (2010: 3) for example:

> In most societies, qualifications are used by students, trainees, employees, employers and admissions tutors (and, of course, education and training providers) both as a proxy for what someone knows and can do and as a 'currency' in the labour market; the more learning is expressed in qualifications, the more it can be 'bought' and 'sold'.

The take up of vocational qualifications (use value) can be can be substantially affected by their currency in the labour market (whether employers want them), their links to professional and career pathways (whether they are valued by professional bodies), the age of candidates and by state-imposed requirements (Unwin *et al* 2004).

Different actors can have different purposes for vocational qualifications. Government policy makers mainly use qualifications as drivers of educational and economic goals (Young and Allais 2009). Qualifications are used to measure and promote human capital and for evaluating and regulating supply and demand for skills, and secondarily for social reproduction – as is evident in the quotations provided earlier (DFES *et al* 2003, DBIS 2010, DfE 2011a) and in other skills policy documents (e.g. Leitch 2006). For the most part, policy makers assume that these economic and social purposes are compatible (Young and Allais 2009). Policy makers

are less directly concerned with quality assurance, leaving this to the Sector Skills Councils, awarding bodies and Ofqual.

Employers see certified vocational qualifications as a guarantee of safe and 'competent' practice (Wolf 1995, Cox 2007). Employers may also use NVQs to promote learning in their workplaces, sometimes as part of a human resource strategy (Cox 2007, Unwin *et al* 2008).

NVQ candidates value the learning interventions associated with NVQs as useful in expanding their knowledge and practices (Tolley *et al* 2003, Torrance, *et al* 2005, Cox 2007). Candidates/learners may seek qualifications as a 'positional good' to get a job or to progress in their chosen field. Even when not pursuing qualifications for their 'currency', candidates may pursue them for motivational reasons or to gain a feeling of accomplishment (Unwin *et al* 2008).

Education and training providers use qualifications as broad guidelines for the design of courses and programmes, so that NVQ holders are ready for work in a particular employment sector (Tolley *et al* 2003).

In sum, vocational qualifications have multiple purposes for different stakeholders, and some commentators have noted the tensions between some purposes (e.g. Young and Allais 2009). As the Cedefop study pointed out, problems can arise if the 'purposes and functions of a single qualification become too numerous, diverse and contradictory. The first casualty of this diversification is the perception of users to the validity of the qualification for their particular priorities' (2010: 20). This brings us to the question of validity.

Validity of Qualifications in Relation to Purposes

Validity refers to the 'degree to which the evidence and theory support the interpretation of test scores entailed by proposed uses of tests' (AERA/APA/NCME 1999: 9). The process of validation involves accumulating evidence for the proposed interpretations of test scores. When test scores are used in more than one way, each intended interpretation must be validated (AERA/APA/NCME 1999[11], Kane 2006).

[11] The *Standards for Educational and Psychological Testing* have been produced through a long-standing collaboration among the American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education to promote the sound and ethical use of tests and to provide a basis for evaluating testing practices. The standards are updated periodically and widely used to evaluate tests, testing practices and the effects of test use. The

There are several aspects of validity to consider with regard to validity of assessments (including performance-based assessments), and these provide an instructive means to consider in what ways vocational qualifications are fit for the different purposes for which they are used. That is, what sorts of accumulated evidence might be used in evaluating a proposed interpretation of a vocational qualification for a particular purpose? Historically, these aspects have sometimes been treated as distinct types of validity (e.g. content validity, predictive validity, construct validity), but validity is now framed as a unifying concept, with the aspects representing different types of validity evidence. It is not necessary to gather evidence on every aspect, but only on those relevant to the proposed use of the assessment result (AERA/APA/NCME 1999, Kane 2006).

One aspect concerns evidence based on test content, which can be obtained from an analysis of the relationship between an assessment's content and the construct it is intended to measure. Does the assessment for the qualification cover the knowledge and skills it is intended to cover? For example, in developing a test that enables license to practice, the major facets of an occupation can be specified (e.g. through a job analysis, functional analysis or accepted occupational standards) and experts in the occupation can be asked to asked to assign assessments to the categories defined by those facets. Another set of qualified experts might then judge the representativeness of the chosen assessments in relation to the occupation as a whole.

A second aspect of validity is evidence based on response processes, or the fit between the construct being measured and detailed nature of the performance or response that examinees engage in (AERA/APA/NCME 1999). For example, if an assessment is meant to measure mathematical reasoning it is important to determine that examinees are, in fact, reasoning about the material and not just applying a memorised algorithm. Evidence based on response processes is generally obtained from analyses of individual responses (e.g. questioning test takers about performance strategies). Experts who are knowledgeable about the underlying process models and their relationship to the construct can review the performance tasks and judge whether the tasks require use of the processes. When assessments rely on observers or judges, as in vocational qualifications, then validation may include review of the guidelines

---

standards are currently under revision, and the author also had access to draft versions that had been posted for comment.

for making subjective judgments; these need to show that scores are based on the successful completion of a process (Miller and Linn 2000).

Validity evidence may be based on an examination of the scoring system as it relates to the construct domain. Multiple options exist for scoring in performance assessments, and scoring is often done with rubrics that focus on pieces of the assessment or on the performance as a whole (Miller and Linn 2000, Young 2010). Scoring methods need to be consistent with the domain and scoring procedures must be implemented consistently to obtain comparable scores across scorers. Approaches to increase reliability of scoring include having multiple scorers, anchors or benchmarks (i.e., standards), adjudication, training and calibration checks. As was pointed out earlier, studies examining the reliability of assessment and verification processes have identified a number of problems with reliability of NVQ scoring (e.g. Wolf 1998, Konrad 2000, Greatorex 2005, Johnson 2008).

A fourth aspect of validity is generalisability, or consistency of assessment results across different conditions of both raters and tasks. It is especially relevant to performance assessments where assessment conditions are usually complex. Generalisability studies may look at the consistency with which raters use scoring rubrics or whether examinees respond differently to different tasks, resulting in a high variance component. It is possible to improve generalisibility of performance tasks by defining the construct and task more narrowly, but this would also limit the inferences that could be made (Miller and Linn 2000, Young 2010). Generalisability can also be improved by increasing the number of tasks, but this may prove impractical for performance assessments because of the extra time required (Kane 2006).

Analyses of the relationship of assessment scores to variables external to the test provide another important source of validity evidence. The idea here is to demonstrate that test results are consistent with the knowledge base and theory of the construct – in this case that having a vocational qualification correlates with success in further related training or job performance. For example, an employer who wants to use a test to assess candidates' suitability for a particular job needs some accepted, valid measure of job performance or criterion, to compare with test performance (Kane 2001, Lissitz and Samuelson 2007). Empirical studies indicate that competence-based assessments in the context of selecting candidates for training generally relate more strongly to job proficiency measures than other types of assessments (Wolf 1995).

The final aspect of validity concerns evidence based on the consequences of testing or assessment. Assessments are administered in the expectation that some benefit will be realised from the intended use of the scores (e.g. placement of workers in suitable jobs or prevention of unqualified workers from entering a profession). A fundamental purpose of validation is to indicate whether specific benefits are realised. Claims are sometimes made for benefits of testing that go beyond the assessments themselves. For example, it is sometimes claimed that tests will improve student motivation or encourage changes in teaching. If such a claim is made, then the validation process should include evidence that these anticipated benefits of assessment are being realised.

For qualification assessments that include performance-assessment scores, such as an NVQ, Kane and colleagues (1999, Kane 2006) suggest that validating the interpretation of performance assessment scores involves three steps or inferences – scoring, generalisation and extrapolation. Their approach recognises the six aspects of validity presented above, but presents them as a set of three inferences. The first inference is from a performance to an observed score and rests on two assumptions: that the criteria used to score the performance are appropriate and applied as intended and that the performance occurred under conditions consistent with an interpretation in terms of the examinee's skill (e.g. there are no inappropriate impediments to performance or inappropriate advantages). Factors that can undermine scoring are numerous: scoring rubrics reflect inappropriate criteria or fail to include relevant criteria; selection or training of scorers may be flawed; faulty equipment; and inappropriate help from the assessor. A second inference concerns generalisation from actual performance on a task to expected performance on similar tasks. Generalisability is threatened when there is variation in observed scores, due to sampling of tasks, raters and occasions or settings. Since NVQ assessments take place in different setting, with varying characteristics, inferences about generalisation may be particularly problematic.[12] Variability of scores across tasks is the most serious problem as it suggests that the scores cannot be generalised beyond the specific set of tasks. A third inference concerns extrapolation of assessment scores to

[12] Studies of the generalisability of performance assessments has generally not been encouraging, especially in terms of sampling error associated with the use of a relatively small number of tasks (Kane *et al* 1999). A similar problem applied to unstructured or 'free choice' portfolios because lack of standardisation makes it impossible to construct scoring rubrics (Stecher 2010).

actual performance in a domain. As the value of a performance test lies in its closer approximation to reality, it is important to determine the degree to which the test requires skills (e.g. problem definition, problem solving) that are essentially the same as those needed for effective performance in practice. For example, if the interest is to assess skill in using mathematics in a variety of 'real-world' situations, evidence needs to be provided to demonstrate that students who are successful in the assessment will be successful in the real situation. If evidence fails to support any of these inferences then validity has not been established. It is not sufficient to assume that assessments that have the appearance of high fidelity or authenticity with the target domain are valid (Messick 1994).[13]

This review indicates that issues concerning the validity of performance assessments, such as those used in awarding vocational qualifications, can be technically complex. Indeed there have been active debates among scholars as to the various definitions of validity and their usefulness (e.g. see Linn 1997 and Mehrens 1997 for discussion of construct versus consequential validity and Kane 1991). However, it is not essential that such differences be sorted out here. The key message is that validity of assessments used to award vocational qualifications cannot be assumed or taken for granted. Rather, it is necessary to carry out validity investigations to ensure that assessment results are properly interpreted in relation to their purposes.

Evidence for validity of vocational qualification assessments

Ofqual has developed five common assessment criteria for evaluating assessments. One of these is validity, which Ofqual defines as follows:

> Validity is the central concept in evaluating the quality of assessment outcomes. It is the overarching concept and subsumes the other concepts.[14] Validity pertains to the arguments or interpretations placed on assessment outcomes, results or scores. It does not relate to the test

[13] Note the tradeoff between the second and third inferences. Extrapolation can be strengthened at the expense of generalisation by making the assessment tasks highly similar to those in the target domain. Conversely, generalisability can be strengthened by employing larger number of tasks, but with possibly somewhat lower fidelity. In designing assessments the goal is to standardise the assessment procedure in ways that strengthen scoring and generalisation, without undermining extrapolation (Kane *et al* 1999).

[14] The others are reliability, comparability, manageability and minimisation of bias (The five common assessment criteria: Quality of assessment outcomes). See http://comment.ofqual.gov.uk/regulatory-framework-for-national-assessments/part-a-the-principles-of-regulation/section-2-criteria-for-regulation/. According to the regulations, Ofqual must inform the Secretary of State of any risks to the validity of assessments. (Accessed 26 July 2011).

> or other assessment procedure itself, nor to the scores generated by that assessment… The evaluation of validity will amount to working out whether the outcomes of the assessment (grades, profile of child and so on) provide adequate information to sustain the argument being made. One needs to evaluate for each interpretation being made. Validity will rarely be an absolute condition; for example one would be more likely to decide that 'the outcomes from this assessment are sufficiently valid' rather than saying they are valid in absolute terms.[15]

This definition agrees broadly with the six aspects of validity just reviewed. Our review of the literature on vocational qualifications, however, revealed few mentions of validity of assessments and even fewer instances to explicitly ask whether the interpretation of assessment results are valid for their purposes. It is quite possible, of course, that awarding bodies routinely validate performance measures as part of their work to develop qualifications and their associated assessments. Cedefop's (2010) review reported that qualification developers do pay attention to construct validity to be clear about what is being assessed and certified, and developers review and evaluate the performance of assessments in relation to both measurement characteristics and purpose. But as commercial organisations awarding bodies are not obliged to publish this work, nor may they wish to for proprietary reasons.

This section summarises the results of a few publicly available studies that examined different aspects of validity. None of them examined all six aspects of validity discussed in the previous section, and some define terms in different ways than those presented. Therefore, the findings have been interpreted in some cases to better align with current conceptions of validity. Problems identified in these studies can be seen as threats to validity.

Achievement of a qualification is often considered as a proxy measure for learning – the evidence for learning gathered in the assessment meets the standards for the qualification (Cedefop 2010, Young 2010). The Cedefop study (2010) rightly pointed out that a qualification's validity needs to be judged on these terms. Although the review does not provide specific validity evidence, it discusses several types of validity: 'construct validity (the thing to be assessed), criterion validity (the thing being assessed) and content validity (whether an assessment covers the knowledge

[15] *Regulatory Framework for National Assessments.* Available at: http://comment.ofqual.gov.uk/regulatory-framework-for-national-assessments/part-a-the-principles-of-regulation/section-2-criteria-for-regulation/ (Accessed 16 August 2011).

and skills it is meant to cover)' (p. 35). As for consequential validity (inferences being made about a qualification can be supported by the assessment), Cedefop found that it is increasingly being explored by assessment agencies and qualification bodies. In discussing different types of validity, note that the report does not follow current conceptions of validity as a unified concept (e.g. AERA/APA/NCME 1999, Kane 2006). Interestingly, it suggests that validity may be interpreted differently in different country contexts. In France and dual system countries like Germany, validity of the certification process also relates to building of professionalism, through immersion in work practices, and therefore assessments might differ for those in traditional programmes versus those seeking certification to recognise prior attainment. In Britain, the same qualifications are meant to serve both purposes.

Torrance and colleagues (2005) investigated whether or not different assessment methods affect learner achievement in the learning and skills sector. Their review found that the greater transparency of competence-based vocational qualifications with respect to learning outcomes and the criteria by which they are judged has supported learners' achievement (number of learners and awards received). But this transparency has also fostered widespread practices that may reduce the validity of qualifications achieved. The assessment practices identified in their research include:

- Assessments reduced to completing a checklist of competences, whereby candidates may not even be aware that they are being assessed.
- Tutors and assessors at the local level sometimes interpret the awarding body specifications and criteria to translate to candidates 'what they really mean'. They may help candidates identify appropriate evidence to record in portfolios. If the users need to translate the assessment specifications into plain English, then to what extent do they really 'represent' workplace standards and competences? Similarly, observed competences have to be translated back again into acceptable evidence statements.
- Examiners provide assistance or help through coaching (e.g. asking leading questions during workplace observations), practice (detailed assignment templates that follow tutors' 'translation' of criteria) and provision of formative feedback (e.g. drafting and re-drafting of assignments) to boost both individual and institutional achievement.
- Opportunities for assessment of competences vary because not all work-based placements offer the same set of experiences (threat to generalisability).

- Work-based assessments are modified because the site does not have the same equipment (threat to generalisability). For example, small garages may not have the electronic equipment to conduct the diagnostic work called for at NVQ Level 3. Therefore, assessors must find 'ways and means' to observe and assess competence. Such modification is one of the 'ways and means' that assessors correct for variability in experiences just discussed.[16]

- Assessors' training is uneven.

Tolley *et al* (2003) examined assessment of NVQ2 and/or NVQ3 in four occupational areas – Hairdressing, Child Care, Engineering and Business Administration – using a variety of qualitative methods (including video recordings). They defined validity of NVQs as 'the adequacy of the standards and the fidelity with which assessment methods reflect those standards' (p. 4). That is, they focused on the first two aspects of validity discussed above. They also examined one aspect of reliability, defined as 'the degree to which the judgments made by assessors and verifiers are free from errors and inconsistencies' (p. 4).

Their research confirms some of the findings reported by Torrance *et al* (2005). In Child Care, they found evidence that candidates were 'overwhelmingly' reliant on college and workplace staff to interpret the standards and their relationship to collecting evidence for assessment purposes (and also Engineering candidates to a lesser extent). Business Administration candidates helped each other complete the materials that are provided as evidence in their portfolios. College tutors helped the Business Administration candidates' record evidence in the appropriate forms and, according to video evidence, sometimes prompted them with clues to help them provide correct answers to questions.

The study also noted variations in opportunities for assessment in the workplace. Business Administration candidates, for example, were mainly assessed under simulated conditions at colleges rather than in workplaces, due to lack of work experience places in the field.

[16] Variability in experiences may lead to use of simulations both in the learning programme and in the assessment situation. Then the question arises as to whether, and to what extent, the simulation matches the occupational standard. For example, changing a clutch in a college workshop because the job has not occurred in the workplace seems a reasonable simulation. But 'pretending' to do a fitness appraisal on a colleague one works with every day seems less appropriate (Torrance *et al* 2005). It is perfectly acceptable for Hairdressing candidates to practice on dummies, as it is not always possible to practice and demonstrate skills on paying customers (Tolley *et al* 2003).

Their research also questioned the fidelity with which assessments reflected standards. In Child Care, for example, the NVQ was contentious: college tutors saw it as a means to ensure readiness to work in the sector, whereas employers were more concerned with training to meet the needs of their own workplaces. In effect, these groups had different ideas about 'standards' and thus different views about the relationship between assessments and standards. These differences in views also indicate that qualifications have different purposes for tutors and employers. The same tension revealed itself in debates over the representativeness of candidates' experiences, whether undertaken in simulated or real workplaces.

Wolf (1998) examined whether portfolio assessments achieved the objectives of validity and national standardisation. Her review noted wide variations in the time candidates spend in workplaces, which meant that it was difficult to integrate assessment with actual workplace performance. Also, differences in the nature of work in occupational sectors make the gathering of portfolio evidence easier for some NVQs than others. For example, business administration requires competence in a number of office tasks (e.g. photocopying, filing, telephone skills) that happen quite readily and can be logged relatively easily as a piece of evidence for assessment. In the construction fields, however, construction sites are less-suited to training and assessment, so NVQs are mainly delivered through college courses or industry training workshops where the emphasis can be on gathering evidence for the portfolio. This kind of variation weakens any claim that assessments are 'more valid' because they are based on standards and occur in the context of work practice.

With respect to validity, the research review focused on issues of reliability and comparability of assessors' judgements. A major independent study of assessment decisions in NVQs in construction, engineering and administration concluded that NVQ assessment was unsatisfactory and produced scores with low reliability. For example, 40 per cent of assessors made judgements without enough evidence, and 38 per cent used standards other than the NVQ as their reference points in assessing portfolio evidence (Eraut *et al* 1996, cited in Wolf 1998). In the NVQ system, external verifiers, appointed by the awarding body, are seen as the main source of quality control over assessment procedures and are responsible for confirming assessors' judgements. The same study found that verifiers did not check whether assessors correctly judged the evidence provided. Rather, assessors were most concerned that assessors' decisions were properly documented (Eraut *et al* 1996,

cited in Wolf 1998). According to Wolf, the verification process had not been subjected to independent empirical study, even though the problems with verification were recognised.

What kind of validity evidence is needed and who should provide it?

The Ofqual regulations state that it will judge whether an awarding organisation's assessments are of an appropriate quality. With respect to validity, the 'focus of awarding organisations' processes and procedures should be on ensuring and generating evidence to support the intended interpretation' (Regulatory Framework for National Assessments, Part A, Section 1). Ofqual has responsibility for monitoring organisations and for determining the frequency and intensity of monitoring based on the organisations' self-reports of their processes and procedures. The regulatory guidance states that the validity argument should be built on statements of proposed interpretation and that supporting evidence be collected from all stages of the assessment process. It expects responsible bodies to maintain and regularly update a 'risk register' that addresses threats to validity of assessments and the quality of the assessment materials. The extent to which Ofqual systematically gathers this evidence is unclear.

QCA, the former regulatory body, carried out 'comparability' studies as part of its programme of quality assurance monitoring of NVQs. These studies examined the consistency and quality of assessments for units in specific qualifications (e.g. QCA 2004b, 2005). In these studies teams of 'scrutineers' visited assessment centres and sampled candidates' portfolio evidence at each. The scrutineer's job was to assess the evidence in terms of validity, authenticity, currency and sufficiency. The studies defined 'validity' as whether the evidence is relevant to what is being assessed. Scrutineers checked the assembled candidate materials (e.g. witness statements, work records, oral questioning, written questioning) to see if the material was referenced to the national occupational standards. In other words, studies focused on evidence based on content and did not develop an interpretive argument about the proposed interpretations and uses of assessment results. The comparability studies identify weaknesses in assessment practices and make recommendations to awarding bodies to improve.

As discussed earlier, a fundamental purpose of validation is to ensure that results of assessments are being used in appropriate ways. There has been

considerable debate as whether those who propose to use the results of an assessment in a particular way (e.g. to make a decision about job placement) need to establish consequential validity. While some argue that this is essential (e.g. Miller and Linn 2000), others (e.g. Popham 1997) argue for a more limited, technical description of validity that involves the descriptive interpretation of test scores. Even those who support the view that evaluation of consequences is an essential component in the validation of test use (including Ofqual, according to its definition of validity) admit that consequences can be far reaching and hard to determine. Furthermore, gathering this type of evidence is challenging. If it is unreasonable or counterproductive to hold a test user or developer responsible for every consequence of test use, then who is responsible (Kane 2001)?

A likely candidate to validate the descriptive interpretation of published tests is the test developer, because some of the descriptive inferences must be examined as part of the development process (e.g. evaluation of scoring keys or rubrics). Test developers that label a test as a 'placement' or 'readiness' test (either explicitly or implicitly) are claiming that a test can be used in a particular way. It seems reasonable to expect the developer to produce a validation argument supporting this use. Developers of placement testing programmes, for example, have traditionally been expected to report data on how the use of the 'placement' scores affected achievement of students placed in different courses (Kane 2001, 2006).

In the NVQ system, and following the expected practice of standardised test developers, then perhaps awarding bodies should be providing validation information? On the other hand, validation research performed by developers of the assessment instrument may be positively biased (Kane 2001). Ofqual's job is to quality assure the awarding bodies, and it (or its predecessors) provides guidance and establishes rules and regulations to do so. As this review has shown, however, independent research and comparability studies point out significant threats to validity in those qualifications studied. With thousands of vocational qualifications in the NQF and thousands of units (the smallest element that can be assessed), this picture does not inspire confidence.[17]

[17] Ofqual carried out a two year study, the Reliability Programme, which examined the consistency of assessments and factors that affect the reliability of results. Our review of the programme's documentation indicates that it mainly examined non-vocational qualifications and did not examine aspects of validity. See http://www.ofqual.gov.uk/standards/research-reports/92-articles/20-reliability (Accessed 4 August 2011).

Decision makers, as qualification users, might also be expected to evaluate the likely consequences, as they know how they are using the assessments, the population being tested and the intended outcomes/consequences. In the NVQ context, government policy makers are a key user of NVQ assessments, because they encourage and support uptake of NVQs for both economic and social purposes. Indeed, government policies can be said to force providers, via their funding systems, to use NVQs rather than other types of vocational qualifications. There is ample evidence that some NVQs are not living up to expectations: occupationally specific vocational qualifications at level 2, for example, generally offer poor or even negative wage returns (DfE 2001a). Yet successive governments have maintained that increasing NVQs, awarded via competence-based assessments, will lead to economic prosperity.

Scholars and psychometricians continue to grapple with the questions of what kind of validity evidence is needed and who should provide it. Current standards advise that both test developer and test users bear some responsibility (AERA/APA/NCME 1999). This paper does not attempt to resolve the issue, but to merely to highlight the questions as still relevant to the continuing debate about national vocational qualifications, their purposes and their uses.

Conclusions

When Gilbert Jessup first argued for awarding vocational qualifications based on occupational standards and assessments of performance on job-related tasks, his main argument focused on a type of validity. If qualifications are to have meaning to employers as representative of performance of work-related skills, he reasoned, then they would be more useful to employers, who would be more likely to accept the qualification as an appropriate signal for labour market preparation. There would be no need to question the construct validity of the assessments because they directly measure what is intended.

As this paper has shown, judging the validity of vocational qualifications based on performance assessments is far more complicated. The purposes of vocational qualifications have expanded beyond the mere demonstration that the candidate can perform up to a standard. And even if the attainment of a qualification is accepted at face value, there are other aspects of validity that can threaten the basic assumptions

that Jessop championed. Over the 20 or so years since competence-based NVQs have been developed, surprisingly little attention has been paid to understanding the implications of anchoring the qualifications on performance-based assessments. Aside from a few noteworthy studies, the validity of claims related to an expanding set of purposes for vocational qualifications has not been thoroughly explored.

Although the qualifications regulator discusses and presumably aims for validity in the broadest sense, in actuality their inspection is somewhat narrow in focusing on determining whether the accumulated evidence for achieving a qualification is complete and relevant. In this regard it does not appear to have advanced much from earlier critiques that question the basic premise: it is possible to develop detailed outcome specifications from standards that can be 'applied reliably, by multiple assessors, to multiple assessment situations' (Wolf 1998). At the same time, the acceptance of performance-based assessments based on occupational standards, and indeed the British NVQ model itself, has been spreading internationally for the same reasons Jessop espoused (Young 2010, Cedefop 2010).

Vocational qualifications still remain a problem in the British educational system, and the Wolf Review's recommendations for improving them focuses on broad issues – their place in 14-19 education; funding arrangements and access; and regulation (DfE 2011a). It partly touches on the discussion in this paper, with regard to the processes for developing and awarding qualifications. The review has been criticised for not getting to the heart of the matter – what counts as good quality vocational education? (Fuller and Unwin 2011). This paper adds another essential, largely unexplored question: To what extent do vocational qualifications support valid inferences in relation to their purposes?

Acknowledgements

My thanks to Laura Hamilton, Brian Stecher and Ewart Keep, who reviewed an earlier draft of this paper and provided many helpful comments and suggestions.

Bibliography

Allais, S.M. (2011) 'The impact and implementation of national qualification frameworks: a comparison of 16 countries', *Journal of Education and Work*, 24 (3/4), 233-258.

AERA/APA/NCME (American Educational Research Association, American Psychological Association, National Council on Measurement in Education) (1999) *Standards for Educational and Psychological Testing*, Washington, DC: AERA.

Brockmann, M., Clarke, L. and Winch, C. (2008) 'Can performance-related outcomes have standards?' *Journal of European Industrial Training*, 32 (2/3), 99-113.

Cedefop (European Centre for the Development of Vocational Training) (2010) *Changing Qualifications: A Review of Qualifications Policies and Practices*, Luxembourg: Publications Office of the European Union.

Cox, A. (2007) 'Revisiting the NVQ debate: 'bad' qualifications, expansive learning environments and prospects for upskilling workers', SKOPE Research Paper No. 71, Cardiff: Cardiff University.

DBIS (Department for Business, Innovation and Skills) (2010) *Skills for Sustainable Growth: Strategy Document.* London: DBIS.

DfE (Department for Education) (2010) *The Importance of Teaching: The Schools White Paper 2010*. London: The Stationery Office.

DfE (2011a) *Review of Vocational Education – The Wolf Report*, DFE-00031-2011, The Stationery Office: Department of Education.

DfE (2011b) *Wolf Review of Vocational Education: Government Response*. Available at: http://www.education.uk.gov/publications (Accessed 25 June 2011).

DfES (Department for Education and Skills/HM Treasury/Department for Trade and Industry/Department for Work and Pensions) (2003) *21st Century Skills - Realising Our Potential - Individuals, Employers, Nation*, CM5810, London: The Stationery Office.

Eraut, M. (2001) 'The role and use of vocational qualifications', *National Institute Economic Review*, 178 (1), 88-98.

Eraut, M., Steadman, S., Trill, J. and Porkes, J. (1996) 'The assessment of NVQs', Research Report No. 4, University of Sussex Institute of Education.

Fuller, A. and Unwin, L. (2011) 'Don't forget the good among the bad and the ugly', *The Times Educational Supplement*, 18 March 2011, p. 6-7.

Greatorex, J. (2005) 'Assessing the evidence: different types of NVQ evidence and their impact on reliability and fairness', *Journal of Vocational Education and Training*, 57 (2), 149-164.

Grugulis, I. (2002) 'Skill and qualification: the contribution of NVQs to raising skill levels', SKOPE Research Paper No. 36, Coventry: SKOPE, University of Warwick.

Innes, E. and Straker, L. (1999) 'Validity of work-related assessments', *Work: A Journal of Prevention, Assessment and Rehabilitation*, 13 (2), 125-152.

Johnson, M. (2008) 'Assessing at the borderline: Judging a vocationally related portfolio holistically', *Issues in Educational Research*, 18 (1), 26-43.

Johnson, S. and Wolf, A. (2009) 'Qualifications and mobility in a globalizing world: Why equivalence matters', *Assessment in Education: Principles, Policy & Practice*, 16 (1), 3-11.

Kane, M. T. (2006) 'Validation', in R. Brennan (ed.) *Educational Measurement* (4th ed.) pp. 17–64, Westport, CT: American Council on Education/Praeger.

Kane, M. (2001) 'Current concerns in validity theory', *Journal of Educational Measurement*, 38 (4), 319-342.

Kane, M., Crooks, T. and Cohen, A. (1999) 'Validating measures of performance', *Educational Measurement: Issues and Practice*, Summer, 5-17.

Keep, E. and James, S. (2010) 'Recruitment and selection – the great neglected topic', SKOPE Research Paper No. 88, Cardiff: SKOPE, Cardiff University.

Konrad, J. (2000) 'Assessment and verification of National Vocational Qualifications: policy and practice', *Journal of Vocational Education and Training*, 52 (2), 225-242.

Leitch, S (2006) *Prosperity for All in the Global Economy – World Class Skills, Final Report*, London: H M Treasury.

Linn, R.L. (1997) 'Evaluating the validity of assessments: The consequences of use', *Educational Measurement: Issues and Practice*, 16 (2), 14-16.

Lissitz, R.W. and Samuelson, K. (2007) 'A suggested change in terminology and emphasis regarding validity and education', *Educational Researcher*, 36 (8), 437-448.

Mehrens, W.A. (1997) 'The consequences of consequential validity', *Educational Measurement: Issues and Practice*, 16 (2), 16-18.

Messick, S. (1994) 'The interplay of evidence and consequences in the validation of performance assessments', *Educational Researcher*, 23 (2), 13-23.

Miller, M.D. and Linn, R.L. (2000) 'Validation of performance-based assessments', *Applied Psychological Measurement*, 4 (4), 367-378.

Ofqual (Office of Qualifications and Examinations Regulation) (2009) *Identifying Purposes for Qualifications in the Qualifications and Credit Framework: Testing and Trialling.* Available at: http://www.ofqual.gov.uk/files/Identifying_ purposes_for_qualifications_in_the_qcf.pdf (Accessed 15 August 2011).

Ofqual (2011a) *The 2010 Evaluation of the Qualifications Credit Framework (QCF) Regulatory Arrangements*. Available at: http://www.ofqual.gov.uk/files/2011-06-03-qcf-evaluation-report.pdf (Accessed 15 August 2011).

Popham, W. J. (1997) 'Consequential validity: Right concern - wrong concept', *Educational Measurement: Issues and Practice*, 16(2), 9-13.

QCA (Qualifications and Curriculum Authority) (2004) *The Statutory Regulation of External Qualifications in England, Wales and Northern Ireland*, London: QCA.

QCA (2004b) *Comparability study of NVQ Assessment Practice (Lifting and Transferring Units)*, London: QCA.

QCA (2005) *Comparability Study of NVQ Assessment Practice (Call Handling Operations)*, London: QCA.

Stasz, C., Hayward, G., Oh, S. and Wright, S. (2004) *Outcomes and Processes in Vocational Learning: A Review of the Literature*, London: Learning and Skills Research Centre.

Stecher, B. (2010) *Performance Assessment in an Era of Standards - Based Accountability*, Stanford, CA: Stanford Center for Opportunity Policy in Education (SCOPE).

Sykes, R (2009) *The Sir Richard Sykes Review*, Commissioned by the Conservative Party. Available at: http://www.conservatives.com (Accessed 26 November 2010).

Tolley, H., Greatbatch, D., Bolton, J. and Warmington, P. (2003) 'Improving occupational learning: the validity and transferability of NVQs in the workplace', Research Report 445, London: Department for Education and Skills.

Torrance, H., Colley, H., Garratt, D., Jarvis, J., Piper, H., Ecclestone, K and James, D. (2005) *The Impact of Different Modes of Assessment on Achievement and Progress in the Learning and Skills Sector*, London: Learning and Skills Research Centre.

UKCES (UK Commission for Employment and Skills) (2009) *Ambition 2020: World Class Skills and Jobs for the UK*, London: UKCES.

Unwin, L., Fuller, A., Turbin, J. and Young, M. (2004) 'What determines the impact of vocational qualifications? A literature review', Centre for Labour Market Studies, University of Leicester.

Unwin, L., Fuller, A., Bishop, D., Felstead, A., Jewson, N. and Kakavelakis, K. (2008) 'Exploring the dangers and benefits of the UK's permissive competence-based approach: the use of vocational qualifications as learning artefacts and tools for measurement in the automotive sector', Learning as Work Research Paper, No 15, Cardiff University.

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

Wolf, A. (1998) 'Portfolio assessment as national policy: the National Council for Vocational Qualifications and its quest for a pedagogical revolution', *Assessment in Education*, 5 (3), 413-445.

Young, M. (2008) 'Towards a European qualifications framework: some cautionary observations', *Journal of European Industrial Training*, 32 (2/3), 127-137.

Young, M. and Allais, S. (2009) *Conceptualizing the Role of Qualifications in Educational Reform*, Geneva: International Labour Organization.

Young, M. (2010) *National Vocational Qualifications in the UK: Their Origins and Legacy*, Geneva: International Labour Organization.

Appendix A: Summary of National Qualifications Framework

Table A.1: The NQF and the FHEQ with Selected Examples

<table>
<tr><th colspan="2">National Qualifications Framework (NQF)</th><th>Framework for Higher Education Qualifications (FHEQ)</th></tr>
<tr><th>Previous levels (and examples)</th><th>Current levels (and examples)</th><th></th></tr>
<tr><td rowspan="2">5
Level 5 NVQ in Construction Management
Level 5 Diploma in Translation</td><td>8
Specialist Awards</td><td>D (doctoral)
Doctorates</td></tr>
<tr><td>7
Level 7 Diploma in Translation</td><td>M (masters)
Masters degrees, postgraduate certificates and diplomas</td></tr>
<tr><td rowspan="3">4
Level 4 NVQ in Advice and Guidance†
Level 4 National Diploma in Professional Production Skills
Level 4 certificate in Early Years</td><td>6
Level 6 National Diploma in Professional Production Skills</td><td>H (honours)
Bachelor degrees, graduate certificates and diplomas</td></tr>
<tr><td>5
Level 5 BTEC Higher National Diploma in 3D Design</td><td>I (intermediate)
Diplomas of HE and FE, foundation degrees, HNDs</td></tr>
<tr><td>4
Level 4 Certificate in Early Years</td><td>C (certificate)
Certificates of higher education</td></tr>
<tr><td>3
Level 3 Certificate in Small Animal Care
Level 3 NVQ in Aeronautical Engineering
A Levels</td><td></td><td></td></tr>
<tr><td>2
Level 2 Diploma for Beauty Specialists
Level 2 NVQ in Agricultural Crop Production
GCSEs Grades A* - C</td><td></td><td></td></tr>
<tr><td>1
Level 1 Certificate in Motor Vehicle Studies
Level 1 NVQ in Bakery
GCSEs Grades D - G</td><td></td><td></td></tr>
<tr><td>Entry
Entry Level Certificate in Adult Literacy</td><td></td><td></td></tr>
</table>

Note: † Revised levels are not currently being implemented for NVQs at levels 4 and 5.

Source: QCA/QCAAW/CCEA March 2006

Appendix B: Example of a Unit Qualification

Unit Reference Number	K/501/7703
Qualification Framework	QCF
Title	Negotiate and agree terms and conditions for the sale of floristry products and services
Unit Level	Level 3
Unit Sub Level	None
Guided learning hours	28
Unit Credit Value	5
Date of Withdrawal	
SSAs	2.2 Horticulture and Floristry
Unit Grading Structure	Pass
Assessment Guidance	As far as possible this unit should be assessed in a workplace environment. In completing this unit learners will need to identify customer requirements for all the following occasions and designs: A. Occasions: i) wedding ii) funeral iii) birth iv) corporate event; B. Designs i) single bridal designs ii) multiple venue decorations iii) personalised funeral tribute iv)gift. Note: To facilitate regional, seasonal and business variations candidates may negotiate and produce alternative design evidence, provided that all occasions are covered and designs are of an equivalent level of complexity. In completing this unit the learner must cover the following currents legislation: HASAW, Sale of Goods Act and COSHH

Learning Outcomes and Assessment Criteria

Learning Outcomes – The Learner Will:	Assessment Criterion – The Learner Can:
Be able to discuss the customer requirements	1.1 Provide customer care in accordance with the business procedures
	1.2 Identify and record the customer's exact requirements through discussion, questioning, sketches and pictures
	1.3 Ascertain the customer's budget for the work and discuss viability
Be able to agree the customer's requirements	2.1 Select and recommend the best design type to meet customer requirements, occupation and budget. Suggest necessary modifications
	2.2 Assess and maximise upon the opportunity for selling-up and/or add-on sales, if appropriate
	2.3 Estimate the amount of time it will take to compete the design
	2.4 Estimate the cost of the design and communicate this clearly to the customers
	2.5 Confirm the final design and cost of the work with the customer
	2.6 Produce written estimates and a quotation to send out to the customer
Price products and services	3.1 Price diverse products and services and calculate appropriate profit margins and VAT
Understand the principles of design	4.1 Explain the basic principles of a design schema and its benefits in floral design covering: order categories, styles of design
Understand the importance of discussing and agreeing customer requirements	5.1 Identify the information needed when taking complex customer orders: Client details, resident details, design requirements, delivery instructions, payment method
	5.2 Explain how cultural differences and practices affect requirements and preferences for floristry products and services
	5.3 Explain how seasonal availability of fresh materials can influence the advice given to customers regarding selection, suitability and value
	5.4 Explain why it is important to outline the additional services available to the customer and what they cost
	5.5 Explain how to identify and interpret the requirements of a design from information provided by customers (such as photographs or memories)
Understand the principals of pricing	6.1 Explain the variables that influence the calculation of an estimated price for a complex design: materials, labour, profit
	6.2 Explain the difference between a quotation and an estimate
Understand the business policies and procedures	7.1 Explain the business procedures regarding customer care
	7.2 Describe business policy on terms and conditions of sale and methods of payment including the requirements for deposits or advance payment and credit arrangements
	7.3 Explain the procedure that should be followed when managing a client problem or complaint
Know the relevant legislation and codes of practice	8.1 State current safe working practices and responsibilities contained within relevant legislation

Source: http://register.ofqual.gov.uk/Unit/Details/K_501_7703