RESEARCH AND ANALYSIS

Responding to Criticisms of the CASLO Approach (Report A)

A taxonomy of potential problems

ofqual

Authors

Paul E. Newton, Milja Curcin, Latoya Clarke and Astera Brylka and from Ofqual's Standards, Research and Analysis Directorate.

With thanks to

With thanks to Chris Winch for his expert advice on the literature.

The CASLO Research Programme

This report is part of a series that arose from Ofqual's 2020 to 2024 programme of research into the CASLO approach:

1. The CASLO Research Programme: Overview of research projects conducted between 2020 and 2024.
2. The CASLO Approach: A design template for vocational and technical qualifications.
3. How 'CASLO' Qualifications Work. (This was published in February 2022.)
4. Origins and Evolution of the CASLO Approach in England: The importance of outcomes and mastery when designing vocational and technical qualifications.
5. Responding to Criticisms of the CASLO Approach (Report A): A taxonomy of potential problems.
6. Responding to Criticisms of the CASLO Approach (Report B): Views from awarding organisations.
7. Responding to Criticisms of the CASLO Approach (Report C): Views from qualification stakeholders.
8. Responding to Criticisms of the CASLO Approach (Report D): Properties of qualifications from the CASLO research programme.
9. Understanding Qualification Design: Insights from the 2020 to 2024 CASLO qualification research programme.

Contents

Introduction

Engaging with the literature has been an important part of our research into the CASLO approach, although it proved to be a challenging literature to engage with. This was partly because the peripheral literature is large and spans many decades: it emerged from North America during the first half of the 20th century, spreading internationally during the second half and into the 21st century. This peripheral literature spans educational movements related to:

- educational objectives
- mastery learning
- criterion-referenced assessment
- outcome-based and competence-based education and training
- outcome-based and competence-based assessment and qualifications
- the more recent international embrace of learning outcomes within vocational education and training, and within higher education

We decided that our core literature should include reports that discussed issues of relevance to understanding or evaluating the CASLO approach, with a specific focus on UK vocational and technical qualifications from the 1980s onwards: academic articles, published and unpublished documents from the 'grey literature', books, book chapters, and so on. Although this restricted focus made it easier to engage with the literature, we still faced numerous challenges.

To begin with, there is not actually a literature on the CASLO approach, per se.[1] Instead, there are reports on qualifications that happened to adopt the CASLO approach – including National Vocational Qualifications, General National Vocational Qualifications, Business and Technician Education Council awards, and so on – and only some of these reports discuss issues of relevance to understanding or evaluating the approach. Indeed, when we searched electronic resources for terms like 'NVQ' and 'BTEC' this generated vast numbers of hits, of which very few were relevant to our interests. This militated against a systematic search strategy.

We decided to adopt an unsystematic approach to identifying our core literature, based upon the idea of snowballing: using reference lists from key reports to identify further potentially relevant reports, then using their reference lists for the same

[1] After all, as we explained in report 4, the CASLO approach is a construct that we introduced for the sake of our research programme, to classify qualifications that shared 3 core design characteristics in common – outcomes, criteria, and mastery.

purpose, and so on. The literature that we identified focused around a relatively small number of vocational and technical qualifications that adopted the CASLO approach, or a very similar one (see below).

The reports that we identified proliferated during the early years of these qualifications, and particularly during the 1990s. Raggatt & Williams (1999) noted how reports that emerged during the early years of the NVQ and GNVQ tended to fall into one of 3 categories:

1. prescriptive (for example, Jessup, 1991)
2. evaluative (for example, Robinson, 1996)
3. competence-focused (for example, Wolf, 1995)

Prescriptive reports explained how and why these qualifications had been introduced – and how they were intended to operate – and were typically written by system architects, like Gilbert Jessup. Evaluative reports typically described investigations into the effectiveness of rollout, often having been commissioned (or undertaken) by government agencies. Competence-focused reports tended to target the viability of the qualification model itself, especially the NVQ competence model.[2]

The National Council for Vocational Qualifications, which introduced NVQs and GNVQs, faced "deep-seated hostility and scepticism" from various parts of the educational establishment (Ecclestone, 1997, page 299). This is consistent with – and may to some extent help explain – our observation that the academic literature linked to these prominent CASLO qualifications tends to be skewed against them.

Unfortunately, it is hard to know exactly what to make of this critical corpus. First, these early CASLO qualifications were highly innovative and idiosyncratic. So, it is hard to judge the extent to which problems that beset their introduction were attributable to the CASLO approach itself, to how it was operationalised within these qualifications, to other more peripheral features of these qualifications, or to how they were rolled out. Bear in mind that both NVQs and GNVQs were significantly revised (more than once) during their early years in response to major implementation problems. Second, the core literature is dominated by reports from the 1990s, so it is also hard to judge the extent to which the criticisms that we identified might generalise to CASLO qualifications of the present day. Third, some of the most interesting (and persuasive) criticisms arise from very detailed but very small-scale research projects. So, it is even hard to tell whether their criticisms would generalise to other qualifications of exactly the same type at exactly the same point

[2] Bates (1995) categorised reports from this core literature slightly differently, and thematically, according to their perspective: evaluative, educational, or sociological (the latter being less well developed).

in time, for example, to qualifications targeting different learners, or different centres, or different employers, or different sectors, or suchlike.

Quandaries of this sort motivated the critical strand of our research programme. The purpose of this strand was twofold:

1. to identify criticisms (from the core literature) that had been levelled against the CASLO approach, and
2. to reflect on the extent to which these criticisms might generalise to current CASLO qualifications

The present report outlines criticisms from the core literature (report 5 – critical strand report A). We reflected on the extent to which criticisms of this sort might generalise by discussing them in detail with awarding organisations (report 6 – critical strand report B) and wider stakeholders (report 7 – critical strand report C).[3]

Recognising that there is no literature on the CASLO approach, per se, the idea of producing a conventional literature review seemed misplaced. Instead, we decided to produce a taxonomy of criticisms that had been levelled against qualifications like NVQs and GNVQs in relation to the characteristics that we (subsequently) associated with the approach. Qualifications of this sort included:

- National Vocational Qualifications (NVQs)
- General National Vocational Qualifications (GNVQs)
- Business and Technician Education Council (BTEC) awards
- Advanced Vocational Certificate of Education (AVCE) awards
- Open College Network (OCN) awards
- Qualifications and Credit Framework (QCF) qualifications

We also considered criticisms that had been levelled at:

- Technician Education Council (TEC) awards
- Business Education Council (BEC) awards

As explained in report 4, we decided that TEC and BEC awards did not fully exhibit all of the core characteristics that we associate with the CASLO approach, although they are critical precursor qualifications, and both were explicitly outcome-based.[4]

Our core literature – which comprised reports on these qualifications of relevance to understanding or evaluating the CASLO approach – proved to be very substantial. However, it consisted predominantly of reports related to NVQs and GNVQs, and

[3] The full list of reports from our research programme is presented at the beginning of this report.

[4] The TEC and the BEC merged in 1983 to become the BTEC.

there were surprisingly few relevant reports related to other qualifications. The literature included a wide variety of criticisms, although it was not immediately obvious how to categorise them. After a period of reflection, we decided to classify them in terms of the kind of problem that they potentially (allegedly) gave rise to. This is a fairly crude taxonomy, but it proved to be useful for structuring our conversations with awarding organisations. The following section describes our taxonomy of potential problems for qualifications that adopt the CASLO approach. We then illustrate each category in turn.

The taxonomy

Figure 1 presents a taxonomy of potential problems for qualifications that adopt the CASLO approach.

Potential assessment problems (A)
Inaccurate judgements
Ineffective standardisation
Atomistic assessor judgements
Poorly conceived assessment tasks or events
Lenience
Malpractice
Inappropriate support
Potential teaching and learning problems (T&L)
Local or personal irrelevance
Lack of currency
Content hard to pin down gets missed
Downward pressure on standards
Incoherent teaching programmes
Lack of holistic learning
Superficial learning
Demotivation and disengagement
Potential delivery problems (D)
Undue assessment burden

Figure 1. A taxonomy of potential problems

This 3-facet taxonomy suggests that adopting the CASLO approach to qualification design can – according to the literature – result in problems for:

1. assessment (A)
2. teaching and learning (T&L)

3. delivery (D)

Potential assessment problems relate to the accuracy of results from CASLO qualifications or, more generally, to the validity of their assessment procedures. Potential teaching and learning problems relate to negative educational impacts that arise from the way in which CASLO qualifications are designed, which are often known as 'washback' or 'backwash' impacts. They include consequences for qualification uptake, for qualification completion, for the effectiveness of teaching, and for the quality of learning. Finally, potential delivery problems relate to the viability, or manageability, of operating a qualification, either for centres, for learners, or for awarding organisations.

Subsequent sections explain each of the 16 potential problem categories in more detail. The purpose of this report is purely to illustrate these 16 categories, rather than to discuss each of them in exhaustive detail. As such, it is important to recognise the following caveats.

First, we do not cite every report that mentions a particular problem. The reports that we do cite, we chose because of their prominence in the literature, or because they illustrated an important variant. Consequently, the reports that appear in the reference list comprise a subset of our core literature, selected for pragmatic reasons.[5]

Likewise, although we hope to have exhausted all of the major potential problems from the literature, we have not necessarily illustrated every possible variant. For example, while we discuss the potential problem of inappropriate support leading to 'false positive' judgements within CASLO qualifications, we do not specifically mention the potential problem of inappropriate hindrance leading to 'false negative' judgements. In this instance, although the literature has certainly identified it as a potential problem – for example, where outcomes are deliberately not signed off by way of punishment for a difficult or unpopular trainee (Wolf, 1995) – it has not received much attention, and was therefore not given its own category (or even referenced specifically in the assessor malpractice section). Similarly, although student malpractice is obviously as much of a concern for the CASLO approach as for any approach that relies heavily upon centre-based assessment, it was less frequently referenced in the core literature than assessor malpractice, which may have something to do with it being a more generic potential problem. Because the potential problem of student malpractice was not linked specifically and consistently to CASLO design features, we did not include it in the taxonomy, despite it being

[5] Relatedly, although we read a very large number of reports for this critical strand, we would not claim to have read (and may not even have identified) all of the core literature. However, we do feel confident in having read enough of it to have reached a point of 'theoretical saturation' in the sense of no longer identifying any substantively new (important) criticisms or potential problems.

mentioned occasionally in the core literature: for example, Bates (1998) mentioned it as a potential problem for NVQs and GNVQs.

Second, we decided not to include any critique of the criticisms that we identified from the literature. This was primarily because we were not interested in specific criticisms that were levelled against specific qualifications of the past. Instead, we were (and still are) interested in the extent to which criticisms of past qualifications might generalise to current qualifications that happen to share certain key features in common. So, rather than attempting to carry forward an historical debate, we have decided to produce a taxonomy of potential problems as a resource for constructing a more contemporary debate.

The important point to appreciate is that many of the criticisms that we identified from the literature have been critiqued in their own right. For example, in the sections that follow, we reference criticism of the NVQ approach from Stewart & Hamlin (1992). Yet, we do not refer to the critique of this report from one the principal architects of the NVQ approach, Mansfield (1993), which was followed by a subsequent (somewhat deferential) response from Stewart & Hamlin (1994). Nor do we mention the extended critique of many of the early criticisms that appears in Burke (1995).

On a more personal note, we do not actually agree with all of the criticisms that we have identified from the literature. Nor do we agree with the seemingly omnipresent 'meta' criticism that appears now to have become a matter of dogma within the international community of technical and vocational education and training scholars: the idea that outcome-based approaches are inherently behaviourist and therefore fundamentally unworkable. We confront this criticism at length in report 4, when unpacking the origins of the CASLO approach in England.

Third, and most obviously, we have discussed neither the presumed benefits of the CASLO approach nor any criticisms associated with alternative approaches. This report is entirely, but intentionally, one-sided. Again, for insights into criticisms that have been levelled at the classical approach to qualification design, and for an analysis of goals that appeared to drive adoption of the CASLO approach in England, report 4 should be consulted.

Fifth, plenty of the criticisms that have been levelled against CASLO qualifications could also be levelled against classical ones. For instance, we reference Torrance, Colley, Garratt, Jarvis, Piper, Ecclestone, & James (2005) for its analysis of superficial learning within qualifications like GNVQs, with evidence of 'criteria compliance' displacing deep learning. Yet, their report is equally critical of similar problems arising within classical qualifications of the period, including A levels.

Finally, we do not provide much insight into the actual prevalence of the potential problems that we identified from the literature, even for the 2 qualifications for which we have most information (NVQ and GNVQ). As noted earlier, some of the most interesting (and persuasive) criticisms arose from very detailed but very small-scale

research projects – especially those concerning negative backwash impacts on teaching and learning – yet the generalisability of conclusions from case studies of this sort remains unclear. For certain of our potential problems, such as the problem of undue assessment burden, there is considerable evidence of it having been a serious issue for certain qualifications (GNVQs in particular). However, for other potential problems, such as the problem of poorly conceived assessments, the evidence is far more limited. For some of our potential problems, the allegation is based more on more conceptual analysis than empirical evidence, such as the problem of atomistic assessor judgements.

The following sections illustrate each of the 16 potential problems in turn.

A: Inaccurate judgements

CASLO qualifications have been criticised for being liable to inaccurate assessment judgements (where learners are judged to have achieved outcomes that they have not actually achieved, or judged not to have achieved outcomes that they have actually achieved). This is said to be a particular problem for qualifications of this sort because they rely so heavily upon written criteria statements for articulating standards. Critics have argued that statements of this sort can never be totally unambiguous, leaving a great deal of room for misinterpretation and misapplication (Wolf, 1995). Faced with evidence of inconsistent application, the temptation is to explicate the criteria in even more written detail. Yet, this risks the criteria becoming unwieldy, unmanageable and, in consequence, unused (Wolf, 1993; Allais, 2014).

In their detailed evaluation of NVQ assessment, a team from the University of Sussex found that 38% of assessors and 48% of external quality assurers believed that many candidates pass who should not (Eraut, Steadman, Trill, & Parkes, 1996). Empirical evidence of judgemental inconsistency provided some backing for these beliefs, for both NVQs and GNVQs (Callender, 1992; Wolf, Burgess, Stott, & Veasey, 1994; QCA, 2005). Having said that, Murphy (1995) suggested that judgemental inaccuracy may be no worse for NVQs than for public examinations.

Commentators have explained problems of this sort in different ways, for example:

- the formulaic approach to articulating outcomes and criteria can seem jargonistic, making them hard to understand and apply (Beaumont, 1996)
- the command verbs that play a critical role in defining standards within criteria are blunt instruments for articulating thresholds (Johnson, 2008a; Newton, 2018)
- assessors working in different contexts (and for different employers) sometimes understand the ‘industry standard’ quite differently (Lester, 2014)
- competence is rarely (if ever) a binary concept, so learners may demonstrate competence in one context but not in another, so there is always a judgement call concerning whether or not a learner has reached a threshold (Eraut, 1994)

The stakes associated with judgemental inaccuracy are potentially high for CASLO qualifications – owing to the generic mastery requirement – because if a learner is incorrectly judged not to have achieved a particular learning outcome (that they have actually achieved) then this alone would be sufficient to prevent them from passing the relevant unit and therefore the relevant qualification (Wolf, 1993).

Reports by Alison Wolf discuss the potential problem of inaccurate judgements in considerable detail (for example, Wolf, 1993; 1995; 1994; 1998).

A: Ineffective standardisation

Standardisation refers to steps that are taken to ensure that assessors' judgements are accurate and remain consistently accurate across assessors, candidates, settings, and so on. Evidence of inaccurate judgement provides reason to question the effectiveness of standardisation activities, materials, and practices.

Standardisation is particularly challenging for CASLO qualifications because judgements need to be accurate at the level of each and every assessment criterion. By way of contrast, when a compensatory aggregation principle is applied (as opposed to the mastery principle) there is scope for judgemental errors to cancel each other out to some extent (as long as they are unbiased errors). This possibility does not exist for CASLO qualifications, as each judgement needs to be accurate in its own right. Consequently, the more units, outcomes, and criteria a qualification incorporates, the greater the pragmatic challenge of standardising judgements. This is magnified when outcomes are written at a high level of generality, to enable them to be acquired and demonstrated in different contexts, which is likely to require new standardisation materials for each context (not simply for each learning outcome).

Despite the obvious potential for judgemental inaccuracy, it was often noted that proponents of the NVQ and GNVQ systems downplayed it, suggesting that the clarity provided by assessment criteria radically reduced it (Wolf, 1995; Garland, 1998). This risked giving the impression that neither training, standardisation, nor quality assurance needed to focus specifically on this threat. Indeed, it was frequently noted that NVQ and GNVQ training, standardisation, and quality assurance practices focused too heavily on processes and too little on judgements (Wolf, Burgess, Stott, & Veasey, 1994; Garland, 1998; Greatorex & Shannon, 2003).

Greatorex & Shannon (2003) investigated the challenge of standardising NVQ assessors' judgements empirically. Their most striking finding was that only one participant in their study had ever taken part in a formal standardisation exercise. A similar lack of standardisation activity was noted by the Joint Awarding Body Steering Group (2001), the QCA (2005), and Curcin, Boyle, May, & Rahman (2013). Greatorex & Shannon identified a variety of practical challenges for NVQ standardisation, including:

- the authenticity of candidate performances (for example, building a staircase, or driving a forklift truck, which might involve observation and questioning) made it impractical for a large number of assessors to scrutinise them simultaneously
- where this problem was circumvented by the use of video recordings, for instance, the candidate would not be available for supplementary questioning
- even within a centre, assessors can be spread across a wide area, and it can be expensive and time consuming to bring them together

A: Atomistic assessor judgements

A fairly common criticism of the CASLO approach relates to the fact that learning outcomes (and assessment criteria) are presented separately, item-by-item, as though they represented entirely discrete achievements. This risks assessors overlooking the interconnectedness of proficiency within a domain of learning, which is normally critical to operating effectively within it. So, the risk associated with atomistic assessment – assessing elements of knowledge or skill, one-by-one, without attending explicitly to their integration and co-ordination – is that students who are unable to apply their learning in a meaningful way still end up being judged competent (as their lack of competence has not actually been revealed). As such, the qualification fails "to provide a guarantee of integrated competence" (Winch, 2023, page 25). Concerns of this sort were raised in policy reviews from the 2010s:

- "Employers tell me that individuals could tick off the many tasks involved but not, at the end, be genuinely employable and fully competent." (Richard, 2012, page 50)
- "The [QCF] format has also encouraged a 'tick box' approach to curriculum and discouraged assessment that confirmed the overall standard had been reached." (Whitehead, 2013, page 18)

As discussed by Murtonen, Gruber, & Lehtinen (2017), this criticism is often linked to the idea that outcome-based qualification design is grounded in a behaviourist epistemology. Drawing upon Bereiter & Scardamalia (1998) they described this as a naïve view of learning outcomes, which treats the acquisition of knowledge and skill as though it were akin to storing records in a mental filing cabinet, which ignores the dynamic development of proficiency structures and conceptions. This behaviourist epistemology has been described by one of its most strident critics as: "not just viciously reductionist but also utterly naive and simplistic" (Hyland, 1993, page 61).

Having concluded that the charge of behaviourism was not actually coherent, Lum (2013; 2015) provided a more sophisticated critique of the outcome-based approach that lay at the heart of the NVQ model. He proposed that assessors who worked in a 'prescriptive mode' were limited to judgements of identity, to confirm whether or not pre-specified criteria had been satisfied, with no further intellectual scrutiny. Conversely, assessors who worked in an 'expansive mode' made judgements of significance, considering all of the available evidence, with no risk of their judgement being constrained by an explicit, but inevitably partial, set of criteria. The NVQ model, he believed, disposed assessors toward the prescriptive mode, tightly constrained by criteria, which risked inaccuracy when circumstances called for the expansive mode. The NVQ approach was particularly likely to err, he believed, owing to its decision to model competence in terms of successful performance rather than in terms of naturally expansive constructs, like knowledge, skill, and understanding.

A: Poorly conceived assessments

For qualifications like NVQs, a key reason for adopting the CASLO approach was to improve the validity of assessment judgements in circumstances that might otherwise have been prone to inaccuracy and inconsistency – including when trainers with little assessment expertise were responsible for in situ assessments. Unfortunately, the extensive scaffolding that was provided by detailed learning outcomes and assessment criteria sometimes made it seem as though the assessment process was essentially "unproblematic" because it simply involved "comparing behaviour with the transparent 'benchmark' of the performance criteria" (Wolf, 1995, page 64). Ironically, by making the assessment process more accessible to assessors, this risked trivialising the challenges that any assessor would face. Furthermore, these outcomes and criteria focused primarily on what to assess, providing no indication of how to assess, including the challenges associated with eliciting evidence in various formats (observation, simulation, questioning, and so on). Note that these qualifications devolved a lot of responsibility for developing and administering assessments to centres (teachers and trainers).

Evidence of poorly conceived assessment practices has been recorded in various reports. For instance, in a study of NVQ assessment, Eraut, et al (1996) raised concerns over the use of ill-designed simulations that led to suboptimal assessments. Murphy (1995) raised similar concerns regarding the potential artificiality of college-based 'authentic' assessments of workplace competence. He also questioned the quality of questioning within NVQ assessments, noting the prevalence of leading questions, unfamiliar questions, predictable questions, and so on (see also Torrance, et al, 2005; Colley & Jarvis, 2007).

Relatedly, in a study of BTEC Nationals, the QCA (2005) identified a variety of issues for centres, which included ensuring that:

- assessment activities were appropriately contextualised to sector pathways, and gave students opportunities to achieve higher grades
- engineering teaching staff were using up-to-date and effective assessment instruments
- teaching staff (particularly part-time and newly qualified teachers) received appropriate training and support in assignment-writing skills
- assessors stopped the inappropriate recycling of assessment from other types of qualification

A: Lenience and malpractice

Although it is important to differentiate between lenience and malpractice (for both conceptual and moral reasons) the kind of malpractice that we have in mind, here, is essentially a more extreme and more intentional manifestation of lenience. Obviously, the potential for lenience and malpractice exists with all qualifications. For CASLO qualifications, these risks might be heightened owing to:

- a tendency to rely heavily upon college-based or workplace-based assessment (associated with the operation of perverse incentives to lower standards)
- ambiguity concerning qualification standards if relying too heavily upon written criteria
- the challenges associated with detecting and correcting incorrect judgements

Assessor lenience

For the purpose of this report, we will define lenience as: inappropriately confirming the acquisition of learning outcomes for learners who have not quite satisfied the relevant criteria, whether doing so intentionally or unintentionally. Wolf (2011) suggested that teachers were incentivised by the post-16 payment system to lower standards within teacher-assessed vocational qualifications (see also Smithers, 1993; Steedman & Hawkins, 1994). Evidence of grade inflation within 'older style' BTEC Nationals suggests that lenience might have been an issue for these qualifications at higher grades (Cuff, Zanini, & Black, 2016).

The potential for lenience was acknowledged by a college lecturer who participated in a case study of a BTEC National in engineering: "nobody is going to disagree with it, so I could let people through that I felt were weak […] there are some grey areas" (Carter & Bathmaker, 2017, page 9). Consistent with observations from Carter & Bathmaker, Johnson (2008b) noted a tendency for OCR National assessors to display a mindset of searching for positive evidence of achievement, and giving students the benefit of the doubt, demonstrating a reluctance to fail students that raised questions concerning the legitimacy of these practices.

In the NVQ context, Wolf explained how social pressures can impact on assessment judgements, especially when the roles of assessor and mentor were not clearly distinguished. She described: "the reluctance of people working in small groups, with a common culture, to criticize, let alone 'fail', the colleagues with whom they work" (Wolf, 1995, page 97). Eraut, et al (1996) noted that assessors sometimes gave students the benefit of the doubt when they believed that assessment criteria were unrealistically demanding. Colley & Jarvis (2007) noted how NVQ motor vehicle assessors took active steps to get 'the good bloke' through – for example,

scaffolding their answers using leading questions – despite evidence that they lacked sufficient competence. This would seem to be at least on the borderline between lenience and malpractice, if not wilfully overstepping it. Garland (1998) noted similarly questionable practices in the GNVQ context.

Assessor malpractice

For the purpose of this report, we will define malpractice as: knowingly confirming the acquisition of learning outcomes for learners who have clearly not yet satisfied the relevant criteria.

Field (1995) described an example of (pig weaning) supervisors signing off relevant NVQ requirements purely on the basis of satisfaction with a trainee's overall level of performance, with little concern for evidencing individual outcomes. This would seem to cross the boundary between lenience and malpractice.

NVQs were judged to be particularly susceptible to fraud, leading Stanton & Bailey to describe them as the "worst possible type of qualification to be used as part of output-related funding" with assessment often being conducted by a supervisor or trainer whose salary depended on the candidate passing (Stanton & Bailey, 2004, page 23). Claims of fraud and corruption related to the award of NVQs were widely discussed during the early years (Hyland, 1997; Garland, 1998).

A: Inappropriate support

One of the principal reasons for adopting an outcome-based approach to qualification design is to facilitate learning by improving the transparency of qualification requirements through detailed explication of learning outcomes and assessment criteria. These outcomes and criteria clarify goals for learning, and the role of the teacher or trainer is to help learners to understand and internalise these goals, and to support them in closing the gap between where they are now and where they need to be. This is the essence of formative assessment (Black & Wiliam, 1998; Sadler, 1989; Wiliam, 2011), which has been described as assessment 'for' learning, in contrast to summative assessment, assessment 'of' learning (Assessment Reform Group, 2002).

Potential problems arise when the line between formative and summative is blurred. While the idea of scaffolding learning is fundamental to formative assessment, it is important for summative assessment not to be similarly scaffolded, to avoid students who have not yet reached the standard being judged to have reached it. Ecclestone has detailed the blurring of this line in the context of GNVQs, BTECs, and related qualifications (Ecclestone, 2002; Ecclestone, Davies, Derrick, & Gawn, 2010).

In the GNVQ context, the risk of inappropriate support would have been high when work of an unsatisfactory standard had been 'referred' back for improvement following an initial summative assessment. Ecclestone (2002) noted how students wanted teachers to be very precise in their feedback, to explain exactly how to 'fill the gaps' related to specific criteria (page 151). This not only risked providing too much support, it also risked focusing attention on the details of specific assessment criteria – as though they were desired outcomes in their own right – thereby diverting attention away from the learning outcome itself. Ecclestone noted that it became futile for GNVQ teachers to offer ideas for improvements that did not relate directly to the criteria. She concluded that teachers in her study had reduced formative assessment to "little more than a pre-emptive extension of summative checking, tracking and evidencing" (page 167).

Carter & Bathmaker (2017) described a number of strategies used by BTEC National lecturers from their case study to scaffold successful performances, particularly at the pass threshold. These included using classroom-based worked examples that were very similar indeed to the tasks that would be used for the subsequent summative assessment (meaning that students could achieve a pass without resorting to additional study outside of the classroom). They also noted the tendency to provide detailed criterion-focused feedback for students whose work had been referred back to them, along the lines of: "you need to do this; this is no good, turn to page 86 in the workbook" (Carter & Bathmaker, 2017, page 11).

T & L: Lack of relevance and currency

The principal purpose of a qualification specification is to explain what candidates need to learn. CASLO qualification specifications do this in great detail, strongly influencing both what gets learnt and what does not get learnt.

Problems can arise when CASLO qualifications specify outcomes that learners are not likely to need in the future. However, if there are no major obstacles to achieving those (superfluous) learning outcomes, then they should still be able to achieve the qualification without too much problem. More serious is the situation in which CASLO qualifications specify outcomes that learners are likely to need in the future, but that are not straightforwardly achievable in their current learning context (for instance, being required to change a clutch in a workshop that had not dealt with repairs of this sort for some time). This was identified as a significant threat to the effective delivery of NVQs in certain situations (for example, Hyland, 1994a; Torrance, et al, 2005).

Potentially even more serious are problems that can arise when CASLO qualifications fail to specify outcomes that learners are likely to need in the future. The obvious consequence of the omission of outcomes of this sort is that learners are likely not to acquire them. Again, this has been identified as a potential problem for NVQs (and National Occupational Standards) as we discuss in more detail below.

Local or personal irrelevance

CASLO qualifications have been criticised for lacking local or personal relevance because of how tightly they pin down learning outcomes, which prevents them from being tailored to the needs of individual learners or employers.

The 2013 Whitehead report discussed “tightly prescribed” National Occupational Standards, which had limited relevance for “small employers or larger employers operating with unique systems and structures” (page 18). This echoed concerns that had been expressed from the outset. For instance, Stewart & Hamlin (1992) questioned the very idea of national standards – given the particular and varied demands of local contexts – citing their experience of management qualifications. Quoting anecdotal evidence from lead bodies, they suggested that many companies supported the general idea of producing national standards, yet deemed them unsuitable for their own organisations. Wolf (1994) proposed that many employers did not recognise these products as embodying the industry standard.

Oates (2004) identified a more subtle, but related, challenge, owing to the potential for building biased (for example, gender-based) presumptions concerning the nature of competent performance into explicit statements of outcomes and criteria, thereby excluding learners with different (yet still legitimate) manifestations of competence.

Lack of currency

CASLO qualifications have also been criticised for lacking real-world currency – once again because of how tightly they pin down learning outcomes – which prevents them from equipping learners adequately either for current or future employment.

Grugulis (2002) identified a problem associated with specifying outcomes tightly, when focusing at the wrong level of detail, noting a tendency for management NVQs to focus upon the minutiae of performing a role in a particular setting (for example, arranging an office ergonomically, or ordering name badges) rather than focusing on the overall meaning and function of the role (for example, developing IT systems, or negotiating pay rates). If candidates are only required to demonstrate the former, then this risks them ending up unable to demonstrate the latter.

Callender (1992) noted a tendency for construction NVQs to be unduly limited in terms of the breadth of activities they referred to and the demands they made on skills, knowledge, and understanding. An emphasis on job performance was tending to squeeze out general job knowledge, and was encouraging rigidity rather than the flexible application of skills. She added that employers "traditionally are more concerned with their short-term needs rather than the long-term needs of their industry, the economy, and individual trainees" (Callender, 1992, page 23). This helped to explain the narrowness of employer-defined standards.

The narrowness of NVQ standards is an issue that Winch has discussed at length (for example, Winch, 2023). He has argued that the NVQ approach focused too much on training and too little on educating, leaving learners ill-prepared for working in broad occupational sectors (as opposed to narrow occupational roles). It is fair to say that there is room for debate concerning the intended breadth of NVQ certification. It is certainly true that NVQs were designed to certificate competence in a specific occupational role, and nothing more than that. In other words, they were intentionally directed at occupational training rather than vocational education. Yet, according to Mansfield & Mitchell (1996), they were always supposed to certificate a broad conception of role competence (derived from functional analysis) and not a narrow one (derived from task analysis). Unfortunately, in practice, they often appeared to reflect a narrow approach, with a consequent negative backwash impact on the breadth of NVQ learning (for example, Brockmann, Clarke, & Winch, 2008).

Finally, and most pragmatically, Hodkinson (1996) observed that NVQ standards were derived from existing (past) work patterns. Combined with the fact that qualification standards, by their very nature, are resistant to being changed quickly, their detailed specification in terms of existing practices makes them ill-suited to the rapidly changing requirements of the labour market.

T & L: Hard-to-pin-down content missed

Ironically, CASLO qualifications have also been criticised for failing to pin down learning outcomes in sufficient detail. In this case, the issue at stake is not the breadth of the learning domain – which might have been appropriately defined – but a failure to specify outcomes that are clearly within scope, yet hard to articulate clearly. Again, the claim is that these outcomes, which are important to acquire yet hard to pin down, end up not being taught or learnt. This criticism appears in a variety of different forms.

One version is the idea, from Stenhouse (1975), that outcome-based approaches are limited to representing predictable (intended) learning outcomes, and that this radically undervalues the importance of unpredictable (unanticipated) outcomes. It is fair to say that this argument has been made more forcefully in the higher education context (for example, O'Brien & David Brancaleone, 2011) than in the technical training context, where the idea of being able to pre-specify learning outcomes has been far less controversial (Kelly, 2009). Having said that, Elliott (2001) has made a similar point in relation to 'higher order' occupational activities, such as management.

A different version of this criticism questions the underpinning logic of outcome-based approaches, which is to elucidate the basis for inferring competence (proficiency, or suchlike) by explaining how it is manifested, that is, what it looks like in practice. Elliott (2001) suggested that there are many aspects of learning that are not directly observable, including "higher order domains of beliefs, attitudes, and values" (page 87). He claimed that 'soft' workplace skills – such as exercising judgement, intuition, weighing up ethical issues, behaviour under stress, intellectual ability, and balancing competing demands – are particularly challenging for outcome-based approaches to deal with (see also Hyland, 1994b).

Yet another version of this criticism relates to the idea of informal, tacit, knowledge. Eraut (1998), for instance, noted the importance of expertise that is developed over time through experience, such as how to handle a particular kind of problem or how to respond to a certain kind of client. Despite being necessary for performance on the job, it exists largely in the form of tacit knowledge, rendering it "relatively impervious to the techniques used by competency-based trainers" (page 132).

Finally, various researchers have discussed the challenge of representing value commitments that are crucial to effective action in many professions, such as teaching, or health and social care (for example, Issitt, 1995; Oates, 2004). Hyland (1997) proposed that the "excessive individualism" of competence-based education and training models marginalised the "collective values of professional work" and thereby served to de-professionalise public sector occupations (page 492).

T & L: Downward pressure on standards

Although not widely discussed in the literature, the potential problem of downward pressure on qualification standards is important. It featured specifically within the 2011 Wolf report.

This potential problem is certainly not unique to CASLO qualifications. It reflects incentives that operate in the post-16 education and training sector independently of how vocational and technical qualifications are designed. Informed by a submission to her review from Norman Gealy, Wolf described how a qualification-based funding regime, devised for institutions that recruit from less able student populations, puts pressure on awarding organisations (that compete for market share) to limit the demands of their qualifications. If they succumb to pressure of this sort, it reduces the effectiveness of education and training provision.

Although not unique to CASLO qualifications, Wolf suggested that this pressure is more acute for them, owing to their stringently applied mastery principle, which means that passing a qualification depends on each and every learning outcome (from each and every unit) having been achieved. This means that "no single element can be difficult" (Wolf, 2011, page 88). In other words, if there is an expectation that any diligent student ought to be able to pass a qualification that they have been appropriately recruited to, this means that no single learning outcome can be pitched at a level that is beyond the reach of the slowest learner from that targeted cohort. Note that this is distinct from the issue of assessor lenience. In this instance, the potential problem relates to the qualification standard itself.

It is fair to say that this criticism holds less force for qualifications that are designed to certificate on-the-job competence, where lowering the standard would render the qualification obviously unfit-for-purpose, and where trainees are given as much time as they need to acquire full occupational competence. It holds more force for off-the-job qualifications that are more removed from the demands of specific jobs, and that offer less flexibility for slower learners who might need additional teaching or training time.

T & L: Incoherent teaching programmes

The principle underpinning outcome-based qualification design is to begin with a comprehensive and authentic specification of intended learning outcomes. Yet, while an outcome specification delineates the end point of a learning journey, it does not explain the nature of progression within a domain. This risks teachers and trainers paying insufficient attention to the challenge of planning for progression, resulting in incoherent teaching and learning programmes.

Hordern (2021) argued that we need to base qualification design on a coherent philosophy of vocational education. Using the example of recently developed apprenticeship standards, he noted that they provided no indication of sequencing of knowledge acquisition, nor of how the apprentice might progress from novice to competent practitioner. Discussing the development of standards for the construction industry, some 3 decades earlier, Callender (1992) noted how (even back then) the declining role of educationalists had led to pedagogical concerns being overlooked, and a failure to map clear paths of vertical progression from one level to the next. Rather than starting with low-level standards and working upwards (or vice versa), the industry had started by developing mid-level standards. Other scholars have developed essentially the same criticism from a more theoretical stance, arguing that the entire NVQ framework was based on a spurious model of progression (Brockmann, et al, 2008; Winch, 2020; Winch, 2023).

The NVQ approach was sometimes ridiculed for appearing to suggest that the very idea of a teaching or learning programme made no sense from an outcome-based perspective (Hyland, 1994a). Although this was not actually true, it is fair to say that architects of the NVQ system left the development of coherent teaching and learning programmes to other people. This risked teachers and trainers mistakenly treating NVQ unit specifications as though they constituted, or could substitute for, coherent teaching and learning programmes, and this mistake had adverse consequences for learners (Stanton, 2012; West, 2004). Observations of this sort led the Commission on Adult Vocational Teaching and Learning to conclude that we “need to put curriculum development and programme design back at the heart of vocational teaching and learning” (CAVTL, 2013, page 14).

Clearly, the task of sequencing a teaching and learning programme is no trivial matter, whether for an outcome-based qualification or for a classical one. For instance, an account of challenges faced by Health and Social Care lecturers noted a common concern that GNVQ specifications did not “explicitly espouse a notion of progression” (Garland, 1998, page 333). It was not at all obvious to them which outcomes built on which, nor how long it might take to reach any particular standard.

Finally, some have argued that a firmer basis in disciplinary knowledge can mitigate the threat of incoherent teaching programmes (Wheelahan, 2016; Hobley, 2016).

T & L: Lack of holistic learning

The potential problem of lacking holistic learning is the flipside of the potential problem of atomistic assessment. This is the risk that teaching elements of knowledge or skill one-by-one, without attending explicitly to their integration and co-ordination, results in students being unable to apply their learning in a meaningful way. This derives from some of the earliest critiques, which focused on the idea that standards for BTECs, NVQs, and GNVQs lacked integration:

- “Lists of objectives, by any standard, are poor representations of the overall structure of knowledge, ideas, and values. They seem to portray little heaps of knowledge, rather than an integrated structure or matrix.” (Bull, 1985, page 77)
- “atomism remains in the idea that elements of competence sum to give overall competence” (Ashworth & Saxton, 1990, page 12)

Discussing the secretarial curriculum, Waymark (1997) proposed that the atomisation of NVQ specifications led to “a diminution of the secretarial role” (page 114). For Waymark, the holistic aspect of secretarial competence – which included prioritising tasks, completing them against deadlines, and simultaneously coping with interruptions – held special importance. Yet, as a consequence of atomised NVQ specifications, this aspect of the secretarial role was being neglected.

Helsby, Knight, & Saunders (1998) made similar observations in the context of Advanced GNVQs. They proposed that complex GNVQ course specifications encouraged atomisation of content, which militates “against holistic understanding by students and coherent course planning by teachers” and constrains “their power to create connections and meanings” (page 71).

Carter & Bathmaker (2017) reached similar conclusions in the context of a BTEC National in engineering. This included a suggestion from one lecturer that the “piece-meal approach” of the National award reduced preparedness for the Higher National, which he linked to the idea of “passing that particular bit and forgetting it [which results in students who] cannot put it together” (page 9). The risk of establishing training regimes that make it possible for apprentices to “pass and forget” learning outcomes from one unit to the next has been identified as a particular threat for CASLO qualifications (see Alton, Boyle & Limmer, 2021, page 8). This relates both to the previous potential problem (incoherent teaching) and to the following one (superficial learning).

T & L: Superficial learning

CASLO qualification specifications explicate what candidates need to learn in great detail, as well as specifying the exact criteria by which their learning will be judged. Although this has the potential to empower learners, enabling them to take greater control over their learning, it also risks degenerating into instrumentalism, where learners become unduly focused on satisfying criteria and insufficiently focused on learning itself. This can result in shallow, superficial learning.

Tolley, Greatbatch, Bolton, & Warmington (2003) argued that it was essential for NVQ candidates to engage in learning that went "beyond the immediate requirements of the situation" (page 8). Yet, they also noted that both candidates and their employers (although not college tutors) saw assessment primarily as a mechanism for 'signing off' on portfolios rather than also as a site of further learning. Hyland (1994b) proposed that the NVQ model effectively replaced genuine learning and development with the generation and collection of assessment evidence.

The charge of instrumentalism has received particular attention in the context of GNVQ delivery. Smith (1998), for example, suggested that the bureaucracy of assessment may have fostered undue dependency on GNVQ teachers, encouraging students to become a "task completer" (page 547) rather than an independent reflective learner. Bates (1998) also lamented the bureaucracy of assessment, which had turned students into "hunters and gatherers" of information (page 199). This consumed excessive amounts of time and disincentivised working independently.

The charge of instrumentalism was developed most forcefully by Ecclestone and colleagues during the 2000s. Ecclestone's 2002 study of autonomy in GNVQ learning included a chapter on 'getting through' that began with a reminder that GNVQs were part of a solution to the new challenge of having to "engage the 69 per cent of 16-year-olds" who now stayed on in full-time education (page 107). She believed that increasing numbers were likely to be 'pragmatically compliant' at best and 'hanging on' at worst. For instance, she noted that most of the students in her study aimed low, only targeting easily achievable grades. They tended to produce "compliant statements geared directly to each criterion" and resisted "extraneous or irrelevant" work (page 142). Similarly, Garland (1998) described GNVQ students who used a "formula" (page 338) to evidence high-level criteria, which enabled them to jump through the necessary hoops despite their assignments not actually being very good.

Following a detailed review of assessment in the learning and skills sector, Torrance, et al (2005) argued that transparent assessment criteria encourage instrumentalism. Coupled with extensive coaching and practice, this risked removing the challenge of learning. Pejoratively, they described a situation in which 'criteria compliance' had come to replace genuine learning as "assessment as learning" (page 2).

T & L: Demotivation and disengagement

From one perspective, the CASLO approach has the potential to motivate and engage learners. For instance, the idea of unitisation is often assumed to be attractive to learners, as it helps to break the learning journey into a series of accomplishments, which can be achieved gradually over time (as opposed to the terminal 'single shot' exam format). The idea of breaking units down even further, into sets of learning outcomes, offers more potential to segment the learning journey into even smaller steps, which can be particularly motivating for learners in challenging circumstances, such as 'returning' adult learners who are taking a 'second chance' on education and training (Wilson, 2010). Finally, where learning outcomes are written with a sufficient level of generality to permit tailoring of content to personal interests or local employment needs, this has been assumed to sustain motivation to learn.

The GNVQ experience is important in relation to this potential problem, as one of its key purposes was to secure motivation and autonomy through qualification design features, including its emphasis on independent learning and formative assessment. While acknowledging the potential for successful rollout, a case study of Advanced GNVQ by Bates (1998) identified challenges related to securing motivation and autonomy, encapsulated in the idea of students "resisting" becoming empowered (page 187). She observed an attitude of instrumentalism in completing assessment requirements, which was (perversely) incentivised by: the requirement to produce large amounts of evidence, the potential for contriving evidence, large backlogs of work, and so on.

Ecclestone (2002) tackled similar issues in her own detailed case study of the Advanced GNVQ. She emphasised the need to be more precise, and then more strategic, about how to secure autonomy, noting how low expectations and micro-disciplinary practices in assessment can lead to compliance, and risk aversion, rather than genuine empowerment (see also Torrance, et al, 2005; Wellings, Spours, & Ireson, 2010). Ecclestone also mentioned issues related to how the CASLO approach was operationalised within GNVQs, such as the demotivating impact of having to re-submit "atomised bits" of large assignments in order to pass (page 163).

The challenge of sustaining motivation and engagement in the face of a demanding assessment regime has also been identified in relation to NVQs. For instance, Colley & Jarvis (2007) identified students who were considering abandoning their NVQ studies, particularly where private training providers failed to provide adequate support in areas like portfolio building. One final issue of relevance to the NVQ model relates to its lack of grading. Wolf (1993) argued that pass or fail qualifications were problematic owing to a tendency for teachers to teach for minimum competence, which fails to motivate faster learners to achieve deeper learning.

D: Undue assessment burden

The risk of undue assessment burden is obviously high for CASLO qualifications because of their stringent application of the mastery principle, which requires the assessment of all specified learning outcomes – not to mention all of the ramifications of this exhaustive assessment process related to record keeping, storing of evidence, and quality assurance. Assessment burden has received a lot of attention in the literature. The more granular the learning outcomes and assessment criteria, the greater the risk of undue burden.

Depending on how undue burden is dealt with, it can manifest as a teaching and learning problem as well as a straightforward delivery one. This happens when assessment, recording, storing, and quality assurance activities eat into time that has been allocated for teaching and learning, thereby lowering student achievement and threatening the successful completion of a qualification.

Undue burden was identified as a major challenge during the early years of both NVQ and GNVQ rollout. For instance, the GNVQ Assessment Review Project report described the amount of time required for assessment as “unsustainable” and described documentation and paperwork as “inherently unwieldy” (Wolf, Burgess, Stott, & Veasey, 1994, page 2). It also noted that centres were struggling to find space to physically store assessment evidence. The burden of GNVQ assessment was exacerbated by a requirement to record achievement at a level of detail below the assessment criterion, that is, at the range statement level. The 1995 Capey review noted agreement from all sources, particularly practitioners, that GNVQ assessment and recording requirements were “far too detailed and counter-productive in terms of teaching and learning” (Capey, 1995, page 23).

Stewart & Hamlin (1992) discussed NVQ assessment in similar terms, describing a complex paper-chain of bureaucracy spread across multiple stages: referring to the standards, collecting evidence, comparing evidence against the standards, recording the comparison, signing-off outcome achievement and unit completion, submitting confirmatory statements for the overall award. The 1996 Beaumont review reached a similar conclusion, observing that NVQ bureaucracy had been widely criticised, and proposing that excessive bureaucracy would have to be eliminated. Some years later, the 2012 Richard report railed against the continuous “and time consuming assessment” of apprentices via NVQs and related qualifications (page 67). Indeed, it proposed that much of the time that apprentices spent ‘training’ was actually spent “with their assessor providing evidence of their ability to meet competency requirements” (page 87).

Conclusion

The purpose of this report was simply to unpack the various criticisms that we identified from the literature, which we elaborated in terms of potential problems for CASLO qualifications. As such, there is little to say by way of conclusion, other than to remind the reader that the purpose of developing our taxonomy was to support conversations with awarding organisations concerning their current CASLO qualifications. Outcomes from these conversations, and from conversations with wider stakeholders, can be found in reports 6 and 7 (critical strand reports B and C).

References

Allais, S. (2014). *Selling Out Education: National qualifications frameworks and the neglect of knowledge*. Rotterdam, The Netherlands: Sense Publishers.

Alton, A., Boyle, A. & Limmer, H. (2021). *Research into Assessments of Occupational or Practical Competence Outside of Regulated Qualifications: Review and case studies*. Manchester: AlphaPlus.

Ashworth, P.D. & Saxton, J. (1990). On 'Competence'. *Journal of Further and Higher Education*, 14(2), 3–25.

Assessment Reform Group (2002). *Assessment for Learning: 10 principles. Research-based principles to guide classroom practice*. London: Assessment Reform Group.

Bates, I. (1995). The competence movement and the National Vocational Qualification framework: The widening parameters of research. *British Journal of Education and Work*, 8(2), 5–13.

Bates, I. (1998). Resisting 'empowerment' and realizing power: An exploration of aspects of the GNVQ. *Journal of Education and Work*, 11(2), 187–204.

Beaumont, G. (1996). *Review of 100 NVQs and SVQs*. London: National Council for Vocational Qualifications.

Bereiter, C. & Scardamalia, M. (2005). Beyond Bloom's taxonomy: Rethinking knowledge for the knowledge age. In M. Fullan (Ed.). *International Handbook of Educational Change: Fundamental change* (pp.5–22). Dordrecht, The Netherlands: Springer.

Black, P.J. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.

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

Bull, H. (1985). The use of behavioural objectives: A moral issue? *Journal of Further and Higher Education*, 9(3), 74–80.

Burke, J. (1995) (Ed.). *Outcomes, Learning and the Curriculum: Implications for NVQ's, GNVQ's and other qualifications*. London: The Falmer Press.

Callender, C. (1992). *Will National Vocational Qualifications Work? Evidence from the construction industry*. IMS Report No 228. East Sussex: Institute of Manpower Studies.

Capey, J. (1995). *GNVQ Assessment Review: Final report of the review group chaired by Dr John Capey*. London: National Council for Vocational Qualifications.

Carter, A. & Bathmaker, A.M. (2017). Prioritising progression over proficiency: Limitations of teacher-based assessment within technician-level vocational education. *Journal of Further and Higher Education*, 41(4), 460–474.

CAVTL (Commission on Adult Vocational Teaching and Learning) (2013). *It's About Work… Excellent adult vocational teaching and learning*. London: Learning and Skills Improvement Service.

Colley, H. & Jarvis, J (2007). Formality and informality in the summative assessment of motor vehicle apprentices: A case study. *Assessment in Education: Principles, Policy & Practice*, 14(3), 295–314.

Cuff, B.M.P., Zanini, N., & Black, B. (2018). *An Exploration of Grade Inflation in 'Older Style' Level 3 BTEC Nationals: 2006 to 2016*. Coventry: Office of Qualifications and Examinations Regulation.

Curcin, C., Boyle, A., May, T., & Rahman, Z. (2013). *A Validation Framework for Work-Based Observational Assessment in Vocational Qualifications*. Coventry: Office of Qualifications and Examinations Regulation.

Ecclestone, K. (1997). On the frontline or at the margins? Finding a place in post-compulsory research. *Research in Post-Compulsory Education*, 2(3), 299–312.

Ecclestone, K. (2002). *Learning Autonomy in Post-16 Education: The politics and practice of formative assessment*. Oxford: RoutledgeFalmer.

Ecclestone, K., with Davies, J., Derrick, J., & Gawn, J. (2010). *Transformative Formative Assessment in Lifelong Learning*. Berkshire: Open University Press.

Elliott, G. (2001). An alternative model of competence-based assessment. In C. Velde (Ed.). *International Perspectives on Competence in the Workplace* (pp.85–98). Dordrecht, Netherlands: Springer.

Eraut, M. (1994). *Developing Professional Knowledge and Competence*. London: RoutledgeFalmer.

Eraut, M. (1998). Concepts of competence. *Journal of Interprofessional Care*, 12(2), 127–139.

Eraut, M., Steadman, S., Trill, J., & Parkes, J. (1996). *The Assessment of NVQs*. Brighton: University of Sussex Institute of Education.

Field, J. (1995). Reality testing in the workplace: Are NVQs 'Employment-led'? In P. Hodkinson & M. Issitt (Eds.). *The Challenge of Competence: Professionalism through vocational education and training* (pp.28–43). London: Cassell.

Garland, P. (1998). Assessment in GNVQs: learning the hard way. *Research in Post-Compulsory Education*, 3(3), 329–344.

Greatorex, J. & Shannon, M. (2003). *How can NVQ assessors' judgements be standardised*? Presented at the British Educational Research Association Annual Conference, 11–13 September, Heriot-Watt University, Edinburgh.

Grugulis, I. (2002). *Skill and Qualification: The contribution of NVQs to raising skill levels*. SKOPE Research Paper No.36. Coventry: Centre on Skills, Knowledge and Organisational Performance, University of Warwick.

Helsby, G., Knight, P., & Saunders, M. (1998). Preparing students for the new work order: The case of Advanced General National Vocational Qualifications. *British Educational Research Journal*, 24(1), 63–78

Hobley, J. (2016) 'Here's the iPad'. The BTEC philosophy: How not to teach science to vocational students. *Research in Post-Compulsory Education*, 21(4), 434–446.

Hodkinson, P. (1996). National Vocational Qualifications – A way forward? In R. Halsall & M. Cockett (Eds.). *Education and Training 14–19: Chaos or coherence* (pp.17–32)? London: David Fulton Publishers.

Hordern, J. (2021). Are apprenticeship standards in England supporting expert vocational practice? In Cedefop/OECD (Eds.). *The Next Steps for Apprenticeship* (pp.144–151). Luxembourg: Publications Office of the European Union.

Hyland, T. (1993). Competence, knowledge, and education. *Journal of Philosophy of Education*, 27(1), 57–68.

Hyland, T. (1994a). Tilting at windmills: the problems of challenging the National Council for Vocational Qualifications. *Educational Studies*, 20(2), 251–265.

Hyland, T. (1994b). *Competence, Education and NVQs: Dissenting perspectives*. London: Cassell.

Hyland, T. (1997). Reconsidering Competence. *Journal of Philosophy of Education*, 31(3), 491–503.

Issitt, M. (1995). Competence, professionalism and equal opportunities. In P. Hodkinson & M. Issitt (Eds.). *The Challenge of Competence: Professionalism through vocational education and training* (pp.70–85). London: Cassell.

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

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

Johnson, M. (2008b). Exploring assessor consistency in a Health and Social Care qualification using a sociocultural perspective. *Journal of Vocational Education & Training*, 60(2), 173–187.

Joint Awarding Body Steering Group (2001). *Joint Awarding Body Guidance on Internal Verification of NVQs*. London: Department for Education and Skills.

Kelly, A.V. (2009). *The Curriculum: Theory and practice* (6th edition). London: SAGE Publications Ltd.

Lester, S. (2014). Professional versus occupational models of work competence. *Research in Post-Compulsory Education*, 19(3), 276–286.

Lum, G. (2013). Competence: A tale of two constructs. *Educational Philosophy and Theory: Incorporating ACCESS*, 45(12), 1193–1204.

Lum, G. (2015). On the functions of a system of vocational qualifications. *Journal of Education and Work*, 28(2), 187–206.

Mansfield, B. & Mitchell, L. (1996). *Towards a Competent Workforce*. Aldershot, Hampshire: Gower Publishing Ltd.

Mansfield, B. (1993). Competence-based qualifications: A Response. *Journal of European Industrial Training*, 17(3), 19–22.

Murphy, R. (1995). *The Reliability of Assessment of NVQs: Report presented to the National Council for Vocational Qualifications*. Nottingham: School of Education, University of Nottingham.

Murtonen, M., Gruber, H., & Lehtinen, E. (2017). The return of behaviourist epistemology: A review of learning outcomes studies. *Educational Research Review*, 22, 114–128.

Newton, P.E. (2018). *Grading Vocational & Technical Qualifications: Recent policies and current practices.* Ofqual/18/6441/3. Coventry: Office of Qualifications and Examinations Regulation.

O'Brien, S. & Brancaleone, D. (2011). Evaluating learning outcomes: in search of lost knowledge. *Irish Educational Studies*, 30(1), 5–21.

Oates, T. (2004). The role of outcomes-based national qualifications in the development of an effective vocational education and training system: The case of England and Wales. *Policy Futures in Education*, 2(1), 53–71.

QCA (Qualifications and Curriculum Authority) (2005). *Comparability Study of Assessment Practice: BTEC National qualifications in business, engineering and media*. QCA/05/1681. London: Qualifications and Curriculum Authority.

Raggatt, P. & Williams, S. (1999). *Government, Markets and Vocational Qualifications: An anatomy of policy*. London: Falmer Press.

Richard, D. (2012). *Richard Review of Apprenticeships*. Bootle: School for Startups.

Robinson, P. (1996). *Rhetoric and Reality: Britain's new vocational qualifications*. London: Centre for Economic Performance, LSE.

Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144.

Smith, J. (1998). Beyond the rhetoric: Are General National Vocational Qualifications (GNVQs) doing students any good? *Journal of Vocational Education & Training*, 50(4), 537–548.

Smithers, A. (1993). *All Our Futures: Britain's education revolution*. London: Channel Four Television.

Stanton, G. & Bailey, B. (2004). Fit for purpose? Sixty years of VET policy in England. In G. Hayward & S. James (Eds.). *Balancing the Skills Equation: Key issues and challenges for policy and practice*. Bristol: Bristol University Press, Policy Press.

Stanton, G. (2012). *Course Design and Pedagogy in Adult Vocational Education: Where are we now and where have we come from?* CAVTL 1st Commissioners' Meeting, 27 June 2012, paper 6. Unpublished.

Steedman, H. & Hawkins, J. (1994). Shifting foundations: The impact of NVQs on youth training for the building trades. *National Institute Economic Review*, 149(1), 93–102.

Stenhouse, L. (1975). *An Introduction to Curriculum Research and Development.* London: Heinemann Educational Books Ltd.

Stewart, J. & Hamlin, B. (1992). Competence-based qualifications: The case against change. *Journal of European Industrial Training*, 16(7), 21–32.

Stewart, J. & Hamlin, B. (1992). Competence-based qualifications: A reply to Bob Mansfield. *Journal of European Industrial Training*, 18(1), 27–30.

Tolley, H. Greatbatch, D., Bolton, J., & Warmington, P. (2003). *Improving Occupational Learning: The validity and transferability of NVQs in the workplace*. Research Report RR425. London: Department for Education and Skills.

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

Waymark, M. (1997). The impact of National Vocational Qualifications on the secretarial curriculum. *Journal of Vocational Education and Training*, 49(1), 107–120.

Wellings, D. Spours, K., & Ireson, J. (2010). Advanced GNVQs, AVCEs and Level 3 Diplomas in England: A motivational analysis. *Research in Post-Compulsory Education*, 15(4), 387–403.

West, J. (2004). *Dreams and Nightmares: The NVQ experience.* CLMS Working Paper No 45. Leicester: Centre for Labour Market Studies, University of Leicester.

Wheelahan, L. (2016). Patching bits won't fix vocational education in Australia – a new model is needed. *International Journal of Training Research*, 14(3), 180–196.

Whitehead, N. (2013). *Review of Adult Vocational Qualifications in England*. London: UK Commission for Employment and Skills.

Wiliam, D. (2011). *Embedded Formative Assessment*. Bloomington, IN: Solution Tree.

Wilson, P. (2010). *Big Idea, Small Steps: The making of credit-based qualifications*. Leicester: National Institute of Adult Continuing Education.

Winch, C. (2020). Learning outcomes, progression, and qualifications: Considerations for vocational and higher education in the United Kingdom. In C. Arnold, M. Wilson, J. Bridge, & M.C. Lennon (Eds.). *Learning Outcomes, Academic Credit, and Student Mobility* (pp.167–197). Kingston, Canada: McGill-Queen's University Press.

Winch, C. (2023). Learning outcomes. The long goodbye: Vocational qualifications in the 21st century. *European Educational Research Journal*, 22(1), 20–38.

Wolf, A. (1993). *Assessment Issues and Problems in a Criterion-Based System*. London: Further Education Unit.

Wolf, A. (1994). Measuring competence: The experience of the United Kingdom. *European Journal of Vocational Training*, 1, 29–35.

Wolf, A. (1995). *Competence-Based Assessment*. Maidenhead: 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: Principles, Policy & Practice*, 5(3), 413–445.

Wolf, A. (2011). *Review of Vocational Education – The Wolf Report*. London: Department for Education.

Wolf, A., Burgess, R., Stott, H., & Veasey, J. (1994). *GNVQ Assessment Review Project. Final Report*. London: Employment Department.

Published by:

ofqual

2nd floor, 1 Friargate
Station Square
Coventry
CV1 2GN

0300 303 3344

public.enquiries@ofqual.gov.uk

www.gov.uk/ofqual

November 2024

Ofqual/24/7177