RESEARCH AND ANALYSIS

Understanding Qualification Design

Insights from the 2020 to 2024 CASLO qualification research programme

ofqual

Authors

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

Acknowledgements

We are extremely grateful to all of our colleagues within Ofqual who have supported the CASLO research programme in so many ways. For insights of particular relevance to the present report, we would also like to thank Alison Wood, Chris Winch, and Stan Lester.

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 many vocational and technical qualifications in England.
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

The CASLO approach

At the outset of our research programme, circa 2019, we noted that large numbers of Vocational and Technical Qualifications (VTQs) shared 3 core design characteristics:

1. unit content is specified via learning outcomes
2. the unit standard is specified via assessment criteria for each learning outcome
3. to pass each unit, a learner must acquire all of the specified learning outcomes, which we refer to as the mastery requirement

We proposed that, when a qualification incorporates these 3 core characteristics – tightly specified outcomes, tightly specified criteria, and a stringently applied mastery principle – we will describe this as having adopted the CASLO approach.[1] In effect, this approach constitutes a high-level template for designing qualifications.

As recently as 2020, we had no generic label for qualifications of this sort. For the purpose of the present research programme (and subsequent scrutiny) we decided to call them CASLO qualifications, because they are all designed to Confirm the Acquisition of Specified Learning Outcomes.

Our research into the origins and evolution of the approach revealed the strategic importance of labelling it, and thereby demarcating it from other approaches. Despite its prevalence in England, it is just one way of operationalising an outcome-based approach to qualification design, and just one way of operationalising a mastery-based approach.[2]

The present report is the capstone of our 2020 to 2024 research programme, outcomes from which are described within 9 reports (which are listed in the present report prior to its contents page) linked to 4 strands:

1. descriptive strand – which set out to explain what we mean by the CASLO approach to qualification design and, therefore, what we mean by a CASLO qualification

[1] Note that certain high profile VTQs of that period did not adopt the CASLO approach. For instance, the T Level Technical Qualifications and Occupational Specialism Assessments were (and still are) marked numerically rather than being judged directly against assessment criteria, and they did not (and do not) apply the mastery requirement in the stringent manner required by the CASLO approach. Their design is more akin to the ‘classical’ approach that is adopted for GCSEs and A levels.

[2] Outcome-based approaches preceded mastery-based ones, historically, although they often incorporate a mastery-based approach. As such, we sometimes refer simply to ‘outcome-based’ as the superordinate category of most importance to our analysis. We distinguish between outcomes and mastery where necessary.

2. functional strand – which set out to describe how CASLO qualifications work (in contrast to the more widely recognised family of ‘classical’ qualifications, which includes GCSEs and A levels)
3. historical strand – which set out to understand the origins and evolution of the CASLO approach within the landscape of VTQs in England
4. critical strand – which set out to consider criticisms that have been levelled at the CASLO approach

The present report does not attempt to summarise everything that we have learnt over the course of the past few years. Instead, it attempts to draw together some of the most fundamental insights from the programme – related, in particular, to criticisms of the approach and to how potential problems might be mitigated – with the use of some theoretical scaffolding. Its 4 main sections are headed:

1. ‘policy journey’ – a brief account of how views of the CASLO approach have evolved over time, particularly from a policy perspective, which ultimately explains the rationale for our research programme
2. ‘fledgeling theoretical framework’ – an attempt to sketch an outline for an overarching integrated theory of educational certification, within which to situate a framework for understanding qualification design
3. ‘anticipatory qualification design’ – suggestions for steps that can be taken to improve the likelihood of effective qualification reform, by anticipating likely requirements and threats more thoroughly
4. ‘moving forward’ – our concluding section, which explains how we might begin to use insights from our research programme to improve qualification policies and practices in the future

Note that the present report is titled ‘Understanding Qualification Design’ as the insights that we have drawn from the programme extend beyond the CASLO approach.

Policy journey

England embraced outcome-based qualification design during the 1970s within a variety of Technical and Vocational Education and Training (TVET) settings. It became a matter of national policy during the 1980s, as the CASLO approach was embedded at the heart of the National Vocational Qualification (NVQ) system. During the 1990s and into the 2000s, it continued to be supported by policy makers. Yet, in the wake of policy reviews from the early 2010s, it began to fall out of favour. The following 2 subsections summarise this historical journey.

Historical journey

The origins and evolution of the CASLO approach are detailed in our report on the historical strand, report 4, which should be consulted for further details. The following bullet points provide a highly condensed summary of this story:

- during the 1950s, industrial training and certification were not in a good state:
    - off-the-job college qualifications were respected but were far removed from occupational competence, and they were plagued by drop out and failure
    - on-the-job apprenticeship and training was patchy, of variable quality, and structurally straightjacketed (by the unions in particular)
- during the 1960s, the state started driving reform of industrial training by, for example, establishing Industrial Training Boards to roll out top-down initiatives
- during the 1970s, the state started driving reform of TVET qualifications by, for example, establishing the Technician Education Council (TEC) and the Business Education Council (BEC) to co-ordinate the provision of technician-level awards, taking over from a plethora of existing organisations in an attempt to rationalise the system
- during the 1980s, the state established the National Council for Vocational Qualifications (NCVQ) to further rationalise the TVET qualification landscape
- the NCVQ introduced the NVQ framework, to which it was anticipated that all TVET qualifications would be accredited
- the NVQ framework put the CASLO approach on the map by specifying its core characteristics as accreditation criteria for all NVQs (although its outcome-based and mastery-based approach was clearly prefigured in BEC and TEC awards)
- there was a lot of criticism of NVQs: scholars criticised its model, while many criticised its rollout

- fairly early on, it was confirmed that certain TVET qualifications would not have to be accredited to the NVQ framework (as NVQs)
- yet, many of these qualifications that remained beyond the framework were still designed (or redesigned) in accordance with the CASLO approach, including General National Vocational Qualifications (GNVQs), Business and Technician Education Council (BTEC) awards, Open College Network (OCN) awards, and so on
- while scholars continued to criticise the model at the heart of NVQs and GNVQs – which included criticism of the CASLO approach itself – policy reviews of the period largely supported the underlying model, despite acknowledging teething problems
- during the early 2000s, the state embarked on a new qualification rationalisation exercise, this time introducing the Qualifications and Credit Framework (QCF), to which it anticipated that all TVET qualifications would be accredited
- it was soon confirmed that certain TVET qualifications would not have to be accredited to the QCF, although the vast majority were accredited
- this was significant because (as for the NVQ framework) accreditation criteria for the QCF required the CASLO approach
- just as the vast majority of regulated VTQs were being accredited to the QCF, during the early-2010s, the new coalition government initiated a series of high-profile policy reviews
- unlike earlier reviews, they tended to be less positively disposed to the CASLO approach, and fall-out from these reviews and from Ofqual's own evaluations led to the QCF being withdrawn and to certain qualifications no longer being permitted to adopt the approach (see following section)

In short, the CASLO approach:

- was formally introduced by a government agency (although awarding bodies had been experimenting with similar approaches since the 1970s and 1980s) and continued to be promoted by successive government agencies
- achieved almost hegemonic status as a VTQ design template (partly owing to government pulling funding levers, but also due to the support of awarding organisations, and to a certain amount of support on the ground)
- fell out of favour with policy makers during the 2010s

Present day

Shortly after the 2011 Wolf report had been published, the Department for Education (DfE) announced that it would tackle CASLO-related criticisms by taking steps to increase the 'rigour' of qualifications that were eligible for being counted in school and college performance table calculations. This led to DfE requirements for external assessment and synopticity. Although this did not specifically proscribe the CASLO approach, the approach is challenging to operationalise via terminal external assessments. Consequently, many of the qualifications that were redesigned for inclusion in performance table calculations (from 2016 onwards) included classical units alongside CASLO ones, rendering them hybrid qualifications.[3]

Following a critical review of the QCF in 2014, Ofqual withdrew this regulatory framework, continuing to regulate all former QCF qualifications under its General Conditions of Recognition. From that point on, Ofqual no longer required any regulated qualification to adopt the CASLO approach.

Pursuing a qualification 'strengthening' agenda, during the late-2010s, Ofqual began to require numerical marking for centre-based (as well as external) assessments for certain performance table qualifications, which included Level 3 (T Level) Technical Qualifications, and Level 1/2 Tech Awards. In effect, this proscribed the CASLO approach for those units and qualifications.

In some ways, the Institute for Apprenticeships and Technical Education (IfATE) has also distanced itself from the CASLO approach. This was in response to concerns raised in the Richard report related to the "time consuming" nature of assessment via NVQs and related qualifications, linked to the idea of having to "tick off a very long list of competencies" (quotations from Richard, 2012, pages 67 and 87 respectively). Having said that, the new apprenticeship standards are still defined in terms of outcomes – which are now described as elements of knowledge, skill, and behaviour – and IfATE still respects the principle of mastery in the sense that End-Point Assessments (EPAs) are intended to certify full occupational competence, which corresponds to having achieved all specified elements. Although, strictly speaking, the new apprenticeship model is not based on the CASLO approach – for instance, it does not specify assessment criteria for each learning outcome – the EPA model is still both outcome-based and mastery-based.

[3] A classical unit would incorporate numerical marking (rather than criterion-linked judgements) and would apply a compensatory aggregation principle (rather than a mastery one). See report 2.

Down but not out

When we discussed the recent history of the CASLO approach in chapter 6 of report 4, we concluded that it was currently 'down but not out'. It had fallen out of favour with VTQ policy makers in England during the 2010s, and it was no longer promoted as a matter of VTQ policy in the way that it had been. Yet, Ofqual continued to regulate many high-certificating CASLO qualifications.

One of the curious features of outcome-based approaches to curriculum and qualification design is that they have a tendency to be loved by some while being hated by others, and they have a tendency to fall into and out of favour. For instance, although the CASLO approach fell out of favour with policy makers in England, policy makers elsewhere appear to have become increasingly enamoured with outcome-based approaches. In a recent critique of outcome-based qualifications, Winch reflected on this international trend with an air of perplexity:

> The language of learning outcomes and associated terminology such as 'competence', 'output' and 'assessment criterion' is becoming a global phenomenon [...]. More than three decades after its adoption, its fortunes still seem to be increasing, despite a dubious record of achievement and some hard-hitting but relatively ineffectual criticism. It is difficult to find a clear answer as to why. As the basis for policy, the adoption of learning outcomes approaches has ranged from relatively harmless to disastrous and take-up has generally occurred without careful and thorough evaluation of previous experiences by early adopters.

(Winch, 2023, page 21, footnote references removed)

Scholars of education have frequently criticised outcome-based approaches (see report 5) and the academic literature has tended to focus on examples of outcome-based qualifications failing, sometimes insisting that they are inherently doomed to fail. This literature has revolved around experiences from the UK, as an early and highly influential adopter, although experiences from other countries have also helped to shape this narrative, including South Africa and Australia (for example, Allais, 2014; Wheelahan, 2016).

Conversely, in recent years, the European Centre for the Development of Vocational Training (Cedefop) has repeatedly reported on the international 'shift to learning outcomes' since the late 2000s. For instance, earlier this year, a Cedefop policy brief opened with the following observation:

> The idea of learning outcomes increasingly seems to dominate education policy at European and international levels; many countries around the world seem to be shifting towards a greater role for learning outcomes in their education and qualification systems (Cedefop, 2008; 2009). Currently all countries from Europe

> and many beyond are actively using learning outcomes or competence statements when defining, reviewing and refining the content and profile of their education, training and skills provisions and strategic practices (Cedefop, 2016; 2017; 2022; UNESCO, et al., 2023).
>
> (Cedefop, 2024a, page 3, footnote reference removed)

In this report from Cedefop, the shift to learning outcomes – which has involved focusing both curriculum and assessment squarely upon the acquisition of specified learning outcomes – was described as one of the most significant trends to have influenced vocational education and training over the past 2 decades. Moreover, this trend is not confined to technical and vocational qualifications, having occurred in the higher education sector too (Biggs, Tang, & Kennedy, 2022).

So, while policy makers in England seemed, during the 2010s, to become more sympathetic to concerns expressed within the academic literature, policy makers elsewhere seem to have travelled in the opposite direction (albeit often characterising learning outcomes in quite different ways across countries).

As suggested by Winch, this disjunction between academic antipathy and international policy enthusiasm is curious, to say the least, and worthy of analysis in its own right. We suspect that both sides may have their blind spots. We suspect that policy makers often pay insufficient attention to how outcome-based qualification design can fail, while scholars often pay insufficient attention to how outcome-based qualification design can succeed. Indeed, the Cedefop briefing offered some support for this hypothesis, noting a tendency for policy makers to underestimate the extent of real-world challenges, and emphasising how the “interests and decisions taken by different stakeholders directly influence the way, and the extent to which” outcome-based qualifications succeed, especially for learners (Cedefop, 2024a, page 3).

Our research programme

Although it is fair to say that the CASLO approach fell out of favour with policy makers during the 2010s, this does not mean that policy makers had become entirely antagonistic towards it by the end of the decade. Ofqual, for instance, had proscribed it in certain contexts (particularly for qualifications that were counted in school performance tables) but not in others. Consequently, following a tightening of regulations during the latter part of the 2010s, 2019 seemed (to Ofqual) to be a good point at which to take stock of the approach.

Recognising that the VTQ landscape was poorly documented, poorly researched, and poorly theorised, we embarked upon the present research programme to provide a solid foundation of knowledge and understanding that could help us to take stock. We reasoned that the better we (the sector) understand the CASLO approach:

- the better our policy making will be
- the better our regulatory practices will be, and
- the better our qualification design, development, and delivery will be

The following sections of this report are based on the assumption that the CASLO approach, and other outcome-based approaches, can work, but that their successful operation cannot be taken for granted. Indeed, in some ways they can be harder to operate successfully than classically designed qualifications. So, we need to reflect deeply on how best to ensure their effectiveness.

Fledgeling theoretical framework

To understand better the pros and cons of outcome-based qualification design, we recognised the potential value of scrutinising the origins and evolution of the CASLO approach through the lens of a broader theoretical framework. Unfortunately, one of the key issues that motivated our research programme was the observation that VTQs in England are poorly theorised. More radically, our report on the historical strand suggested that we may simply lack an overarching integrated theory for qualification design – a theory of qualifications or, more generally, a theory of educational certification – so this is not simply a problem for VTQs. This proposition is worth spelling out in more detail.

Although qualification design rightly draws upon insights from educational measurement theory, curriculum theory, and to a lesser extent pedagogical theory – each of which might be described as an established field or discipline in its own right – what we seem to lack is a distinct body of scholarship that formally integrates insights from these multiple perspectives, and from others too. This, we propose, is exactly what educational certification theory ought to do.

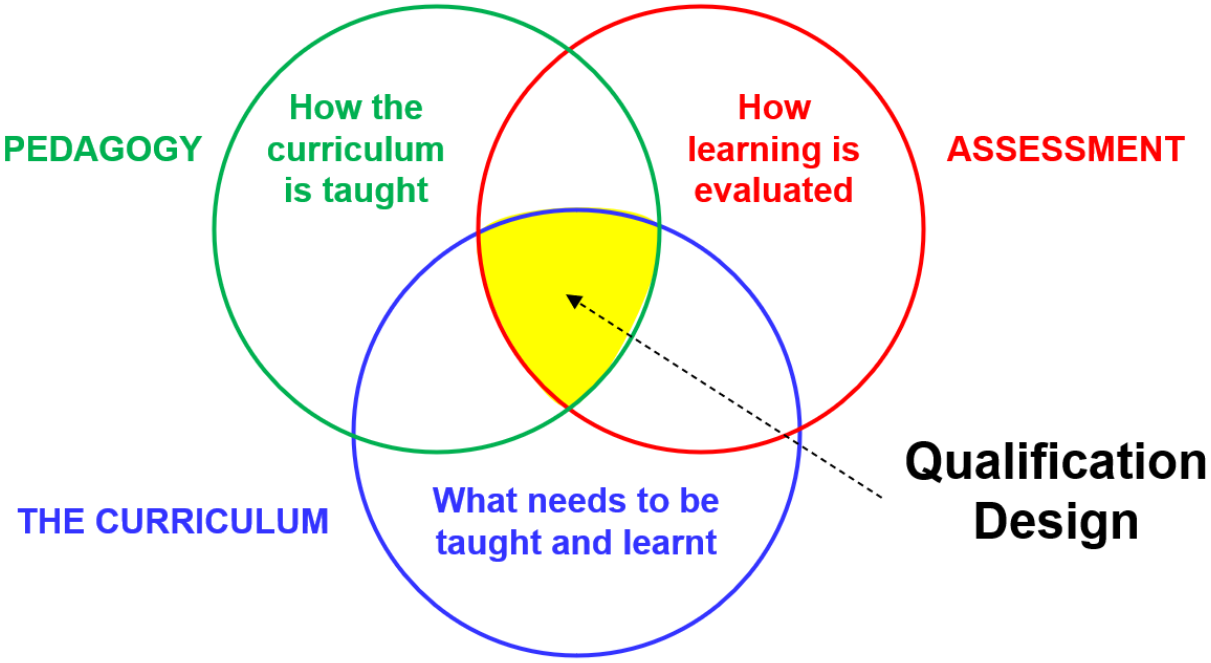

Figure 1. Qualification design at the intersection

This analysis is premised on the assumption that qualification design operates at the interface between curriculum, pedagogy, and assessment – as illustrated in Figure 1

– also informed by insights from subdisciplines as disparate as industrial and behavioural economics, industrial and organisational psychology, industrial and political sociology, moral and political philosophy, policy science, and so on. From this perspective, the distinctive role of educational certification theory should be to explain how these multidisciplinary insights ought to be marshalled and co-ordinated in the effective design of educational certification systems and procedures.

The following subsections might be understood as early contributions to a nascent discipline of educational certification, informed by insights from the present research programme (particularly the historical strand) as well as from earlier work on qualification purposes (for example, Newton, 2007; 2017a; 2023a; 2023b).

Framework outline

We propose that designing a system or procedure for certifying educational attainment (for instance, a school-leaving qualification) comprises 3 critical stages:

1. We identify the full range of purposes (or goals) that explain why the particular qualification is deemed to be necessary for a particular cohort of learners in a particular set of circumstances (these circumstances elucidate the anticipated contexts of teaching, learning, and assessment). Then we prioritise between those purposes. The resulting profile of prioritised purposes will hold direct implications for what will need to be taught, learnt, and certified (the target proficiency) and for how it will need to be assessed (the assessment procedure).
   - This should be informed by policy statements, curriculum theory, pedagogical theory, measurement theory, and by other theoretical perspectives too.
2. We construct a model of the learning that will need to be certified in order to maximise the likelihood of satisfying the profile of prioritised purposes. That is, we construct a ‘proficiency model’ to characterise our ‘target proficiency’ in order to explain exactly what it is that our qualification will need to certify. This will necessitate a decision concerning the overarching approach to qualification design, for instance, whether to adopt the CASLO approach or a different one. It will also involve working out how best to communicate the proficiency model to teachers, trainers, learners, assessors, certificate users, and other stakeholders, whom we will collectively refer to as the wider ‘certification community’ for our qualification.
   - This should be informed by experts in the relevant domain of learning alongside experts in qualification design, and with consideration of pedagogical and assessment implications.
3. We design an assessment procedure for our qualification in order to govern the development of assessment materials, the delivery of assessment events, the

processing of assessment evidence, and the award of qualification certificates. The design of the procedure will need to be as consistent as possible with the proficiency model and with the profile of prioritised purposes, while also bearing in mind a variety of pragmatic, economic, legal, moral, and political constraints on qualification design.

- This should be informed by assessment experts alongside experts in qualification design, with support from experts in the relevant domain of learning, and with consideration of curriculum and pedagogical implications.

This approach embodies the guiding principle that qualification design operates at the interface between curriculum, pedagogy, and assessment. For exactly this reason, it needs to be driven (not just by multiple purposes but) by multiple perspectives on qualification purposes – curriculum and assessment perspectives, in particular, but others too, including pedagogical ones. Likewise, for exactly this reason, the key to effective qualification design is to begin by deliberating explicitly upon these purposes and prioritising between them. Because different stakeholders instinctively prioritise different purposes, this recommends an approach that is based upon consensus building and therefore also compromise.

Our idealised framework for qualification design is illustrated in Figure 2. The black arrows emphasise that this is a sequential process: only once a profile of prioritised purposes has been identified (stage 1) can we move on to the more technical tasks of specifying what the proposed new qualification will need to certify (stage 2) and how this will need to be assessed (stage 3).

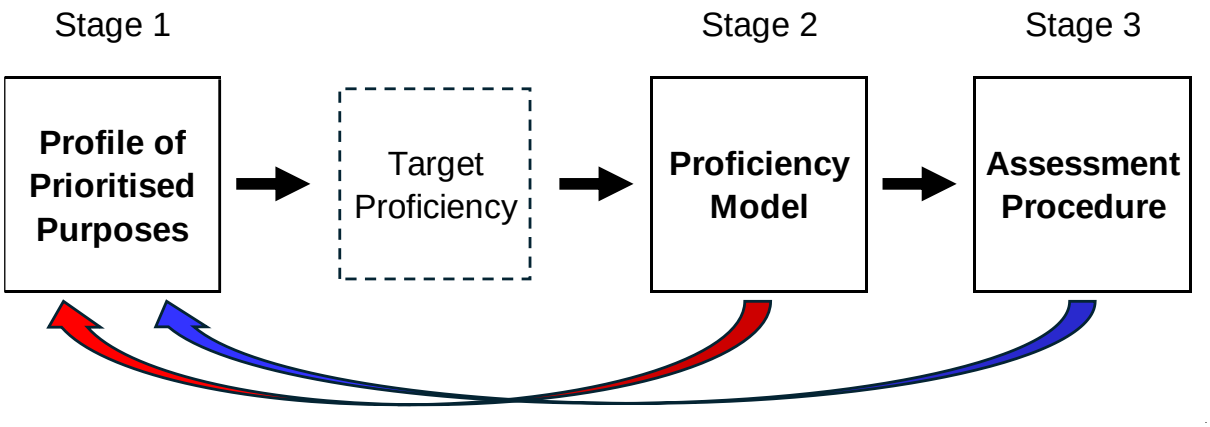

Figure 2. An idealised framework for qualification design

The blue and red arrows emphasise that, in practice, this will also need to be an iterative process. Stages 2 and 3 are concerned with working through the

implications of the purpose profile that was identified during stage 1. The inclusion of a wide range of stakeholders at stage 1 helps to ensure that many of those implications will already have been anticipated. Inevitably, though, these implications will only be fully worked though during the later stages. Unfortunately, unanticipated implications can sometimes raise doubts concerning the legitimacy or utility of the previously agreed purpose profile, for instance, where implications from certain of the prioritised purposes prove to be in conflict, or where operationalisation proves not to be viable. In circumstances like this, the purpose profile may need to be revisited and reprioritised.

Although stage 3 is often (rightly) acknowledged to be highly technical, it needs to be appreciated that stage 2 is technical too. It is true that decisions concerning what needs to be taught, learnt and certified are ultimately grounded in value judgements. Yet, the principal debates concerning value judgements will occur during stage 1. As such, the profile of prioritised purposes that results from stage 1 deliberations will hold direct, logical implications for the nature of the target proficiency, and this renders explication of the proficiency model a far more technical exercise. Stage 1 deliberations will not entirely eliminate the need to make value judgements during stage 2, but they will set critical parameters for making those judgements.[4] Stage 2 is also technical in the sense that the proficiency model will need to represent the target proficiency as authentically and comprehensively as possible without becoming so complex and detailed that it fails to fulfil its principal function (which is to communicate the target proficiency to members of the certification community as effectively as possible).

A model

We should say a few words on why this report refers to the construction of a proficiency ‘model’ and to the idea of ‘modelling’ more generally. After all, in previous reports, we have described the same concept as a proficiency ‘specification’ rather than a ‘model’ (Newton, 2017b).[5] Furthermore, if all that we are really referring to is the process of ‘defining’ or ‘describing’ what our qualification will need to certify, then perhaps we should not unduly mystify it with terms that seem more technical.

---

[4] Stage 2 is likely to be more technical and to require fewer value judgements for technical qualifications (which relate to occupational roles) than for general qualifications (which relate to subject areas). Yet, these lower-level value judgements will still need to remain within the parameters established by the higher-level value judgements that were agreed during stage 1.

[5] Note that the term ‘specification’ has prior significance in the context of regulated qualifications in England, as Ofqual regulations require awarding organisations to publish a ‘specification’ for each qualification that it provides (to indicate critical information concerning the design, development, and delivery of the qualification). As such, the term ‘specification’ has superseded the term ‘syllabus’.

Conversely, the concept of modelling – which refers to the activity of constructing an explicit model – is actually quite meaningful and useful in this context.

First, it is important that we distinguish the thing that our qualification needs to certify (the target proficiency) from our representation of that thing (our proficiency model). The target proficiency refers to the proficiency as manifest in the real world (for example, the ability to assemble a bicycle from component parts) while the proficiency model is a representation of that proficiency, which is provided for anyone with a stake in the certification process (students, teachers, assessors, certificate users, and so on – our certification community for bicycle assembly). The purpose of the model is therefore to act as an explicit point of reference for the certification community – a basis for sustaining a shared understanding of whatever the qualification needs to certify. The idea of a model is useful for emphasising that it is a tool that is developed to serve a clear purpose.

Second, as a representational tool, the proficiency model will foreground those aspects of the target proficiency that its developers deem to be relevant and important. This emphasises that fact that, for a proficiency model to have currency, it will need to have represented its target proficiency in a manner that makes sense to, and is generally acceptable to, members of the wider certification community. In other words, we should think of the proficiency model as a social construct, which is intended to represent a concept on behalf of a certification community.

Third, the idea that a proficiency model is intended to represent an aspect of the (socially constructed) world helps to underscore the fact that it might be either a good representation or a bad one. In fact, as we will soon see, some of the criticisms that have been levelled against outcome-based qualification design can be understood in terms of sub-optimal proficiency modelling.

Modelling

A proficiency model might be constructed through various means. For instance, we might:

- label the proficiency (for example, ‘the ability to assemble bicycles’)
- specify what the proficiency includes and excludes (for example, the basic skills of assembling such as torquing, the sorts of bicycles that an assembler would or would not be expected to assemble, whether they would be expected to be able to fabricate bespoke parts, whether they would be expected to manage team members, and so on)
- distinguish between core and peripheral abilities (for example, bicycle assembly versus working with customers versus understanding the bicycle market)

- exemplify different levels of proficiency (for example, related to speed of assembly or to autonomy of working)

This final bullet point is important to emphasise. A proficiency model is not simply a description of an area of learning, or a list of topics to be studied, it is a model of the anticipated learning itself. Critically, this includes information concerning what it means to have achieved a satisfactory level of learning within the domain, that is, how broad and deep the learning needs to be for a learner to be judged to have passed the qualification. In relation to the CASLO approach, all 3 of its core characteristics – outcomes, criteria, and the mastery requirement – refer to dimensions of the proficiency model.

This modelling is likely to involve written descriptions. But it might well incorporate a variety of representational formats, especially when the proficiency is exemplified via performance evidence (for example, using audio or video recordings). Exemplification of this sort is particularly important for filling in the gaps that remain when outcomes and criteria are specified primarily in terms of written statements.

Alignment

In our idealised framework for qualification design, the proficiency model has a critical role to play within a certification system. In fact, it has 2 distinct roles, which both relate to ensuring consistent accuracy. First, the model helps to ensure that a qualification certifies exactly what it needs to certify by establishing this explicitly from the outset. So, the first step in ensuring consistent accuracy is to ensure that the proficiency model represents the target proficiency with sufficient accuracy.[6] Second, the model helps to ensure that all members of a certification community interpret the target proficiency in the same way. The model becomes the principal point of reference to which curriculum, pedagogy, and assessment plans should all be aligned, as well as being the principal point of reference for students and certificate users too. So, the second step in ensuring consistent accuracy is to ensure that all members of the certification community are singing from the same hymn sheet in the sense that they are all basing their understanding of what the qualification certifies on exactly the same proficiency model.

Alignment is the critical concept here: alignment between the proficiency model and the target proficiency, and alignment between all members of the certification community in terms of their understanding of what the qualification is supposed to

[6] Remember that the profile of prioritised purposes sets critical parameters for the modelling, and if those parameters are not respected – and the proficiency model is constructed in a manner that fails to accord with agreements reached during stage 1 – then it will not represent the target proficiency accurately.

certify. Furthermore, because the proficiency model derives directly from the prior analysis of qualification purposes, it is also the principal mechanism for securing continuing alignment to those purposes throughout the lifecycle of the qualification. So, it serves a longitudinal perspective (alignment over time) as well as a horizontal one (alignment across community members).

Misalignment

The idea of constructing an explicit model of the proficiency that a qualification needs to certify is critical to understanding both the origins of the CASLO approach in England and threats to its effective implementation.

As explained fully in report 4, outcome-based approaches to curriculum and qualification design began to gain traction in England during the 1970s as a direct response to perceived inadequacies of extant approaches. One of the problems with these approaches was that they were not based on an explicit model of what needed to be learned and assessed, that is, they lacked an explicit proficiency model.

These ‘classical’ approaches to qualification design revolved around minimally specified exam syllabuses. As late as the 1970s, syllabuses of this sort comprised little more than a list of topics that students were likely to be examined on, with past exam papers providing further insight into the competencies that examiners sought evidence of (Schools Council, 1973). This was true for extant VTQs as well as for general qualifications, like O levels and A levels. Consequently, the target proficiency for a classically designed qualification of the 1970s would never have been articulated explicitly and, as such, both teachers and examiners were liable to interpret those proficiencies differently.

This lack of an explicit proficiency model came to be associated with alignment-related problems. The classic critique of classically designed qualifications was that they tended to over-assess ‘lower-level’ competencies (such as recall and comprehension) and under-assess ‘higher-level’ ones (such as evaluation and application). To some extent, this was a consequence of a traditional reliance on the written exam format, which often defaults to assessing ‘lower-level’ competencies, as they are easy to assess in this format. However, the lack of an explicit proficiency model helped to conceal this. Where written exams under-assessed ‘higher-level’ competencies, this had an inevitable negative backwash impact on what teachers taught and learners learnt.

The shift to outcome-based approaches during the 1970s was a direct response to concerns of this sort. In the TVET context, these concerns were bolstered by a desire to ensure that VTQs certified what the workforce actually needed them to certify (rather than being skewed to the idiosyncratic preferences of teachers and trainers, for example). This new emphasis on ‘occupational competence’ (as

opposed to just ‘book knowledge’) raised the question of how it ought to be defined and articulated. This put the explication of learning outcomes firmly on the agenda for VTQ designers of the 1970s and 1980s.[7]

The shift towards making learning outcomes explicit was intended to facilitate alignment, as everyone with a stake in the certification process would now be singing from exactly the same hymn sheet. As such, it would to help ensure that assessors assessed:

- all that they ought to assess (mitigating the risk of construct underrepresentation), and
- nothing that they ought not to assess (mitigating the risk of construct-irrelevant variance) [8]

And by essentially the same mechanism, it would help to ensure that teachers taught what needed to be taught and learners learnt what needed to be learned.

The principle of outcome-mediated alignment between curriculum, pedagogy, and assessment has been recognised for a long time, having been advocated by Tyler, in particular, nearly a century ago (Tyler 1931; 1932a; 1932b). Indeed, it remains widely recognised and respected to the present day, for example, in the theory and practice of constructive alignment, which has been developed across 5 editions of the now classic text ‘Teaching for Quality Learning at University’ (Biggs, et al, 2022). The concept of alignment has risen in prominence internationally in recent decades, as a direct consequence of concerns over misalignment and its negative backwash impact on teaching and learning (for example, Looney, 2011; Koretz, 2017; Mulder, 2017a; Care, Kim, Vista, & Anderson, 2018). The concept is common to various accounts of how to improve the effectiveness of education systems. For example, Oates has emphasised that entire systems need to be characterised by alignment and mutual reinforcement: curriculum content, textbooks, teaching content, pedagogy, assessment, accountability, inspection, funding, and so on (Oates, 2010; 2013).

CASLO modelling

Even among philosophers, there remains disagreement over what terms like ‘knowledge’ and ‘belief’ actually refer to. Yet, what seems less debatable is that we

[7] Making learning outcomes explicit was also intended to help make the process of becoming qualified more efficient, by facilitating innovations such as the Recognition of Prior Learning (see report 4 for more detail).

[8] Within the literature on educational measurement, construct underrepresentation and construct-irrelevant variance are considered to be the 2 principal threats to validity (see Newton, 2020).

attribute ‘mental’ constructs like these on the basis of what people do: how they act, what they say, and – during formal assessment events – how they perform in response to what assessment tasks invite them to do.

It therefore seems natural that we should model learning – which we describe at a high level of abstraction in terms of having acquired knowledge, skill, and so on – in terms of what people can do (now) that they could not do (previously). This idea lies at the heart of the outcome-based approach to qualification design, and at the heart of taxonomies that have been developed to support them, including Bloom’s taxonomy (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956; Anderson & Krathwohl, et al, 2001) and the SOLO taxonomy (Biggs, et al, 2022).

Table 1 presents an extract from a unit specification from a qualification in plastering, which is typical of how learning outcomes tend nowadays to be modelled within CASLO qualifications.[9] For any particular CASLO qualification, its core proficiency model will include the set of all of its learning outcomes and assessment criteria across all of its units. The mastery aggregation principle is also a part of this core proficiency model, because it specifies that being proficient involves having achieved all of the specified learning outcomes to the standard associated with their assessment criteria.

Learning Outcomes	**Assessment Criteria**
The learner will:	**The learner can:**
2. Know how to comply with relevant legislation and official guidance when applying finishing plaster to prepared surfaces.	2.1. Describe their responsibilities regarding potential accidents, health hazards and the environment whilst working: in the workplace, below ground level, in confined spaces, at height, with tools and equipment, with materials and substances, with movement/storage of materials and by manual handling and mechanical lifting.
	2.2. Describe the organisational security procedures for tools, equipment and personal belongings in relation to site, workplace, company and operative.
	2.3. Explain what the accident reporting procedures are and who is responsible for making reports.

Table 1. Example of how CASLO units tend to be specified

[9] It relates to the NOCN_Cskills Awards Level 2 NVQ Diploma in Plastering (Construction) – Solid (603/2368/1), which we described in more detail in the second report from our research programme.

Sometimes CASLO unit specifications model learning outcomes directly in terms of what learners are now able to do, such as ‘maintain safe and healthy working practices when applying finishing plaster to background prepared surfaces’ (which happens to be the third learning outcome from the same plastering unit). Other times, they model learning outcomes at a higher level of abstraction – in terms of ‘knowing that’ or ‘knowing how’ or suchlike – leaving it to the assessment criteria to explain what this means in terms of what learners are actually able to do (as in Table 1). The fact that outcome-based qualifications model proficiency in terms of what learners are now able to do (as a consequence of learning) explains why they make such heavy use of verbs – doing words – like describe, explain, analyse, and evaluate.

Modelling risks

Although there are serious risks associated with designing a qualification in the absence of an explicit proficiency model, there are also risks that stem directly from the process of modelling a target proficiency. These are linked to the fact that modelling involves constructing a simplified representation of the world. First, the fact that the model is supposed to provide a simplified account means that it necessarily glosses over detail in order to provide a less rich but more useful description. In other words, models are never perfectly accurate, by design. Unfortunately, this runs the risk of providing an oversimplified account, where the glossed over detail turns out to be important for effective practice. Second, the fact that the model represents the world means that suboptimal modelling runs the risk of straightforwardly misrepresenting it, either by omission or by commission. As such, the risks of construct underrepresentation and construct irrelevance are just as prevalent at the proficiency modelling stage as they are at any other stage of the qualification lifecycle. In fact, they are particularly important to avoid at the modelling stage, if the model is to become the principal point of reference for the remaining stages.

The risk of misrepresentation is particularly acute in the context of educational certification. Although awarding organisations routinely certify the acquisition of knowledge and skill on a weekly basis, even philosophers disagree over what these terms actually refer to. So, there is nothing trivial about the activity of proficiency modelling. As we will illustrate below, some of the problems that have been linked to the CASLO approach in the literature (as described in report 5) can be understood in terms of modelling risks.

Deconstruction

While classical qualification syllabuses deconstruct learning domains into subject content areas, CASLO qualification specifications take this deconstruction process a step further by explicating intended learning outcomes and assessment criteria. Deconstruction is fundamental to effective qualification design: it enables the certification community to reach some kind of consensus over what (exactly) the qualification in question is supposed to be certifying.

The risk associated with deconstructing a target proficiency (via its proficiency model) is pragmatic. In report 4 we referred to this as the ‘grain size’ challenge:

- if there is too little deconstruction, then the proficiency model will lack utility by failing to explicate the target proficiency in sufficient detail, and thereby failing to provide sufficient scaffolding for the certification community
- if there is too much deconstruction, then the proficiency model will lack utility by providing the certification community with too much scaffolding – more than they can reasonably utilise – and it may also risk misleading the certification community by attempting to fragment beyond the point at which the target proficiency can no longer be meaningfully deconstructed

Unfortunately, the desire for transparency has a tendency to tempt designers of outcome-based qualifications into providing too much deconstruction, which Wolf described as the “never-ending spiral of specification” (Wolf, 1995, page 55).

Disintegration

Deconstruction also risks giving the impression that a target proficiency can be understood as little more than the sum of its individual parts, as though learning were akin to stamp collecting. This risk has been associated with classically designed qualifications, especially when assessed using short-answer questions, which can give the impression that learning amounts to little more than the accumulation of discrete facts.

Bereiter & Scardamalia criticised traditionally formulated educational objectives on the same basis, arguing that they construct a misleading proficiency model, as though the brain operated like a “mental filing cabinet” (Bereiter & Scardamalia, 2005, page 5). What a model like this clearly fails to represent is the organization and co-ordination of knowledge and skill (see also Messick, 1984; Sadler, 1987; Wood & Power, 1987; Mislevy, 1993; van der Vleuten, 1996; Eraut, 2004). The risk, here, relates to the disintegration of the target proficiency that occurs through the deconstruction process. A disintegrated model provides important information on the elements of knowledge and skill, but not on how they interconnect – it lacks information on structure, organisation, integration, and coordination (Newton, 2020).

The force of this critique may depend to some extent on how the learning outcomes that comprise an outcome-based unit are articulated, including their grain size and complexity. For instance, if only a few outcomes are articulated, and they are framed in terms of complex constructs linked to high-level abilities – like supervising or evaluating – then the critique may have less force. Conversely, where a unit contains lots of outcomes, all framed in terms of simple constructs linked to low-level abilities – like assembling or recalling – then it will surely have more force. It seems fair to conclude, however, that deconstruction will always incur a certain amount of risk related to the omission of information concerning interconnectedness.

Simplification

Bereiter & Scardamalia were particularly critical of Bloom's taxonomy, and their critique highlights risks associated with simplifying the target proficiency (via a proficiency model). The point of Bloom's taxonomy was to provide a tool for distinguishing between less complex and more complex forms of engagement with subject content. For example, should learners be expected to be able to evaluate content x, y, or z, or should they merely be expected to be able to comprehend it? Yet, Bereiter & Scardamalia argued that the taxonomy is too simplistic, and the manner of its presentation risks misleading curriculum planners:

> Few would dispute that a good educational program will engage students in plenty of comprehending, applying, analyzing, synthesizing, and evaluating. But these do not constitute a curricular sequence. No sane educator would propose starting with knowledge in grade 1, moving to comprehension in grade 2, application in grade 3, and so on. Rather, the levels of the *Taxonomy* refer to processes that need to go on in concert at all levels, supposedly leading to the attainment of worthy objectives.
>
> (Bereiter & Scardamalia, 2005, pages 13 to 14)

As such, they questioned the very idea of domain-independent, or content-independent, levels of understanding, which seems (at least) to lie at the heart of Bloom's taxonomy. These researchers are far from alone in questioning the adequacy of this taxonomy (see Pollitt, Ahmed, & Crisp, 2007, for instance). Since Bloom's taxonomy has heavily influenced the design and development of outcome-based qualifications, this is an important critique (and we will return to it later).

Proficiency modelling

In the preceding subsections, we have sketched an outline for an overarching integrated theory of educational certification, within which to situate a framework for

understanding qualification design. We suggest that this is useful for helping to identify:

- what sets outcome-based qualification design apart from other approaches (for example, the proficiency model for an outcome-based qualification will be explicated as a complete set of learning outcomes that becomes the principal point of reference for planning curriculum, pedagogy, and assessment)
- critical high-level threats to outcome-based qualification design, which will need to be addressed (for example, outcome-based proficiency models have predictable weaknesses, given their particular approach to modelling, which can potentially be mitigated)

The idea of constructing an explicit proficiency model – which constitutes the second stage of our idealised framework for qualification design – makes a lot of sense in relation to outcome-based qualification design. But is it important enough to be considered a fundamental design principle for any qualification design process? We would argue that it is, while also accepting that the nature and uses of proficiency modelling are likely to differ across qualification types.

Interestingly, within the educational measurement profession, the need to base the design of an assessment procedure on an explicit model of its target proficiency is largely taken for granted nowadays, and what remains open for debate is simply the nature of the required modelling (see Perie & Huff, 2016, for instance).[10] Indeed, if we think of the proficiency model as the 'construct' at the heart of construct validity, then we can appreciate just how central the idea of proficiency modelling has been for decades (see Messick, 1989, for example).[11]

The critical importance of an explicit proficiency model is perhaps easier to see through a measurement, or certification, lens than through a curriculum lens. After all, the core function of a qualification is to provide a credible confirmation of the acquisition of a certain level of a certain kind of proficiency. If members of a certification community for any particular qualification are left in the dark concerning the nature of the proficiency that it certifies, or the standard(s) at which it certifies proficiency, then it is hard to see how the qualification could be fulfilling this core function.

What tends to set outcome-based qualification design apart is the focal role that the proficiency model plays when planning curriculum, pedagogy, and assessment. This

---

[10] This often defaults to a content-by-skill specification of one sort or another (see Schmeiser & Welch, 2006, for example) although it sometimes involves a far more sophisticated model.

[11] Unfortunately, the term 'construct' is often used imprecisely in the literature, so it is frequently unclear whether the author is referring to the (real-world) target proficiency or to the (constructed) proficiency model or even to something in between (Newton, 2012; Slaney & Garcia, 2015).

relates directly to the idea of securing alignment between teaching, learning, and assessment, facilitated by a common proficiency model. This is perhaps why outcome-based qualification design is often associated with contexts in which much of the responsibility for planning curriculum, pedagogy, and assessment is devolved locally – for example, to colleges or to employers – where the need for conceptual scaffolding of this sort is particularly high, in order to help secure accurate and consistent interpretation from one role to the next, and from one setting to the next.

Yet, even when there is a clear division of responsibility between, say, teachers and trainers (operating locally) and a single exam board (operating centrally), there is still a fundamental requirement for alignment between teaching, learning, and assessment. Historically, exam boards in England failed to model the proficiency at the heart of each qualification syllabus, other than implicitly through meagre content lists. This raised a significant threat to alignment: as noted earlier, the classic critique of classically designed qualifications was that they tended to over-assess 'lower-level' competencies (such as recall and comprehension) and under-assess 'higher-level' ones (such as evaluation and application), with negative backwash impacts on teaching and learning.

Nowadays, even classically designed qualifications tend to incorporate a variety of features that are designed to explicate their target proficiency, such as assessment objectives, grade descriptions, and sometimes even explicit learning outcomes (see Annex 1 of report 2). Yet, whereas the proficiency model for an outcome-based qualification tends to be explicit, complete, and discrete – which equips it for becoming the principal point of reference for planning curriculum, pedagogy, and assessment – proficiency modelling for classical qualifications tends to be more fragmented and multiform. To some extent, this has arisen from a desire to model the target proficiency in different ways for different members of the certification community given the different uses to which the model will be put. For instance, grade descriptions might be developed as a guide for students and certificate users, whereas content-by-skill grids might be developed as a guide for assessment developers. Although this fragmented and multiform approach is likely to be better than no modelling at all, it is not necessarily optimal for securing alignment across the system. The nature and uses of proficiency modelling for classical qualifications would certainly benefit from further research, analysis, and theorisation.[12] Of course,

[12] We suggest that classical qualifications are better theorised than outcome-based ones in relation to stage 3 (assessment procedure) whereas outcome-based qualifications are better theorised than classical ones in relation to stage 2 (proficiency model). This brings us back to the question that we asked in report 4 concerning what an 'outcome-based' approach was supposed to stand in contrast to, given that the idea of 'input-based' is not very helpful. Addis & Winch (2019) suggested a different kind of contrast – outcome-based versus holistic – which captures something important, although classical qualifications are not specified entirely holistically by any means.

there are many important resources to draw upon toward this end (for example, Sadler, 1987, and Ahmed & Pollitt, 2011, to name just two). But, what is really needed is an integration of insights from across these divergent resources, within an overarching integrated theory of educational certification.

Anticipatory qualification design

In report 4, we concluded that the most important lesson to learn from our historical research strand concerned the risk of conceptualising and operationalising qualification reform too narrowly. This is often associated with focusing too heavily on assessment implications, with insufficient attention to the wider education and training changes that are necessary for a reform to bed in, particularly the need to support teacher and trainer development from the outset. We argued that qualification reforms are best understood as education and training reforms that are initiated through changes to certification requirements. As such, there needs to be a clear line of sight from the point at which qualification design is initiated through to each stage of its implementation. This means considering (up front) the implications of the qualification reform for curriculum planning and instructional intervention, for assessment development and delivery, and for the interpretation and uses of qualification certificates.[13]

We might refer to this as 'anticipatory qualification design' because it involves anticipating all of the arrangements that need to be put in place:

- for the qualification to succeed (anticipating likely requirements for effective rollout and ensuring that they are fulfilled)
- for the qualification not to fail (anticipating likely threats to effective rollout and mitigating them)

Considerations of this sort are relevant to the design of any qualification, not simply to the design of outcome-based qualifications. The following subsections help to illustrate these 2 facets of anticipatory qualification design.

Anticipating requirements

Although it might seem obvious that a qualification designer ought to anticipate likely requirements for effective rollout in order to ensure that they are fulfilled, this has not always happened in the past for a number of reasons. For instance, exam boards in England have traditionally shied away from pedagogical prescription, assuming that teachers ought to be free to choose how they teach. Indeed, this is part of the reason why syllabuses of the 1970s and earlier tended to be so loosely specified, to permit teachers as much flexibility as possible in their approaches to instruction. In other

[13] This theme is entirely consistent with the importance that many of the awarding organisations that we interviewed placed on anticipating teaching and learning needs, as described in report 6.

words, a tendency for pedagogy to be viewed as 'off limits' to qualification designers may have risked this perspective being overlooked entirely.[14]

A slightly different situation arises when regulatory bodies – like the National Council for Vocational Qualifications – specify qualification design principles as accreditation criteria for awarding organisations. This is how the CASLO approach became such a dominant feature of the TVET landscape in England, as a high-level qualification design template. What this situation imposes, unfortunately, is a series of structural barriers in the line of sight from initial qualification design to its implementation in schools and colleges. These barriers are illustrated in Figure 3, which is an adaptation of Figure 17 from report 4, mapped onto stages 2 and 3 of our idealised design framework.

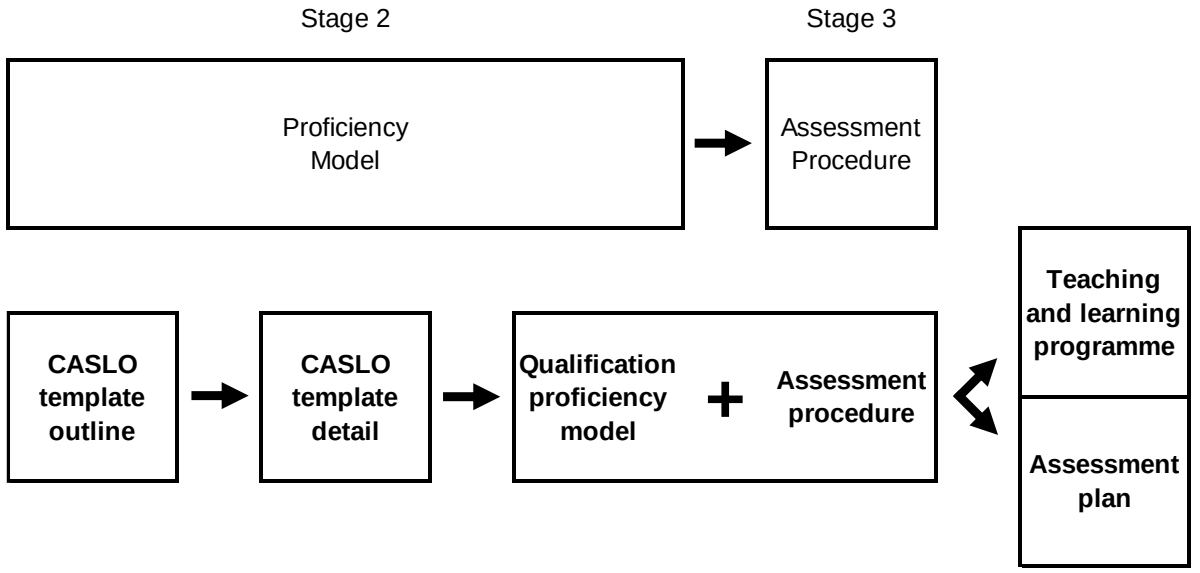

Figure 3. Barriers to securing successful design and delivery

Within the NVQ system, the NCVQ assumed responsibility for designing the NVQ model, that is, it assumed responsibility for the high-level CASLO template (template outline – first square lower left). This bare template was handed over to the Training Agency, which assumed responsibility for co-ordinating the development of National Occupational Standards (NOS) and NVQs. By producing industry-specific NOS, bodies such as the Industry Training Organisations (ITOs) added flesh to the bones of the CASLO template (template detail – second square lower left). Subsequently, these NOS were integrated within NVQs by awarding organisations, with additional

[14] See also Wood (1968) on the related resistance by exam boards to specifying objectives.

elaboration of assessment requirements (central rectangle). Finally, teachers and trainers had to assume responsibility for planning learning programmes and assessment opportunities (vertical squares on right).

Importantly, this meant that responsibility for the critical task of proficiency modelling was distributed across multiple organisations, operating in a linear (siloed) fashion. Thus, the:

- NCVQ designed the blueprint that would be used for developing proficiency models
- ITOs developed bespoke proficiency models for occupational functions (NOS)
- awarding organisations re-worked those proficiency models for occupational qualifications (NVQs), potentially providing additional elaboration via guidance, exemplification materials, and suchlike

The risk associated with this siloed production line approach to qualification design is that the requirements of those toward the end of the line – particularly those responsible for planning and delivering teaching and assessment – may end up being insufficiently understood by designers working at the beginning of the line.

Stanton, a former director of the Further Education Unit (FEU), observed that the linear approach to NVQ design, development, and delivery had caused serious problems for teachers and trainers (Stanton, 2012).[15] He argued that preventing problems of this sort required an iterative approach, not a linear one (see also Stanton, 2016). More specifically, standards needed to be developed in collaboration with teachers and trainers, to maximise the likelihood of them being meaningful and useful when subsequently translated into teaching and learning programmes (see also Callender, 1992; CAVTL, 2013). The lack of a clear line of sight from designers to assessors might also help to explain why NOS were originally viewed as being "marred by complex, jargon ridden language" (Beaumont, 1996, page 13).

These considerations explain why, within the second stage of our idealised framework for qualification design, we proposed that the process of constructing a proficiency model should include consideration of pedagogical and assessment implications. As such, stage 2 is best understood as specifying a proficiency model with a certification community, and not simply for a certification community.

In terms of supporting key protagonists, Stanton (2012) also emphasised the risks associated with handing standards over to teachers and trainers without further elaboration, as though they somehow constituted a programme of teaching and learning, which is how (in the absence of satisfactory support) they were often

[15] The FEU was set up in 1977 under the auspices of the Department of Education and Science to promote good practice in relation to curriculum and pedagogy.

treated. It is clear that system designers (like Gilbert Jessup from the NCVQ) recognised that NVQ standards would need to be translated into coherent and sequenced learning programmes, particularly for younger learners who were starting from scratch. Yet, they did not see it as part of their remit to support this, and other players within the system had insufficient time and resource to support activities of this sort adequately.

The lack of support for teachers and trainers who are tasked with delivering brand new outcome-based qualifications has been a recurrent theme since the very first ones were introduced in England (see report 4). Bear in mind that when a reform process replaces a classically designed qualification with an outcome-based one, this necessitates an entirely different pedagogical approach (and vice versa, of course, see Braun, 2018). Furthermore, according to Blank (1982), the author of a well-known handbook for developing competency-based training programmes, outcome-based approaches are far more pedagogically complex, as they necessitate constant monitoring of progress. This was a theme that we developed in our functional strand (report 3). In a review of competency-based vocational education and training in Australia, Watson (1991) argued that it can only fulfil its potential if it is carefully and effectively implemented, which requires:

1. initial and ongoing commitment to the provision of adequate resources and learning materials
2. adequate preparation and ongoing staff development for teachers
3. adequate preparation of learners for the new approach to teaching, learning, and assessment

If so, then it is perhaps not surprising that 'teething problems' are experienced when teachers and trainers are provided with insufficient (or no) support for pedagogical planning during the early years of radically reformed qualifications.[16] A Cedefop-commissioned research report from 2012 suggested that the relative neglect of pedagogical implications may be an international phenomenon associated with the shift to learning outcomes:

> interviews with policy-makers pointed out that, while national policies have focused mainly on reforms to the written curriculum, comparatively little consideration has been given to how new curricula might affect teaching and learning and what types of pedagogies might best support outcome-oriented curricula.
>
> (Stanley, 2012, page 16)

[16] Note that Ofsted pays far more attention nowadays to the effectiveness of curriculum and pedagogical planning, alongside the possession of adequate expertise for these activities, as laid out in the Further Education and Skills Inspection Handbook (see paragraph 230, for example).

Fullan's discussion of 'change theory' suggests that the lack of attention to required changes in teaching practice is actually a more general problem affecting educational reform efforts internationally (Fullan, 2006).[17]

One final illustration of the importance of anticipating requirements for effective rollout relates to the most fundamental of all requirements: sufficient demand for the new qualification. Stephanie Allais has surveyed the international shift to learning outcomes extensively, and one of her most damning conclusions concerned how frequently many of the outcome-based qualifications that populate new national frameworks end up with very low uptake or none at all (Allais, 2014). Clearly, there is no point in developing a new qualification simply to fill a notional gap in a qualification framework. There needs to be a genuine demand for any new qualification. Qualifications that are expected to be low-certificating, despite being genuinely important, require particular attention to ensure their viability.

Anticipating threats

Anticipating likely threats to effective rollout, then putting controls in place to mitigate them, is what effective qualification design is all about. So, in theory, this ought to be very familiar territory. Yet, anticipatory qualification design reminds us that – as part of the qualification design process – we need to look beyond the usual threats to assessment validity, to consider both threats to teaching and learning effectiveness and threats to the appropriateness of certificate use. Where threats of this sort can be anticipated prior to rollout, controls can be put in place to help mitigate them.[18]

As we have already identified, it is entirely possible to anticipate a variety of threats that stem from the nature of the CASLO model itself, including risks related to deconstruction, disintegration, and simplification. More straightforwardly, we have plenty of evidence from ineffective rollout testifying to the seriousness of threats of this sort (and the importance of exploring options for mitigating them). For instance, our first report on the critical strand of our research programme (report 5) identified 2 potentially serious problems for CASLO qualification rollout, which both relate to the

[17] Research is beginning to unpick the ingredients for effective reforms. For instance, in the context of Kenyan primary education in literacy and numeracy, Piper, et al (2018) found that adding teachers' guides (which included daily lesson plans) to a package of professional development, instructional support, and pupil text books had a dramatic impact on the improvement of learning.

[18] Threats cannot always be anticipated, of course. This is particularly true when the context within which a qualification system is located changes significantly over time. For instance, the use of qualification results for accountability purposes gradually increased in prominence from the 1980s to the 2010s, and the ways in which results were used changed too, both of which made it increasingly challenging to accommodate teacher assessed components (Ofqual, 2012; Newton, 2023b).

disintegration risk: ‘atomistic assessor judgements’ (risking invalid assessment) and ‘lack of holistic learning’ (risking ineffective teaching and learning).

These potential problems both arise from the fact that CASLO qualifications model learning outcomes as discrete elements, which (to a greater or lesser extent) fails to capture important structural information concerning the interconnectedness of the target proficiency. This includes the need for integration and co-ordination across elements of knowledge and skill, which becomes evident when they need to be applied within complex, authentic performances.

The principal threat, here, relates to the risk of atomistic teaching – teaching elements of knowledge or skill, one-by-one, with no attention to their integration and coordination – leaving students unable to apply their learning in a meaningful way. A secondary threat relates to the risk of atomistic assessment – assessing elements of knowledge or skill, one-by-one, with no attention to their integration and co-ordination – which can result in students who are unable to apply their learning in a meaningful way still being judged to be competent (as their lack of competence has not actually been revealed).

We discussed potential problems of this sort with awarding organisations (as detailed in report 6). The use of synoptic or holistic assessment tasks was often described as a mitigation for both of them. For instance, a unit with 3 learning outcomes might be assessed via a single assignment, involving a complex, authentic performance task. This might, for instance, be an extended project for a business qualification, or a professional service for a hairdressing qualification. Although, the CASLO specification would still be presented as a list of learning outcomes and associated assessment criteria – each of which would need to be ‘ticked off’ to pass the unit – they would need to be demonstrated in the context of an integrated and co-ordinated performance. From a validity perspective, this makes it far safer to generalise from atomistic assessment judgements to the conclusion that a learner is genuinely competent. From an effectiveness perspective, preparing for complex, authentic assessment tasks should have a positive backwash impact on teaching and learning.

Finally, whether we describe it as a basic requirement or a threat mitigation, the most fundamental guarantor of effective qualification rollout is the professional expertise of everyone who plays an important role in it. This conclusion received strong support from our critical strand (see reports 6 and 7). It is professional expertise that enables practitioners to fill in the proficiency model ‘gaps’ that are an inevitable consequence of any modelling process. In the VTQ context, teachers and trainers require both occupational and educational expertise, which is doubly challenging. Without occupational expertise, they will not be able to fill in the assessment-related gaps that occur because, for example, assessment criteria inevitably fail to capture the qualification standard with total clarity. Without educational expertise, they will not be

able to fill in the teaching-related gaps that occur, for example, because learning outcomes inevitably fail to set out an optimal trajectory of learning for passing the qualification. Of course, occupational and educational expertise is important for any vocational teacher or technical trainer, so this challenge is not unique to CASLO qualifications.

Ecclestone addressed issues of this sort in the report of an extensive research project with schools and colleges, entitled ‘Transforming Formative Assessment in Lifelong Learning’. With a focus on further education settings, she illustrated the pervasive threat of instrumentalism and low expectations. Rather than assuming that succumbing to this threat is somehow inevitable, she focused instead on how teaching and learning experiences could be improved. As the first of 4 factors that she believed to be crucial to all formative assessment approaches that are genuinely educationally worthwhile, she identified:

> strong and confident expectations of motivation from well-qualified subject teachers who are confident, expert and enthusiastic
>
> (Ecclestone, Davies, Derrick, & Gawn, 2010, page 216).

Building better

Earlier, we suggested that policy makers – nationally and internationally – may have paid insufficient attention to how outcome-based qualification design can fail. This helps to explain the plentiful supply of evidence of outcome-based qualifications having lacked validity or having led to negative impacts on teaching and learning. However, we also suggested that scholars may have paid insufficient attention to how outcome-based qualification design can succeed. The silver lining to our plentiful supply of evidence of failure (from the academic literature) is that we now have a strong basis for proposing a wide range of mitigations with the potential to help us to build better CASLO qualifications (specifically) and better outcome-based qualifications (more generally). This takes us to the concluding sections of our report.

Moving forward

The CASLO approach became so dominant within the landscape of regulated VTQs in England that it assumed the appearance of being 'the' alternative to adopting a classical approach to qualification design. In fact, its combination of tightly specified outcomes, tightly specified criteria, and a stringently applied mastery principle represents just one alternative, even when restricting our purview to outcome-based or mastery-based qualifications. In addition, a more nuanced analysis of the current qualification landscape challenges the idea of the CASLO approach as a straightforward 'alternative' to the classical one, given the growth of hybridised and blended qualifications. Figure 4 provides a more nuanced illustration of the variety of approaches currently available to qualification designers in this space.

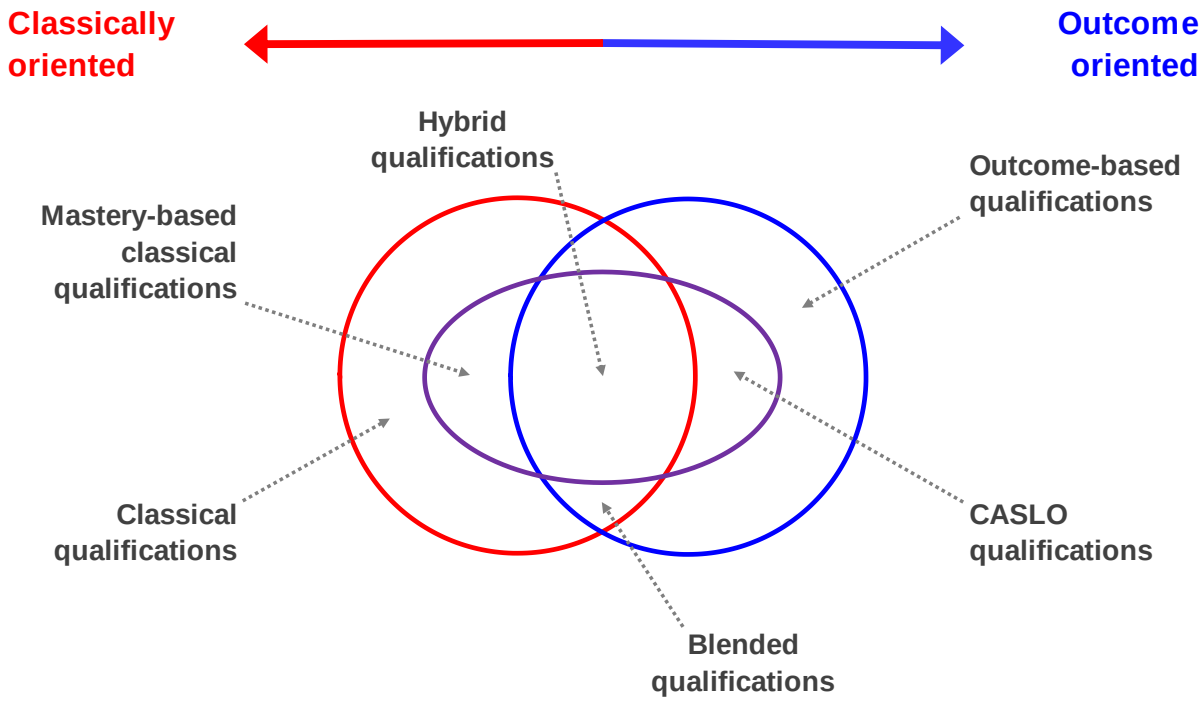

Figure 4. Alternative approaches to qualification design

Within this Venn diagram, the red circle represents a classically-oriented approach and the blue circle represents an outcome-oriented approach. Note that these circles overlap, which implies that the differences between classical and outcome-based approaches are far more blurred nowadays than they might have seemed back in the 1970s. This space is best characterised as a continuum of approaches to qualification design, with some oriented more toward a classical approach and some oriented more toward an outcome-based one.

We have overlaid a purple oval on these 2 circles to indicate qualifications that are designed to certify mastery. CASLO qualifications are located at the intersection between the blue circle and the purple oval (toward the right). Conversely, the intersection between the red circle and the purple oval (toward the left) reminds us that it is quite possible to adapt a classical approach to certify mastery. For instance, accountancy qualifications often adopt this approach: they rely heavily upon the written exam format with numerical marking, but they operationalise mastery via a high pass mark, for example, 70% of the mark total.

It is worth emphasising that adopting an outcome-based approach to qualification design does not commit the designer to also adopting a mastery-based approach. As explained in report 4, the very idea of mastery learning was introduced because of a tendency for teachers to commit considerable effort to articulating important learning outcomes, yet without committing equal effort to ensuring that substantial numbers of students actually achieved them all. Some of the BEC awards (that preceded BTECs) might be said to have adopted an outcome-based approach without formally certifying mastery (see report 4 for further details).

Hybrid qualifications sit right in the middle of the Venn diagram, typically incorporating classical components alongside outcome-based ones. For instance, A level practical science assessments function as outcome-based components within (primarily) classical qualifications. Similarly, when level 3 BTECs were reformed to comply with DfE performance table requirements, they typically swapped a number of CASLO units for classical ones (see report 8 for an example of this).

Sometimes, nowadays, it can be hard to distinguish a classically designed qualification from an outcome-based one purely on the basis of their specification document, especially for qualifications that do not adopt a mastery-based approach. The A level in business discussed in the annex of report 2 provides a good example of this. Where qualifications adopt features associated with both classical and outcome-based qualifications (rather than adopting both approaches side-by-side) we might describe this as a blended approach.

This analysis invites us to think creatively about the significance of outcomes and mastery when designing vocational and technical qualifications for the future. The final 3 sections of this report are headed:

1. ‘revisiting qualification purposes’ – we consider the extent to which the goals that originally drove adoption of the CASLO approach continue to resonate with present-day needs
2. ‘reconsidering qualification design’ – we invite the sector to use findings from our research programme as a resource for interrogating existing practices and for investigating new approaches

3. ‘tackling systemic challenges’ – we identify 3 major threats to the effective operation of VTQ systems in England, which became apparent through conducting our research

Revisiting qualification purposes

In report 4, we reached 2 major conclusions related to fitness for purpose:

1. the CASLO approach cannot be said to be universally fit, nor universally unfit, for purpose
2. it is not easy to render CASLO qualifications fit for purpose, and in some ways they are harder to render fit for purpose than classically designed qualifications

These conclusions were supported by evidence from the other strands too.

If the CASLO approach is not universally fit for purpose, then it makes sense not to specify it as a blanket requirement for all regulated qualifications. Equally, though, if it is not universally unfit for purpose, then we need to consider when it might be the most appropriate approach and when not. Reaching a decision of this sort, for any particular qualification, requires us to scrutinise its intended functions, goals, or purposes (in relation to the cohort for whom it is being designed and the contexts within which it will need to operate). Sometimes a purpose-cohort-context analysis of this sort will recommend an outcome-based approach, such as the CASLO approach, but not always.

We concluded that adoption of the CASLO approach in England has been driven by many different goals over the decades, which have influenced different qualifications and qualification frameworks in differing ways and to differing degrees. The following 4 educational goals appear to have been particularly important. They relate to improving:

1. domain alignment – to align curriculum, pedagogy, and assessment as closely as possible with the intended domain of learning (and therefore also with each other)
2. domain mastery – to ensure that all students achieve a satisfactory level of attainment across the full domain of learning
3. qualification efficiency – to make the process of becoming qualified as efficient as possible
4. domain personalisation – to enable the domain of learning to be tailored to the personal situation, interests, or needs of learners (or customised to meet the needs of local employers)

The critical question, moving forward, is the extent to which these goals continue to resonate with present-day needs. We will illustrate what an analysis of this sort might

look like by considering each of these goals in turn.[19] Note that the following analysis is restricted to these 4 goals, although they are merely a subset of the full set of purposes that would need to be considered when designing a new qualification.[20]

Alignment

Earlier in this report, we observed that alignment was a fundamental consideration for any approach to qualification design. We also explained that outcome-based approaches were developed specifically to counter the threat of misalignment that arises when a domain of learning is explicated primarily via a content list with no indication of what students are expected to be able to 'do' with that content (for example, recall it, apply it, evaluate it).

Whether outcome-based approaches always manage to secure effective alignment is an entirely legitimate question to ask, and it seems fair to conclude that they do not always achieve this goal. However, the goal itself is hard to argue with when certifying achievement, and it is still highly relevant across all qualification contexts today.

Mastery

CASLO qualifications embody a particular conception of proficiency, whereby having reached the overall proficiency threshold means having reached the proficiency threshold across each and every element of the domain of learning in question. We describe this as having mastered the domain of learning (in full). The classic example of this concerns occupational competence, whereby having reached full competence in an occupational role means that there are no longer any aspects of that role that cannot be performed competently. For the purpose of certification, these aspects are represented in terms of discrete learning outcomes, and having achieved full occupational competence means having demonstrated sufficient proficiency across all specified learning outcomes.

The mastery goal – which is to ensure that all students achieve a satisfactory level of attainment across the full domain of learning – can be understood either as a moral principle or as a technical one (or both). The moral imperative for mastery derives from Bloom's original work on mastery learning, which insists that all students ought to be supported to achieve the full set of valued learning outcomes, whatever the

[19] Reports 6 and 7 provide additional insight into their continuing relevance.

[20] For instance, qualifications whose results will be used for accountability purposes – via school or college performance tables – may need to be designed with those purposes in mind, for example, by making them resilient to malpractice or playing-the-system. All sorts of intended purposes may need to be taken into account during stage 1 of qualification design (see Newton, 2007; 2023a).

domain of learning. The technical imperative for mastery simply notes that certain domains of learning – often occupational ones – need to be fully mastered in order to guarantee safe and effective practice.

The technical imperative continues to hold in many certification contexts, although we do not believe that it holds for all VTQs, nor even for all qualifications that are designed to certify occupational competence. So, from a technical perspective, there would still seem to be mileage in the CASLO conception of mastery certification, albeit not for all regulated TVET qualifications.

The question of whether (or the extent to which) the moral imperative continues to hold is trickier to answer. On the one hand, influential policy reviews, including both Dearing (1996) and Wolf (2011), have questioned the value of mastery for students on general or vocational (that is, non-occupational) courses. Both reports argued that mastery certification was only relevant to learners studying for technical (occupational) qualifications. Yet, by rejecting the idea of mastery from a technical perspective – for students on general or vocational courses – they also appeared to dismiss it from a moral one too, whether intentionally or not.

On the other hand, the concept of mastery teaching and learning has received a lot of attention over the past few years, especially in the further education sector, which includes being actively promoted by one of England's former Prime Ministers.[21] This is not mastery in exactly the sense that it is operationalised within CASLO qualifications, but it still embodies very similar ideas and values. Mastery learning also features as a strand in the Education Endowment Foundation Teaching and Learning Toolkit, where it ranks 6th for impact. So, there does still seem to be some enthusiasm for the general concept of mastery in England, today, including within policy making circles.

It is interesting to note, however, that the idea of mastery teaching and learning has not always featured prominently within academic reports on vocational pedagogy. An interesting example is a report by Lucas and colleagues, which claimed to provide proof of concept for a theory of vocational pedagogy (Lucas, Spencer, & Claxton, 2012). It had been commissioned by City & Guilds, not long after the QCF had been introduced, at a time when the vast majority of vocational qualifications would have been premised upon mastery learning. Yet, although it referenced "mastery of everyday working procedures" and "mastery of literacy, numeracy and digital literacy" within 2 of its 6 desired outcomes from vocational education (page 9), it

[21] On 4 October 2023, Prime Minister Rishi Sunak announced an initial investment of £600 million to lay the groundwork for delivering the proposed new Advanced British Standard, including new investment "to promote the highly successful teaching for mastery in maths methods".

made no explicit reference to either mastery learning or mastery teaching.[22] In a section headed "Working competence, but *not* a checklist of 'competences' or skills" (page 38), the report discussed criticisms associated with CASLO qualifications. Given the historical association between outcome-based qualifications and mastery learning, it may be that Lucas felt that neither deserved to feature explicitly within a theory of vocational pedagogy.

In seeking to understand the extent to which mastery teaching and learning are valued in today's educational climate, it will be important to distinguish clearly between different conceptions of mastery, especially as different conceptions will have different implications for both mastery learning and mastery certification. It may also be worth considering whether 'mastery' is the clearest term to be using in this context, and whether other terms might be more useful. For example, the mastery aggregation principle – which is built into the CASLO approach and other outcome-based approaches – does not always tally with the more general concept of having 'mastered' an occupational role. In fact, mastery certification is often associated with minimal (threshold) competence within a role, rather with having fully mastered it, which suggests that 'mastery' is not the best term to be using here.

Efficiency

We defined the efficiency goal in terms of making the process of becoming qualified as efficient as possible, which helps to make qualifications as accessible as possible. Its central mechanism involves breaking down the demands of an overarching qualification into separately certifiable or assessable chunks, that is, separately certifiable units, and separately assessable learning outcomes within units. This facilitates efficiency in a variety of ways, for instance, if a learner:

- can furnish independently verified evidence of having already achieved multiple learning outcomes from a unit, then they might apply for Recognition of Prior Learning (RPL) for those outcomes, to exempt themselves from having to undertake further (unnecessary) teaching and assessment
- starts and finishes their course of learning on a different timeline from other learners within a teaching or training cohort, clarity over the specification of expected learning outcomes makes it easier to track and support progression

[22] The report did, however, reference formative assessment (teaching "through feedback") as the 4th of 18 tried and tested teaching and learning methods. Note that the concept of mastery learning did feature more explicitly within a separate report commissioned by City & Guilds with the Learning and Skills Network around the same time (Faraday, Overton, & Cooper, 2011).

- wants to take a break from learning part way through a course, then they might 'bank' their successfully completed units until they are ready to complete the remaining ones sometime later
- wants to switch to a different qualification pathway – for example, from site carpentry to architectural joinery – within a qualification that begins with units that are common across pathways, then they will not have to start the qualification again from scratch (assuming they switch after completing the common units)

Some of the potential for efficiency is associated with unitisation, per se, and not specifically with adopting an outcome-based approach. Outcome-based design simply adds greater transparency to the unit-level certification. Yet, in certain circumstances, including RPL, outcome-based specification is substantively important to the process, providing a basis for assessing (coherently and defensibly) at a level that is lower than the unit.

Efficient processing is a double-edged sword as far as achieving a qualification is concerned. On the one hand, making qualifications more accessible sounds like a no brainer. If we can achieve this in a coherent and defensible manner, then why not? On the other hand, efficiency of this sort can run the risk of unhelpfully punctuating a learning journey that might be better undertaken continuously. In other words, in certain circumstances, we might risk sacrificing educational value for qualification efficiency. More pragmatically, having to accommodate RPL for multiple learners within a cohort can be challenging for teaching and training providers, with idiosyncratic teaching and learning needs potentially even threatening the viability of course provision. In short, there is clearly a cost-benefit ratio to be considered when evaluating the desirability of building qualifications (and qualification frameworks) to support the efficiency goal.

Personalisation

Outcome-based approaches have the potential to support qualification personalisation (for learners) or customisation (for employers). We identified 2 distinct mechanisms by which this might operate:

1. writing learning outcomes in a certain way, and with sufficient generality, can open the door for learners to acquire and demonstrate those outcomes in different contexts
2. specifying unit content in terms of outcomes and criteria reveals the nature (content), breadth (size), and depth (complexity) of the domain of learning in considerable detail, which helps to facilitate judgements concerning unit comparability

The first mechanism assumes that outcomes can be written in such a way that they can be applied (in essentially the same manner) across multiple learning contexts,

such that one student might acquire and demonstrate them in one context while another might acquire and demonstrate them in another. Imagine, once again, a carpentry qualification with 2 pathways – site carpentry and architectural joinery – with learners on each pathway studying separately in their own context. Then imagine that the first 2 units of this qualification comprise learning outcomes that are written with sufficient generality to enable them to be acquired and demonstrated in either context. These outcomes might be acquired and demonstrated in slightly different ways, from one pathway to the next, but without differing enough to counter the presumption that they represent the same outcomes acquired to the same standard.

An extreme example of this mechanism at work occurs when the target proficiency for a particular qualification is constructed primarily, or even exclusively, in terms of skills. This is sometimes associated with an assumption that skills ultimately matter more than knowledge, that skills can be acquired independently of knowledge, and that the same skillset can be acquired in the context of different bodies of knowledge. If this were true, then it might provide a warrant for treating knowledge as little more than a context for skill acquisition, and for tailoring the teaching and learning programme for any particular group of students to whatever body of knowledge seemed most likely to motivate them. With a narrative grounded in the development of 21st Century Skills, this type of thinking has become fashionable over the past couple of decades in general education settings, albeit triggering a significant backlash (see Priestley, 2017). To avoid confusion, it is important to remember that, while outcome-based approaches can be used to design a purely skills-based curriculum or qualification, they were actually introduced to help secure an appropriate balance between knowledge and skills. More generally, although outcome-based approaches have the potential to support a certain amount of cross-context personalisation or customisation, they do not presume it, and they can straightforwardly be designed to reduce or eliminate the potential for personalisation and customisation.

The second mechanism revolves around the observation that units or qualifications that serve similar functions are typically expected to be comparable in terms of the nature (content), breadth (size), and depth (complexity) of their intended learning outcomes. Adopting an outcome-based approach – particularly one that uses criteria to flesh out the standard – helps to facilitate up-front judgements of comparability.[23] Judgements of this sort are important as a warrant for treating different units as though they are interchangeable or exchangeable. For instance, judgements of

[23] They are 'up-front' in the sense of being based purely upon scrutiny of unit specifications, without also scrutinising student performances. By way of contrast, judgements of comparability for general qualifications tend also to involve scrutinising student performances (after assessments have been taken and the performances have been marked).

comparability provide a basis for defending the use of optional units (alongside mandatory ones) within a single qualification. More extremely, they provide a basis for defending the principle of mixing and matching units to form bespoke qualifications, which was the principle at the heart of the QCF.

Although the QCF formalised the idea of rules of combination, which set important parameters for mixing and matching units, it still ended up being criticised for encouraging the creation of customised qualifications that lacked coherence (Whitehead, 2013). Withdrawal of the QCF signalled a new emphasis on the validity of a qualification, as opposed to its flexibility. As such, demand for radical personalisation, or customisation, of this sort seems to be lower now than it was a decade or so ago. Having said that, the idea of accrediting and accumulating small steps in learning is very attractive, and frequently re-enters national and international debates. Just recently it has resurfaced in the shape of micro-credentialling (see Cedefop, 2023, for example).

Reconsidering qualification design

We embarked upon this programme of research on the assumption that the better we (in the sector) understand the CASLO approach:

- the better our policy making will be
- the better our regulatory practices will be, and
- the better our qualification design, development, and delivery will be

In our programme overview document (report 1) we explained that our methodology was not intended to be evaluative, in the sense of pitting the CASLO approach against the classical approach, or even in the sense of evaluating the CASLO approach independently (in terms of the particular goals that it tends to prioritise). Instead, our research has been descriptive and analytical, in the sense of characterising the CASLO approach both independently and comparatively (compared to the classical approach).

This strategy has generated a wealth of insights into policies, practices, and principles related to the rise of outcome-based and mastery-based approaches over the past half century, in England, particularly within the TVET landscape. The previous section invited us to consider the extent to which goals that drove adoption of the CASLO approach continue to resonate with present day needs. To the extent that they still do, outcomes from our research programme can help us to frame critical questions concerning the effective operation of outcome-based and mastery-based qualifications in England, both now and into the future. The following subsections illustrate questions of this sort, framed in terms of: interrogating existing practices, and investigating new approaches.

Interrogating existing practices

Outcomes from the functional, historical, and critical strands of our research programme constitute a resource that can be drawn upon to provoke and support conversations between certification professionals concerning:

- goals that can be served by adopting the CASLO approach
- principles underpinning the effective operation of CASLO qualifications
- threats to the effective implementation of CASLO qualifications
- strategies that can be adopted to mitigate implementation threats

This resource provides a basis for interrogating CASLO qualification practices – related to their design, development, and delivery – guided by the overarching question of their continuing fitness for purpose. Because CASLO qualifications serve differing purposes, for differing learners, in differing contexts, there are no simple guidelines for ensuring fitness for purpose. However, we have identified a variety of general principles associated with effective practices, and a variety of general observations associated with ineffective ones, both of which are useful for exploring fitness for purpose.

Fitness for purpose

The critical strand of our research programme considered the extent to which criticisms that were levelled against NVQs, GNVQs, and other early CASLO qualifications might generalise to CASLO qualifications of the present day. Our conversations with awarding organisations and other stakeholders suggested that problems of the sort identified by the academic literature remain relevant today. Report 6, in particular, indicated that awarding organisations not only recognised potential problems of this sort, but put steps in place to minimise the likelihood of their occurrence and to mitigate their impacts.

The following 3 subsections illustrate how outcomes from our research programme can be used to interrogate practices associated with existing CASLO qualifications. They are framed in terms of the 3 major categories of ‘potential problem’ for CASLO qualifications that were discussed in reports 5 to 7 of our critical strand: assessment problems (validity concerns), delivery problems (viability concerns), and teaching and learning problems (impact concerns).

Validity

Many learners in England complete their schooling without ever studying a CASLO qualification, as general qualifications tend to be based on a classical approach. This renders the CASLO approach unfamiliar to a large proportion of the general public.

The fact that CASLO qualifications operate quite differently from classical ones raises a straightforward question concerning how, exactly, they do work. Although these qualifications are, of course, very familiar to the many learners who do study them – not to mention their teachers, trainers, and the awarding organisation officers who support them – their practices are surprisingly poorly documented and their underlying principles are even less well articulated.

In the report on our functional strand (report 3) we attempted to surface some of these underpinning principles, based on conversations with awarding organisation officers. For instance, we identified 7 principles that appear to underpin the effective operation of CASLO qualifications:

1. structural integrity (centres that aspire to offer a qualification need to demonstrate their wherewithal to deliver it – from a teaching, assessment, quality assurance, and a management perspective – to the satisfaction of the awarding organisation before being permitted to deliver it)
2. self-regulation (centres need to operate as self-regulating systems, having internalised the awarding organisation's quality standards and quality assurance practices)
3. comprehensive monitoring (quality assurance needs to focus upon the effective delivery of the qualification in the round – teaching, assessment, quality assurance, and management)
4. risk-based sampling (given the potential scope of quality assurance activities, sampling needs to be driven by a risk-based model)
5. incremental improvement (centres need to be supported to improve incrementally – from centre approval onwards and for as long as a centre continues to offer the qualification)
6. supportive surveillance (the relationship between an awarding organisation and its centres – and between an Internal Quality Assurer and their assessors – needs to remain fundamentally supportive)
7. conditional, evidence-based trust (with increasing competence comes increasing trust – but, if a centre were to abuse that trust, then this would need to be dealt with via an appropriate sanction)

Because outcome-based design is premised on the idea of achieving clarity and consensus concerning qualification expectations – for teachers, trainers, learners, assessors, certificate users, and wider stakeholders alike – it is tempting to think that the CASLO approach ought somehow to be transparent, straightforward, and unproblematic. If so, then presumably CASLO qualifications ought, by design, to be easier to implement than classical ones. Conversely, we concluded that they can actually be harder to render fit for purpose, partly as a consequence of how transparent they attempt to be. Report 3 underlined this conclusion, by explaining

why running a CASLO qualification can be more demanding than running a classically designed qualification, owing to the nature of the relationship that has to be established between an awarding organisation and its centres, which is far closer and more hands-on than is generally the case for classical qualifications. This received further support from report 6.

Beyond general principles for effective practice, we identified a variety of general observations concerning ineffective practices, and we invited awarding organisation officers to explain how they attempted to mitigate them. Report 6 is packed full with observations of this sort, which should prove useful for provoking and supporting conversations concerning the fitness for purpose of existing CASLO qualifications.

For instance, although the explication of learning outcomes and assessment criteria is supposed to provide the basis for securing clarity and consensus across the certification community for any particular CASLO qualification, the literature is clear that written statements of this sort are insufficient for securing accurate and consistent interpretation of qualification standards. Awarding organisations in our study recognised this potential problem and discussed how they mitigated it. For example, none of them relied solely on assessment criteria to communicate critical performance thresholds. Most provided guidance documents to help support accurate and consistent interpretation. Some referenced separate industry-specific documents, such as codes of practice, industry regulations, or treatment protocols. Several provided centres with exemplars of student work to support standardisation exercises and ongoing judgements against assessment criteria. Where exemplars were provided, this was sometimes on a priority basis – for core units or for units that had proved hardest to standardise – where providing exemplars for the full complement of units might not be viable (for example, where a qualification offered a large number of optional units). Awarding organisations also recognised the value of a strong community of practice (centred around the qualification) in supporting accurate and consistent judgements. Several awarding organisations explained that they actively promoted networking and community building among their centres – conferences, forums, training events, online groups, and so on – though some acknowledged that there was more that they might be able to do in this respect.

Viability

The biggest practical threat to the effective operation of a CASLO qualification concerns the risk of an unacceptable assessment burden. This can be a very significant risk when developing an entirely new qualification that is likely to be delivered with constraints over timing, for example, as a key stage 5 study programme. This proved to be a significant problem when GNVQs were introduced.

The risk of unacceptable burden is high for CASLO qualifications owing to the need to assess all specified learning outcomes. It is important to mitigate this risk, to

ensure that the time required for teaching and learning is not eaten away by the time required for assessing. That said, CASLO qualifications offer considerable potential for integrating teaching, learning, and assessment, as long as boundaries between formative and summative are respected (such that summative performances are not inappropriately scaffolded by formative feedback).

Beyond the risk of unacceptable assessment burden, there is also the risk of unacceptable administrative burden, given the necessity of record keeping, and the demands of maintaining extensive portfolios. Record keeping is critical, of course, to support internal and external quality assurance activities. The risk is simply that it ends up eating into personal time, which is a significant threat for teachers and trainers of CASLO qualifications, especially where centres do not formally set time aside for quality assurance activities (which was true for some of the participants in our teacher focus group interviews from report 7).

Although comprehensive assessment is a feature of the CASLO approach, most awarding organisations explained that the risk of undue burden had lessened in recent years with the introduction of electronic technologies, including e-portfolio systems. Interviews with teachers and trainers (from report 7) supported this conclusion. They recognised the risk and reality of burden, but also spoke positively about the benefits of awarding organisation e-platforms, particularly those providing a visual display of student progress. These new technologies made portfolio completion and record keeping more user-friendly and motivating than in the past. Having said that, some of the awarding organisations we spoke to explained that their use was optional, and some centres were unable or unwilling to use them.

Finally, some awarding organisations suggested that weaker assessors were more prone to over-assessment, and that the provision of training had helped to mitigate this risk.

Impact

Just as important as the threat of invalidity, if not more so, is the threat of negative backwash impact on teaching and learning arising from qualification design decisions. We have already considered various threats of this sort: for instance, the risk of atomistic teaching – teaching elements of knowledge or skill, one-by-one, with no attention to their integration and co-ordination – leaving students unable to apply their learning in a meaningful way. We also noted steps that awarding organisations have put in place to mitigate threats of this sort: for instance, the use of holistic or synoptic assessment tasks, which are designed to have a positive backwash impact on teaching and learning.

Before moving on, though, it is worth mentioning an issue that sits on the borderline between validity and impact, which helps to illustrate the importance of interrogating the compromises that need to be made when designing qualifications. The issue at

stake, here, concerns the generality with which learning outcomes and assessment criteria are written. From a validity perspective, the greater the specificity, that is, the greater the detail, the greater the potential for clarity and consensus concerning qualification expectations, and (at least in theory) the greater the potential for securing accurate and consistent interpretation of qualification standards. What awarding organisations repeatedly described to us, however, was a need to trade off a validity-driven desire for specificity against an impact-driven desire for generality. Generality, in this instance, was deemed necessary to enable courses to be tailored to the personal situation, interests, or needs of learners, or customised to meet the needs of local employers. The risk of being very specific – which might, for instance, be achieved by embedding outcomes and criteria in the context that happened to be most prevalent within a sector – is that it has a negative backwash impact on those learners who would benefit from situating the acquisition and demonstration of their competence in a different-yet-comparable context. Appropriately situated learning, of this sort, reflects the cross-context personalisation (or customisation) goal, which we described in detail in report 4.

Evaluation

We hope that outcomes from our research programme will help certification professionals to interrogate the design, development, and delivery of existing CASLO qualifications. Design, development, and delivery comprise the first 3 stages of the qualification lifecycle, but we should not forget the fourth stage, which is qualification review. Ofqual has a specific rule related to qualification review, which states that:

> D3.1 An awarding organisation must keep under review, and must enhance where necessary, its approach to the development, delivery and award of qualifications, so as to assure itself that its approach remains at all times appropriate

This includes the sort of informal interrogation that we have just considered. Yet, guidance accompanying this rule also references positive indicators of compliance such as “undertaking in-depth periodic reviews” which hints at a more formal evaluation process. Published outcomes from formal evaluations are important in providing a warrant for the trust that is the bedrock of an effective qualification system (or a stimulus for change when outcomes are less favourable).

Outcomes from the present research programme suggest that there is also scope for formally evaluating CASLO policies (as well as practices). For instance, although Ofqual no longer requires any awarding organisation to adopt the approach anymore, there is still scope for evaluating policies, including Ofqual regulations, that effectively proscribe it (which, in some cases, have been associated with the hybridisation of qualifications).

Investigating new approaches

While the previous section invited awarding organisations to interrogate practices associated with existing CASLO qualifications, this section invites the sector to investigate new approaches to designing qualifications, including outcome-based or mastery-based approaches that might depart from the CASLO template. This takes us into territories that are tricky to navigate, owing to the lack of a clearly demarcated field of research, not to mention the lack of an overarching integrated theory.

Work in this field has a long pedigree, for instance, the idea of job analysis is over a century old (Sanchez & Levine, 2012) and techniques continue to be refined to the present day (see, for instance, Clauser & Raymond, 2017, on practice analysis). The idea of specifying educational objectives has a similarly long pedigree (see Reeve, 1925, for example). Yet, this work is spread across a variety of disparate subdisciplines – from curriculum studies to industrial and organisational psychology – and it has not been effectively integrated. So, it is sometimes hard to know where to turn for insights and guidance. Moreover, only some of the work in this territory pivots specifically around implications for qualification design (while other work focuses more on implications for training or selection, for example). As we proposed earlier in this report, integrating insights from disparate disciplines within a qualification design context is what educational certification theory ought to be fundamentally concerned with.

As our research programme developed over the past few years, it became increasingly clear that the technology of proficiency modelling – stage 2 of our idealised framework for qualification design – is still in its infancy (despite its roots tracing back over a century). Because we believe that it is the highest priority area for future research and development, we have decided to focus exclusively upon it within the following subsections. That is not to say that we believe it to be the only important area for future research and development, of course.

While the field of proficiency modelling for qualification design is still in its infancy, there is important work being undertaken today – both nationally[24] and internationally[25] – just as there has been for many decades. In the following section, we will consider a very small selection of work of this sort, simply to provide a rough map of the territory. The purpose of this section is not to offer guidance on how to choose between the various proficiency modelling methods that have been

[24] For example, by Cambridge University Press & Assessment, including: Child & Shaw (2020), Suto, Greatorex, Vitello & Child (2020), Vitello, Greatorex & Shaw (2021).

[25] For example, by Cedefop, including: Hart, Noack, Plaimauer & Bjørnåvold (2022), Cedefop (2024b). The tome entitled ‘Competence-based Vocational and Professional Education’ (Mulder, 2017b) is another important resource.

developed to date (although the 'modelling approach' subsection identifies a small number of them in Figure 5). Instead, the purpose is to consider the features that any proficiency model might (or might not) incorporate – the general features of a proficiency model – as though we were designing a proficiency modelling method from scratch.[26]

Modelling proficiency

Stage 1 of our idealised framework for qualification design involves identifying – and reaching consensus over – the full range of purposes (or goals) that explain why the qualification that we are working on is deemed to be necessary for our targeted cohort of learners and the contexts or circumstances that they find themselves in (be those educational contexts, training contexts, working contexts, family contexts, or any other relevant context). If these purposes, cohorts, and contexts are sufficiently well elaborated and prioritised, then the issue of what that qualification will need to certify – that is, the nature of its target proficiency – should follow logically from this profile.

Yet, while this might be true in theory, the process of teasing out those logical implications to construct an explicit proficiency model is not at all straightforward. It corresponds to stage 2 of our idealised framework, and the following subsections describe some of the choices and challenges that certification professionals face when employing the technology of proficiency modelling. We begin, though, with a single subsection that concerns decisions that need to be taken during stage 1, which enable us to identify the target proficiency.

Agreeing target proficiency parameters

We start here because of how controversial this stage has proved to be in relation to the design of TVET qualifications in England, especially those that are studied just prior to transitioning from schooling into the workforce. For qualifications of this sort, one of the most important decisions to make concerns the nature of the learning that the target cohort will need to acquire. There are various types of learning that might or might not be suitable for pre-transition learners, for instance, learning that:

- qualifies them to work safely and competently in a specific occupational role

[26] This high-level analysis is intended to be illustrative rather than comprehensive, and does not drill down into the detail of how the 'contents' of any particular proficiency model are best identified and specified. For instance, it does not consider the process of collaborating with domain experts to elicit and capture their understandings of the target proficiency, the level of generality or specificity built into the model, or suchlike.

- provides them with capabilities that enable them to secure a job in a general occupational sector (from where they can develop full occupational competence)
- familiarises them with (and positively disposes them towards) the world of work
- equips them with generic skills for life and work
- empowers them with traditional academic knowledge and disciplinary skills
- empowers them with confidence in their personal ability to learn and develop

Different commentators adopt different positions on the nature of valuable learning for pre-transition learners. For instance, there is a longstanding, and unresolved, debate concerning the relative importance of liberal education versus vocational preparation for learners in the final years of compulsory education, which is discussed in detail by Pring (1995).[27] Winch (2012) explored a slightly different theme, arguing straightforwardly that England puts too much emphasis on narrow vocational training at the expense of broad vocational education.

Decisions concerning the appropriate balance between different types of learning need to be informed by what we think will be good for learners, good for the economy, and good for society. But they also need to be tempered by recognition of the realities of the contexts that students learn in, and of the workforce into which they will transition. For instance, Wheelahan & Moodie (2018) have argued that in wealthy liberal market economies, such as the UK and Australia, links between qualifications and occupations are very weak, with the exception of a limited number of regulated occupations (for example, heating engineers). Quoting from an Australian National Centre for Vocational Education and Research report, from 2016, they noted that:

> only 38 per cent of those who graduated from an apprenticeship or traineeship in a non-trade occupation programme were employed in an occupation directly associated with their qualification, which shows that apprenticeships and traineeships are not necessarily a panacea for ensuring tight matches between qualifications and occupational outcomes
>
> (Wheelahan & Moodie, 2018, page 180)

Note that figures for graduates from professional programmes and clerical and administrative programmes were considerably lower. Observations of this sort not only have implications for the type(s) of learning that might be optimal for pre-transition students, they also have implications for the kind(s) of information that qualification results will need to provide. In heavily regulated markets, there are

[27] For instance, should we think of education as intellectual development for its own sake or as something that ought to be engineered for the sake of economic prosperity.

strong links between VTQs and occupations: certificates are used to identify applicants who have acquired specific bodies of knowledge, skill, and capability. In largely unregulated markets, there are weak links between VTQs and occupations: certificates tend largely to be used to screen (from the rest) those applicants who appear to have the highest potential (Wheelahan & Moodie, 2018).

Modelling approach

Once agreement has been reached concerning the type(s) of learning that need to be acquired and demonstrated – and the broad outline of the target proficiency for the qualification in question has become much clearer – the more technical challenge of articulating a formal proficiency model arises. Report 4 indicated the potential scale of this challenge through the example of National Vocational Qualifications, which were intended to certify full occupational competence (in line with the first of the above bullets).

Not only did the National Council for Vocational Qualifications insist that NVQs ought to adopt the CASLO approach to modelling full occupational competence, it also insisted that outcomes should only be specified in terms of elements of competence, that is, in terms of what performing an occupational role competently actually looked like (the activities it comprised). Importantly, this meant not specifying occupational competence in terms of elements of knowledge or understanding. While not denying the importance of underpinning knowledge and understanding, the NCVQ believed that occupational competence could and should be defined independently of it, as NVQs were supposed to certify competence itself (not the knowledge and understanding that might enable it). This decision proved to be highly controversial.

This example illustrates the more general point that there are many different ways of approaching the challenge of proficiency modelling. Even if the CASLO approach is chosen, it still needs to be decided whether to produce a model that is framed in terms of ‘internal’ constructs like knowledge and understanding (which appear to focus on the cognitive prerequisites for performing an activity successfully), or ‘external’ ones like competence or capability (which appear to focus directly on the ability to perform an activity successfully), or some combination of the two.[28]

[28] The implication is that ‘internal’ constructs are not amenable to direct observation, while ‘external’ constructs are more amenable. Philosophically, this takes us into very murky waters, but the distinction works well enough at a commonsense level.

Approach	**Main development methodologies**	**Common formats**
Internal, individual, attribute-based		
Technocratic or syllabus-led	Derived from knowledge-base or course syllabus	Tasks expressed as application of knowledge
Instructional design	Job analysis, learning needs analysis	Table of (tasks with associated) knowledge, skills and (often) attitudes
Behavioural	Critical incident analysis and variants, repertory grid technique	Behaviours, approaches and attributes associated with effective job performance
External, social, activity-based		
Task-based	Task analysis, work study, DACUM	Descriptions of tasks and their component parts
Role-based	Functional analysis	Descriptions of job functions and detailed activities within them
Profession- or field-based	Analysis of activities across profession, role mapping	Descriptions of activities that enable effectiveness across a profession

Figure 5. Approaches to proficiency modelling (adapted from Lester, 2017)

Figure 5 is adapted from Lester (2017) simply to illustrate the potential for adopting very different approaches in this respect. In the figure, the NVQ approach is classified as external and role-based. In contrast, a classical approach might be classified as internal and syllabus-led. Reflecting on the distinction between internal and external approaches to proficiency modelling, Lester concluded that:

> the external approach to competence has proved particularly useful where standards of practice are needed rather than standards to guide education and training, and where there is a need to assess competence as a practitioner rather than as a novice entering the workplace.
>
> (Lester, 2015, page 4)

An earlier paper by Lester (2014) provided additional detail concerning his distinction between occupational (role-based) and professional models of competence. He proposed that professional (as opposed to occupational) models:

- focused on ethics, professionalism and key standards (rather than the details of roles and functions)
- are designed to apply across the profession (rather than having a core-and-options structure), and
- provide confidence in practitioners' abilities to act as a member of the profession (rather than in a bounded occupational role)

In fact, he proposed that these differences were best understood as characterising 2 ends of a continuum rather than as binary alternatives. Based on results from an earlier study, Lester noted that professions tend to devote significant attention to generic, core aspects of professionalism and professional activity – reflecting a 'centre-outwards' approach to proficiency modelling – with a heavy emphasis on ethics, professional development, self-management and management of work, as well as communication or client relations. They also tend to want to model a form of proficiency that is broader than mere 'competence'. Lester referred to this broader construct as 'capability' and described it as: not just the ability to do, but the ability "to become (more) able to do" (Lester, 2014, page 280) particularly when moving into new areas or responding to changing contexts or demands (Lester, 2015, page 5).

It is important to note that professional certification of this sort tends to be more closely associated with accreditation to formal membership of a profession than with qualification to perform a specific role. A proficiency model that is fit for an accreditation function will not necessarily also be fit for other functions (Lester, Koniotaki, & Religa, 2018).

Modelling breadth and depth

Following the decision to pursue an internal, external, or mixed approach, the next question to consider concerns what the elements of the model will look like. More specifically, we need to distinguish the different kinds of element within the model. For instance, from an internal perspective, it is common to draw a fundamental distinction between knowledge and skill, to which we sometimes add understanding. Hence, the distinction between knowing, understanding, and being able to do.

To unpack a domain of learning in terms of both breadth and depth, Anderson & Krathwohl, et al (2001) proposed a 2-dimensional taxonomy for learning, teaching, and assessment, which classified elements of proficiency – more specifically, learning objectives – into dimensions of knowledge and cognitive processes. Their knowledge dimension was subdivided into 4 categories: factual, conceptual, procedural, and metacognitive (each broken down into multiple subcategories). Their cognitive processes dimension was subdivided into 6 categories: remember, understand, apply, analyse, evaluate, create (again, broken down into multiple subcategories). The authors described the knowledge dimension as a roughly ordered continuum, stretching from most concrete (factual) to most abstract (metacognitive). They described the cognitive processes dimension as a more clearly ordered continuum, stretching from least complex (remember) to most complex (create).

This taxonomy was presented as a revision of the most influential of all such taxonomies, Bloom's taxonomy. However, there are many different taxonomies, models, and frameworks to choose from. For instance, Moseley, Baumfield, Elliott, Gregson, et al (2005) scrutinised 42 systematic approaches to describing thinking and its relation to learning and teaching, including both Bloom's taxonomy and its revision. They explained how these different approaches tend to be based upon different organising principles, arguing that certain approaches were better suited to certain uses. For example, they believed that the most useful framework for assessment was the SOLO taxonomy of Biggs & Collis (1982), particularly when used for formative purposes.

Biggs, et al (2022) described the SOLO taxonomy in a chapter on knowledge and understanding in the context of higher education. It is basically an approach to elucidating levels of understanding. Whereas Bloom's taxonomy and its revision distinguish between more and less complex categories – for example, remember versus create – the SOLO taxonomy attempts to represent how understanding develops, becoming increasingly structured and articulated over time. So, higher levels of this taxonomy subsume lower ones, as listed below (with examples provided in parenthesis):

1. prestructural (the student misses the point)

2. unistructural (the student can perform a simple procedure)
3. multistructural (the student can use algorithms)
4. relational (the student can explain causes)
5. extended abstract (the student can theorise in a domain)

Note that the second and third levels are more about increasing knowledge, while the fourth and fifth are more about deepening understanding.[29]

Modelling complexity and progression

Although this is a new subsection, the next set of ideas is really just an extension of the last subsection. They concern the principle that a proficiency model ought either to represent (explicitly) or at least embody or respect (implicitly) the idea of progression in learning. This is simply because the idea of learning invokes the idea of journeying along a pathway from novice to expert. This, in fact, is exactly how Dreyfus conceptualised skill acquisition, in terms of 5 stages (Dreyfus, 2004):

1. novice (who is limited to adhering rigidly to rules and plans)
2. advanced beginner
3. competence
4. proficiency [30]
5. expert (who no longer relies on rules, guidelines, or maxims, relying instead on vision and deep tacit understanding)

This model has been applied productively to the context of professional learning by Eraut (2008). Lester (2017, page 71) provided an example of a proficiency model developed by the Law Solicitors' Regulation Authority, which transformed the Dreyfus classification into a 2-dimensional matrix using the following row headings (with the Dreyfus categories as column headings):

1. functioning knowledge
2. standard of work
3. autonomy
4. complexity

[29] Biggs has criticised the Anderson and Krathwohl revision for, in effect, trivialising the concept of understanding, which is the organising principle for his own taxonomy (see Biggs, et al, 2022, page 88) but merely the second least complex cognitive process category for Anderson and Krathwohl.

[30] Incidentally, the concept of proficiency as used in the present paper (for example, the idea of a proficiency model) is generic, and not limited to a specific level.

5. perception of context
6. innovation and originality

This model was elaborated by placing relevant descriptors in each of the cells of the matrix, such that the descriptors located in the ‘competence’ column would indicate what it meant to be competent across all 6 dimensions. Lester explained that the minimum standard for sign-off, generally after 2 years of training, was ‘competent’, but solicitors who were partners or practising independently would be expected to have reached the ‘proficient’ level.

Across a number of reports, Winch has developed a more general framework for improving the transparency of professional qualifications, which is designed to articulate different kinds of proficiency as well as to represent progression in proficiency. It is based on the idea of ‘epistemic ascent’ (Winch, 2013; 2014; 2015; 2020). Figure 6 summarises the core features of his framework, and is adapted from Winch (2020).

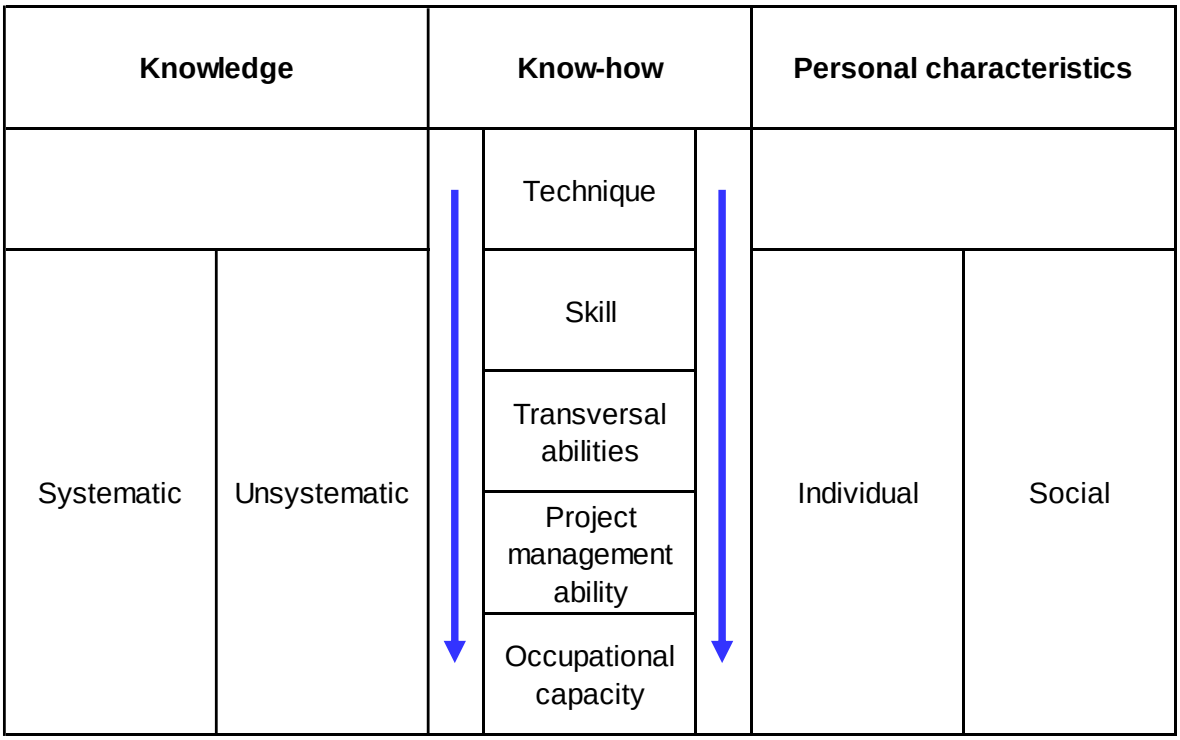

Figure 6. Framework for professional qualifications

This framework draws a distinction between everyday, localised, unsystematised knowledge and systematically organised knowledge, both of which are important to professional development. Winch noted that systematic knowledge is typically

(although not exclusively) associated with academic disciplines. It is organised in terms of: how its conceptual elements are classified, the relationships between them, and the procedures required to gain and validate knowledge. This implies that learning in a domain necessitates a structured approach to progression. He proposed that when occupational competence requires the application of systematic knowledge it can be characterised as either ‘technical’ or ‘professional’ (although this is not ‘professional’ as distinct from ‘occupational’ as defined by Lester).

Winch also proposed that a qualification ought to provide a social guarantee of certain personal characteristics, both individual ones (important when working alone) and social ones (important when working with others). These might include: the exercise of responsibility, the ability to work with others, the ability to manage oneself, the ability to carry out extended work over a significant time period, conscientiousness, courage, and so on.

Finally, Winch illustrated the idea of epistemic ascent most graphically in terms of gradations of practical ability, characterising the trajectory of a learning journey from novice to expert as follows:

1. technique – mastery of the way in which a procedure or task is (supposed to be) carried out
2. skill – the ability to carry out a procedure or task in contextually relevant conditions
3. transversal abilities – the ability to plan, co-ordinate, control, communicate, evaluate, and so on (which require an element of care and attention for successful completion)
4. project management ability – the ability to successfully complete work that is substantially longer and more complex than a procedure or task, and that requires judgement (implying a sense of autonomy from the planning stage, though execution, to evaluation)
5. occupational capacity – the ability to practice an occupation as a whole, which requires deep knowledge of the principles underlying occupational practice and an awareness of the impact of the occupation on society

This extended progression underlines the fact that practical ability – know-how – cannot be reduced to the simplistic idea of skill. Winch explained that these 5 kinds of practical ability tend to be nested, such that possession of the first is necessitated by possession of the second. For work that involves a high degree of autonomy, and a deep understanding of the occupation and its role in society, each of the 5 kinds will need to be acquired. Together, these practical abilities enable a practitioner to put their specialised knowledge into practice through professional judgements in work situations.

Finally, it is important to emphasise that representing knowledge, know-how, and personal characteristics independently does not imply that they are either acquired or demonstrated independently. According to Winch, the reverse is true. Higher levels of practical ability will typically necessitate deeper knowledge and stronger personal characteristics. This point has been underlined by Lester and colleagues:

> simply appending knowledge and skills to practice activities is not generally a good way of developing a curriculum, as it tends to ignore how knowledge builds from general principles to more specific and applied concepts, and how know-how, skills and techniques need to be built into the larger sequences of action that, particularly although not only for higher-level work, are necessary to underpin competent practice
>
> (Lester, et al, 2018, page 298)

Over the past few decades, TVET qualifications in England have relied very heavily on Bloom's taxonomy as a basis for modelling complexity and progression:

- from lower-level qualifications to higher-level ones (operationalised through the concept of a framework level), and
- within a single qualification (operationalised through the award of grades above the pass threshold)

This has not always proved to be optimal and has sometimes seemed paradoxical (see Newton, 2018). There is certainly scope for modelling complexity and progression more meaningfully and usefully in the future (Pellegrino, Chudowsky, & Glaser, 2001). Recent work on learning progressions is consistent with this aim (for example, Wilson, 2009; Shepard, 2018; Gallacher & Johnson, 2019).

Modelling structure

Once again, although this is a new subsection, the next set of ideas is really just an extension of previous ones. They concern the principle that a proficiency model ought either to represent (explicitly) or at least embody or respect (implicitly) the ways in which its component elements fit together. This relates to the need for integration and co-ordination that comes into play when elements of knowledge and skill need to be applied within complex, authentic performances (as discussed earlier).

It is very hard to represent explicitly how elements of knowledge and skill need to be integrated and co-ordinated within complex, authentic performances. However, there is one fundamental dimension of structure that does tend to be modelled explicitly, which is often discussed (from an assessment perspective) in terms of aggregation. An aggregation model specifies how the bits of assessment evidence that have been collected need to be combined when reaching an overall judgement concerning

learner proficiency. Importantly, this is far from an arbitrary decision, because aggregation modelling needs to reflect expectations concerning what it means to develop expertise within a particular domain of learning.

In a very useful report on the relevance of sampling within competence-based qualifications, Alton, Boyle, & Limmer (2021) explored this structural issue directly. They distinguished between 4 approaches to aggregation modelling:

1. compensation – whereby strong performance on one learning outcome can compensate for substandard performance on another (allowing candidates to pass despite not having achieved all specified learning outcomes) [31]
2. conjunction – whereby the standard has to be reached for each and every learning outcome independently (so candidates can only pass if they have achieved all specified learning outcomes)
3. complementarity – whereby one set of learning outcomes can be substituted for another (allowing candidates to pass as long as they have met the standard for one of the exchangeable sets)
4. disjunction – whereby the standard only needs to be reached for a subset of the specified learning outcomes (allowing candidates to pass despite not having achieved all specified learning outcomes) [32]

Again, these are not simply assessment requirements, as they relate to the anticipated structure of competence, or proficiency, within a domain of learning. Consider disjunction, for example. Although this principle does not feature much within the landscape of regulated VTQs in England, it is not actually that uncommon as conception of developing expertise. In fact, it is perhaps more common as a conception at higher levels of expertise, where we do not necessarily expect exceptional performers to be exceptional at everything within their domain of practice (and, when they work as members of teams, they do not need to be).

A critical question for the present report is how to determine which aggregation model (or combination of models) is appropriate within any particular qualification context. There are different kinds of answer to this question. A descriptive approach might refer to how developing proficiency tends to manifest itself in any particular

---

[31] Compensation seems to suggest an algorithmic process, whereby strengths and weaknesses are literally averaged out, one by one. Indeed, this is exactly how it is operationalised when marks for individual assessment items are added to reach an overall mark total. Alternatively, Sadler outlined a non-algorithmic compensatory model, which he associated with standards-referenced assessment. He termed it “configuration” and explained that the key to meeting the standard from this perspective is being able to demonstrate a certain “pattern of performance” over time (Sadler, 1987, page 193).

[32] Their report also described in detail a ‘negative’ conception of disjunction, whereby substandard performance on certain learning outcomes results in automatic failure, for example, not observing health and safety regulations.

context of learning. For instance, when learning how to speak a language, do we envisage this in terms of having to jump through a series of learning hoops (like a conjunctive model) or do we envisage it more as gradually approximating the performance of a native speaker (more like a compensatory or configuration model)? Conversely, a normative approach might set expectations for the structure of successful learning that actively challenged how developing proficiency tends to manifest itself. Bloom's mastery learning principle provides a good example of this, where the acquisition of all specified learning outcomes is essentially a moral principle rather than a technical one. A more technical prescriptive approach might refer to what a practitioner needs to know, understand and be able to do in order to practice safely and effectively.

The report by Alton, et al (2021) explored this more technical, prescriptive approach through a series of detailed case studies within a variety of occupational areas. Perhaps their most important conclusion was that qualifications tend not to map neatly onto these aggregation models, which prevents us from being able to conclude, for instance: "this is a conjunctive system, therefore its design must be thus" (page 3). They proposed that their 4 models were pragmatic rather than ideal, in the sense of not being wholly mutually exclusive, nor wholly internally consistent. Consequently, across each of their case studies, they were able to identify features associated with more than one of the 4 aggregation models. They described a complex, multifaceted, and heavily contextualised approach to aggregation modelling for competence-based qualifications. On the specific issue of sampling, they concluded:

> Perhaps what emerges most clearly is that for sampling within a qualification to be successful the following need to be true: there must be sector buy-in; the content must be as fully specified as possible; there must be validity evidence that what is measured is relatively important, and that what is omitted by sampling is relatively unimportant; what is omitted must not become predictable; and any element that is considered essential is exempted from the general principle.

(Alton, et al, 2021, page 3)

Purposeful variation

If it is true that the field of proficiency modelling for qualification design is still in its infancy, then it seems reasonable to conclude that:

- there must be room for improvement in relation to existing practices
- any new practices that we propose, at least in the near future, are unlikely to solve all of the problems that we may have encountered in the past

In this context, there is a strong argument for supporting purposeful variation, which encourages the use of different models across the qualification system, as long as there is a persuasive rationale for each one. Policy makers (and regulators) sometimes find it hard to appreciate variation. After all, if one approach is better than another, then surely we ought to expect all parties to adopt the superior approach? In addition, the presence of variation makes it more complicated to understand the system. Indeed, the reduction of variation has been pursued as a policy objective since the 1970s (see report 4).

Unfortunately, the problem with excessive rationalisation is that one size never fits all. We have learnt from experience of the NVQ framework and the Qualifications and Credit Framework that attempting to impose a single model on the entire VTQ system results in a certain amount of misfit (see report 4). What tends to happen in this situation is that misfitting qualifications are adapted to comply 'as far as possible' with the single model, which ultimately means that: they are still not actually based (entirely or coherently) on the intended model, but neither are they based (entirely or coherently) on any appropriate model.

A better solution, in situations like this, would be to permit variation in the underpinning model, but to do so purposefully. Purposeful variation implies 3 conditions:

1. there is a clear rationale whenever one model varies from another, and strong logical grounds for believing that each model variant is likely to be fit for its own profile of intended purposes, cohorts, and contexts
2. the fitness for purpose of each of the model variants is monitored over time, as the basis for developing a research-based foundation for better policy making and practice in the future
3. the extent of variation in the system is not so great as to seriously undermine the intelligibility or credibility of the system [33]

In short, to embrace purposeful variation is to embrace a learning culture, which acknowledges and accommodates the limitations of our current level of understanding, and that actively builds a stronger foundation for future generations. Learning from variation is the payback from a more complex system. It is the only way to determine whether certain approaches might ultimately prove to be better than others, in relation to any particular qualification purpose-cohort-context profile.

[33] Part of the work that remains to be done concerns exploring the circumstances that are more or less tolerant of variation, and identifying the criteria that might be used to judge its acceptability.

Tackling systemic challenges

The idea of purposeful variation helpfully introduces the final section of this report, which considers implications from our current programme of research for enhancing the operation of VTQ systems in England. Although the breadth of our programme means that all sorts of insights might be noted at this point, we believe that the following 3 challenges are the most important for the sector to address as we move forward:

1. deeper understanding
2. safer innovation
3. stronger structure

Deeper understanding

The impetus for our 2020 to 2024 research programme came from a recognition that VTQs in England – particularly those that adopt the CASLO approach – are extremely poorly documented, theorised, and researched. This goes some way towards explaining the pervasive lack of policy memory in our sector. We hope that the research that we have reported will play a small role in addressing this deficit, but there is a long way to go. We see deeper understanding as the lynchpin of progress, so we will discuss it in some detail before reflecting on the remaining 2 challenges. We note the role of stronger structure as a foundation for deepening understanding, and the role of deeper understanding as a foundation for safer innovation.

Problems

We began report 4 by noting how the lack of policy memory in our sector has been widely recognised as a serious problem for many years. City & Guilds has published a series of reports on this issue, under the heading ‘Sense and Instability’ (2014; 2016; 2019). The same diagnosis led the Edge Foundation to commission a series of reports in its ‘Learning from the Past’ series.

We also noted a particular problem for the sector related to a lack of documentation of the principles underpinning the effective operation of VTQ systems, especially in relation to the CASLO approach. There are, of course, many official documents that have set out requirements or expectations for VTQs in England, past and present. The ‘NVQ Code of Practice’ (QCA, 2001) would fall into this category, as would ‘Key Questions about Assessment in the QCF’ (QCDA, 2010). But documents of this sort tend to be highly operational, and they rarely go beyond surface-level

recommendations to unpick underpinning principles.[34] There is also a small number of textbooks that have been produced to support teaching and learning in the sector, which have typically been reproduced in multiple editions to keep track of a rapidly changing landscape. These include the 'Vocational Assessor Handbook' (Ollin & Tucker, 2012), 'Principles & Practices of Quality Assurance' (Gravells, 2016), and 'The Best Quality Assurer's Guide' (Read, 2012), as well as the 'The Qualification Manager's Handbook' (Federation of Awarding Bodies, 2017). Yet, even texts like these are fairly high level and pragmatic. In report 3, we attempted to explore beneath the surface of CASLO qualification practices, in an attempt to unpick the underpinning principles that appeared to guide their effective operation. There is much work still to do in this territory, though.

We also noted that VTQs in England – and CASLO qualifications in particular – were poorly researched. This might sound a little paradoxical, given that we identified a large (in fact huge) peripheral literature on topics as diverse as educational objectives, criterion-referenced assessment, competence-based education and training, and so on (see report 5). In fact, even our core literature – which focused on issues of relevance to understanding or evaluating the CASLO approach in the context of UK VTQs from the 1980s onwards – was pretty big (see report 5). Yet, on the other hand, it is important to appreciate that this core literature focused primarily upon just 2 qualifications – the NVQ and the GNVQ – and was concentrated during the 1990s. We found relatively little relevant research in relation to other CASLO qualifications from the 2000s or 2010s.

A significant chunk of the academic literature on NVQs and GNVQs focused on their outcome-based design. These reports were often conceptual, which indicates that there is a certain amount of theoretical work in this area. Yet, as the architects of the NVQ and GNVQ systems failed to unpack the positive case in favour of outcome-based qualification design to the extent that critics unpacked the negative one against it, the literature feels skewed. On the one hand, the conceptual case for outcome-based qualification design was largely assumed but not explicated. Yet, on the other hand, the conceptual case against outcome-based qualification design was often explicated in detail but (in our view) with a tendency to draw conclusions that were too extreme (see report 4). Again, there is much work still to do in this territory.

Finally, as discussed earlier in the present report, we have gone so far as to suggest that we lack an overarching integrated theory of educational certification, which hints at a far broader challenge that stretches beyond the VTQ sector.

---

[34] Incidentally, we also found it very hard to locate syllabus documents from prior to the turn of the millennium, which made it tricky to establish how even fairly recent qualifications operated.

Causes

If we are going to address the challenge of developing deeper understanding, then it helps to begin by reflecting on why problems of the sort just described might have arisen in the first place. Part of our ongoing challenge undoubtedly relates to an interaction between the complexity of the VTQ sector and the lack of foundational education, training, or qualifications for certification professionals in England.

Unlike accountancy, or medicine, or engineering, there is no established educational or training route into the certification profession. So, people do not qualify to become certification professionals before they begin practicing. Practitioners at all levels tend to learn on the job. Moreover, as already noted, there are very few bespoke education or training resources to support certification professionals in England when learning on-the-job. It is true that the profession is supported by over a century of research and development in the field of educational measurement (for example, Wood, 1993; Brennan, 2006), and by research and development from related fields, such as curriculum studies (for example, Kelly, 2009). Yet, there is no generally accepted curriculum for certification professionals – formal or informal – and the profession is not grounded in professional learning to anywhere near the same extent as accountancy, medicine, engineering, or many other professions.

Added to this problem is fact that the VTQ sector can be a black box to anyone who has neither studied, nor taught, a vocational or technical qualification, which is often the case for VTQ policy makers (see, chapter 5 from Lenon, 2018, on ‘Failure in vocational education, 1800–2015’). This is compounded by the complexity of the VTQ sector, and the very different demands that are attributable to a wide variety of:

- applications (general education, applied education, technical training, career progression, personal development)
- clients (teenagers, returning adults, experienced employees)
- learning needs (confidence building, competence building, preparation for life, preparation for university)
- personal challenges (educational, social, financial)
- learning contexts (off-the-job, on-the-job, unemployment, part-time, full-time)
- occupational traditions (the peculiarities of each industrial and commercial area)

Finally, as we discussed in report 4, there are problems arising from the ‘structure of knowledge’ within the VTQ sector, which relates to the intellectual ownership of systems, processes, and practices. Historically, in England, guilds and professions have assumed a prominent role in these respects. Discussing the formation of City & Guilds by the Corporation of London and a number of Livery Companies, Lang (1978) suggested that Great Britain was the only industrial country where

responsibility for technical standards was not managed solely by the state. Thus, City & Guilds was established, in 1878, to provide a national system of technical education at a time when there was none. City & Guilds operated collaboratively, authoritatively, and autonomously, until government began to rationalise the VTQ landscape from the 1960s onwards.[35]

From that point on, government and government-sponsored agencies, including the TEC, the BEC, and later the NCVQ, the QCA, the DfE, Ofqual, and others began to exert greater control over the landscape, which included specifying technical qualification design rules, often set out as qualification framework accreditation requirements. Increasingly, it became the norm for awarding organisations – like City & Guilds and many other much smaller bodies – to develop qualifications in accordance with centrally-specified regulations. Whatever the merits of organising a qualification system like this, one of the fundamental challenges it raises concerns intellectual ownership of the qualifications that populate these frameworks, partly in a commercial sense, but primarily in the sense of who assumes ultimate responsibility for their quality and relevance.[36] Historically, for City & Guilds qualifications, the buck would have stopped with City & Guilds. Nowadays, in the context of distributed responsibilities, it is less clear.

It seems possible that this ambiguity may also help to explain the lack of documentation, research, and theorisation in the VTQ sector, especially in relation to the CASLO approach. The approach achieved national prominence because it was required by a government-sponsored agency, the NCVQ. Yet, it was rolled out by a plethora of independent awarding organisations, including City & Guilds and many others too. Although these organisations assumed operational responsibility for the quality of their qualifications, it is not clear the extent to which they assumed intellectual ownership, and therefore a sense of ultimate responsibility for quality and relevance within the system. This raises the question of where intellectual ownership and ultimate responsibility for quality and relevance ought to be expected to reside when a system is structured like this. A critical test case for this question relates to who assumes operational responsibility for creating and embedding knowledge through research, development, education, training, and publication.

We have a long way to go in understanding how best to design qualifications and qualification systems that are fit for the wide variety of purposes, cohorts, and contexts that characterise the VTQ landscape. As discussed in our section on

---

[35] Stevens quoted a City & Guilds report from the mid-1960s, which described this period as: “a new era in which the Institute will no longer itself determine the priorities, but is nevertheless ready and able to play its part” (Stevens, 1993, page 115).

[36] See Keep (2006) for a discussion of similar issues in relation to education and training more generally.

purposeful variation, a research-based learning culture would go a long way towards deepening and embedding expertise and understanding across the sector.

Safer innovation

The second major challenge for the VTQ sector is simpler and shorter to describe: we need to find ways to innovate more safely. The history of VTQ reform in England over the past half century has not been a story of sustained success. This was apparent in our analysis of the origins and evolution of the CASLO approach, but the conclusion extends beyond outcome-based qualification reform (note that most of the qualifications in the following quotation did not adopt the CASLO approach):

> The middle strand, route or track, of general vocational courses has been a long-standing source of concern and has witnessed successive waves of reform as new qualifications have been piloted and found wanting. Examples include the Certificate of Extended Education (CEE), the Certificate of Pre-Vocational Education (CPVE), General National Vocational Qualifications (GNVQs), the Advanced Vocational Certificate of Education (AVCE), Vocational A Levels and then the 14-19 Diplomas.
>
> (Keep, 2012, page 2)

The complexity of the VTQ sector, combined with the relative immaturity of the certification profession in terms of structure and expertise, renders reform of the VTQ landscape an inherently risky business. Reflecting on recommendations from a series of policy reviews – Wolf (2011), Richard (2012), Whitehead (2013), and CAVTL (2013) – Raffe (2015) pondered how urgently England required another radical qualification reform. Cautioning against rushing too soon into another round, he suggested that:

1. while the criticisms undoubtedly deserved scrutiny, they did not apply universally, many were contested, and even the genuine ones were sometimes exaggerated
2. many of the alleged qualification weaknesses were actually symptomatic of deeper problems with education and training and its social and labour-market context (so reforming qualifications alone may end up having little effect)
3. by rushing into reform, we risk repeating mistakes of the past

These observations resonate with our own conclusions concerning the history of VTQ reform in England.

In report 4, we concluded that an important part of safer innovation was extensive piloting and trialling. Scaling up too soon can cause serious reputational damage to the brand image of a new or reformed qualification, particularly when things go wrong on a national scale, which was true for both NVQs and GNVQs. In the words

of Bent Flyvbjerg and Dan Gardner – think slow, act fast. It is far better to make mistakes in the planning stage than during delivery (Flyvbjerg & Gardner, 2023).

Our earlier discussion of purposeful variation is consistent with the aspiration for safer innovation in at least 3 respects. First, it requires a persuasive rationale for variation, based on an analysis of fitness for purpose. This, in itself, should help to reduce the likelihood of any particular innovation failing. Second, to the extent that innovation of this sort is likely to operate on a smaller scale, this will reduce the scale of harm if an innovation does fail. Third, to the extent that it permits innovation on multiple fronts simultaneously, it provides a steady stream of evidence on the conditions associated with both success and failure, in support of a learning culture.

Stronger structure

The final major challenge for the sector is to co-ordinate better the roles that different actors play in relation to qualification (and qualification system) design, development, delivery, and evaluation. Our research identified different variants of this challenge.

The first variant is: how to achieve as much consensus as possible concerning a suitable direction of travel for the VTQ sector, to secure as much co-ordinated engagement as possible. Reforms of the 1980s and 1990s were confrontational in the sense of intending to disrupt the old order and attempting to impose a new one. Yet, by their very nature, qualifications constitute a communal resource, and they require general assent for their effective operation, hence the idea of a certification community (as discussed earlier).[37] Some within the VTQ sector, including certain large employers, responded to the imposition of change during the 1980s and 1990s by opting out of the national system entirely. Others, including many education scholars, responded by adopting explicitly contrary, adversarial positions. There would seem to be plenty of bridges to rebuild.

The second variant is: how to make sure that all members of the certification community for any particular qualification are playing to their strengths and contributing at the right time. This challenge is exacerbated by the lack of an overarching integrated theory of educational certification, which makes it harder to specify definitively what sorts of strengths are required at any particular stage of the qualification lifecycle. We introduced the idea of a 3-stage framework earlier, as a step in this direction: first identify a purpose-cohort-context profile, then construct a proficiency model, and then design an assessment procedure. This helped us to emphasise 2 points, which both relate to each of the 3 stages:

[37] Note, for instance, how learners can withhold their assent by choosing not to study a qualification, and how users can withhold their assent by failing to use results in the manner intended by the designer. These matters tend to lie beyond the control of policy makers.

1. although each stage may be convened by a particular stakeholder or group of stakeholders (for instance, an awarding organisation at stage 3), it will always require input from a variety of partners from across the certification community
2. each stage will also require input from technical specialists of one sort or another (for instance, stage 1 will need to be guided by an expert in purpose analysis), who will need to work in synergy with the other participants

To provide a more detailed example, stage 2, which involves constructing a proficiency model, might be led by an employer group, assuming that it had been decided that the qualification under development would need to be especially attuned to employer requirements. However, it would also be necessary for the employer group to work in partnership with an expert in proficiency modelling, to ensure that their needs were effectively translated into an appropriate and adequate proficiency model. Furthermore, the group would also require input from teachers and trainers to ensure that the model provided a suitable basis for subsequently developing teaching and training programmes, and it would require input from assessment experts to ensure that the model provided a suitable basis for developing assessment procedures and plans.

The proposal that all 3 stages incorporate technical activities holds important implications for training and development in the sector. It is generally recognised that the sector requires individuals with expertise in assessment, although suitably qualified professionals tend to be in short supply. The demand for individuals with expertise in purpose analysis or proficiency modelling is less well understood, though, and it seems likely that professionals of this sort are in even shorter supply.

A different take on the challenge of participants playing to their strengths is to make sure that the system is designed to accommodate any acknowledged weaknesses. For instance, where circumstances are such that teachers and trainers may struggle to develop adequate teaching and learning programmes, this begs an important question concerning the role of awarding organisations or other agencies in supporting them to do so.

The final variant is: how to embed qualification reform holistically. In report 4, we concluded that the most important risk highlighted by our research is that qualification reforms are conceptualised and operationalised too narrowly, with insufficient attention to the wider education and training changes that are necessary for those reforms to bed in. We proposed that qualification reforms are best understood as education and training reforms that are initiated through changes to certification requirements. When considered from this perspective, the importance of adequately involving teachers, trainers, centres, learners, and users right from the outset becomes harder to overlook. Equally hard to overlook, though, is the necessity of providing adequate support, training, and guidance for those responsible for rolling out the reforms, especially in relation to teaching or training.

These points resonate strongly with the idea of anticipatory qualification design: anticipating as thoroughly as possible the likely determinants of successful and unsuccessful implementation, and laying the groundwork accordingly.

References

Addis, M. & Winch, C. (2019). Introduction. In M. Addis & C. Winch (Eds.). *Education and Expertise* (pp.2–20). London: John Wiley & Sons Ltd.

Ahmed, A. & Pollitt, A. (2011). Improving marking quality through a taxonomy of mark schemes. *Assessment in Education: Principles, Policy & Practice*, 18(3), 259–278.

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.

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). New York: Longman.

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). The Netherlands: Springer.

Biggs, J. & Collis, K.F. (1982). *Evaluating the Quality of Learning: The SOLO taxonomy*. New York: Academic Press.

Biggs, J., Tang, C. & Kennedy, G. (2022). *Teaching for Quality Learning at University* (5th edition). Maidenhead: Open University Press McGraw Hill.

Blank, W.E. (1982). The competency-based approach to education and training. In M. Brown, F. Patrick, D. Tate, & S. Wright (1994) (Eds.). *A Collection of Readings Related to Competency-Based Training* (pp.32–53). Australia: Deakin University Faculty of Education.

Bloom, B.S., Englehart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of Educational Objectives: The classification of educational goals. Handbook I: Cognitive domain*. London: Longmans, Green and Co Ltd.

Braun, M. (2019). A pilot study investigating further education engineering lecturers' readiness for the shift in Business and Technology Education Council (BTEC) assessment. *Research in Education*, 105(1), 42–59.

Brennan, R.L. (2006) (Ed.). *Educational Measurement* (4th edition). Westport, CT: American Council on Education and Praeger Publishers.

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

Care, E., Kim, H., Vista, A., & Anderson, K. (2018). *Education System Alignment for 21st Century Skills: Focus on assessment*. Washington, DC: Center for Universal Education at the Brookings Institution.

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.

Cedefop (2023). *Microcredentials for Labour Market Education and Training: The added value for end users. Luxembourg*: Publications Office of the European Union.

Cedefop (2024a). *Learning Outcomes Going Global: A multifaceted phenomenon*. Thessaloniki, Greece: European Centre for the Development of Vocational Training.

Cedefop (2024b). The Influence of Learning Outcomes on Pedagogical Theory and Tools. Luxembourg: Publications Office of the European Union.

Child, S.F.J. & Shaw, S.D. (2020). A purpose-led approach towards the development of competency frameworks. *Journal of Further and Higher Education*, 44(8), 1143–1156.

City & Guilds (2014). *Sense and Instability: Three decades of skills and employment policy*. London: City & Guilds Group.

City & Guilds (2016). *Sense and Instability 2016*. London: City & Guilds Group.

City & Guilds (2019). *Sense and Instability 2019*. London: City & Guilds Group.

Clauser, A.L. & Raymond, M. (2017). Specifying the content of credentialing examinations. In S. Davis-Becker & C.W. Buckendahl (Eds.). *Testing in the Professions: Credentialing policies and practices* (pp.64–84). New York: Routledge.

Dearing, R. (1996). *Review of Qualifications for 16 to 19 Year Olds: Full report*. London: School Curriculum and Assessment Authority.

Dreyfus, S.E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science, Technology & Societ*y, 24(3), 177–181.

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

Eraut, M. (2004). A wider perspective on assessment. *Medical Education*, 38(8), 803–804.

Eraut, M. (2008). Chapter A2 How Professionals Learn through Work. Learning to be Professional through a Higher Education E-BOOK.

Faraday, S., Overton, C., & Cooper, S. (2012). *Effective Teaching and Learning in Vocational Education*. London: Learning and Skills Network.

Federation of Awarding Bodies (2017). *The Qualification Manager's Handbook*. London: Federation of Awarding Bodies.

Flyvbjerg, B. & Gardner, D. (2023). *How Big Things Get Done: The surprising factors behind every successful project, from home restoration to space exploration*. London: Macmillan.

Fullan, M. (2006). *Change Theory: A force for school improvement. Seminar series paper no. 157*. Victoria, Australia: Centre for Strategic Education.

Gallacher, T. & Johnson, M. (2019). "Learning Progressions": A historical and theoretical discussion. *Research Matters: A Cambridge Assessment publication*, 28, 10–16.

Gravells, A. (2016). *Principles & Practices of Quality Assurance*. London: SAGE Learning Matters.

Hart, J. Noack, M., Plaimauer, C. & Jens Bjørnåvold, J. (2022). *Towards a Structured and Consistent Terminology on Transversal Skills and Competences*. ESCO Publications.

Keep, E. (2006). State control of the English education and training system – playing with the biggest train set in the world. *Journal of Vocational Education and Training*, 58(1), 47–64.

Keep, E. (2012). *Reforming Vocational Qualifications: Some problems ahead?* SKOPE Issues Paper 28. Cardiff: Cardiff University.

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

Koretz, D. (2017). *The Testing Charade: Pretending to make schools better*. Chicago, IL: The University of Chicago Press.

Lang, J. (1978). *City and Guilds of London Institute: Centenary 1878–1978*. London: City and Guilds of London Institute.

Lenon, B. (2018). *Other People's Children: What happens to those in the bottom 50% academically?* Woodbridge: John Catt Educational Ltd.

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

Lester, S. (2015). Competence standards and frameworks: Some lessons from the United Kingdom. Final version published in Edukacja ustawiczna Dorosłych (Polish Journal of Continuing Education), 90(3), 132–141.

Lester, S. (2017). *Professional Competence Standards: Guide to concepts and development*. Taunton, UK: The ComProCom Partnership.

Lester, S., Koniotaki, A., & Religa, J. (2018). ComProCom: a revised model of occupational competence. *Education + Training*, 60(4), 290–302.

Looney, J. (2011). *Alignment in Complex Education Systems: Achieving balance and coherence*. OECD Education Working Papers No. 64. Paris: Organisation for Economic Co-operation and Development.

Lucas, B., Spencer, E., & Claxton, G. (2012). *How to Teach Vocational Education: A theory of vocational pedagogy*. London: City & Guilds Centre for Skills Development.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, 21(3), 215–237.

Messick, S. (1989). Validity. In R. Linn (Ed.). *Educational Measurement* (3rd edition) (pp.13–100). Washington, DC: American Council on Education.

Mislevy, R.J. (1993). Foundations of a new test theory. In N. Frederiksen, R.J. Mislevy, & I. Bejar (Eds.). *Test Theory for a New Generation of Tests* (pp.19–39). Hillsdale, NJ: Lawrence Erlbaum.

Moseley, D., Baumfield, V., Elliott, J., Gregson, M., Higgins, S., Miller, J., & Newton, D. (2005). *Frameworks for Thinking: A handbook for teaching and learning*. Cambridge: Cambridge University Press.

Mulder, M. (2017a). Competence and the alignment of education and work. In M. Mulder (Ed.). *Competence-based Vocational and Professional Education: Bridging the world of work and education* (pp.229–251). Switzerland: Springer International Publishing.

Mulder, M. (2017b) (Ed.). *Competence-based Vocational and Professional Education: Bridging the world of work and education*. Switzerland: Springer International Publishing.

Newton, P.E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170.

Newton, P.E. (2012). Clarifying the consensus definition of validity. *Measurement: Interdisciplinary Research and Perspectives*, 10(1–2), 1–29.

Newton, P.E. (2017a). There is more to educational measurement than measuring: The importance of embracing purpose pluralism. *Educational Measurement: Issues and Practice*, 36(2), 5–15.

Newton, P.E. (2017b). *An Approach to Understanding Validation Arguments*. Ofqual/17/6293. Coventry: Office of Qualifications and Examinations Regulation.

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

Newton, P.E. (2020). Deficiency, contamination, and the signal processing metaphor. *Educational Measurement: Issues and Practice*, 39(1), 82–91.

Newton, P.E. (2023a). Taking purpose into account when designing, using and evaluating educational assessments. In R.J. Tierney, F. Rizvi, & K. Ercikan (Eds.). *International Encyclopedia of Education*, 4th edition, volume 13 (pp.136-146). Elsevier.

Newton, P.E. (2023b). Certificating learning on the basis of teacher assessment. In C. Harrison, C. Leung, & D. Pepper (Eds.). *Educational Assessment: The influence of Paul Black on research, pedagogy and practice* (pp.221–240). London: Bloomsbury Academic.

Oates, T. (2010). *Could Do Better: Using international comparisons to refine the National Curriculum in England*. Cambridge: Cambridge Assessment.

Oates, T. (2013). *Education, Assessment and Developing a Comprehensive Qualifications Framework*. Cambridge: Cambridge Assessment.

Ofqual (2012). *GCSE English 2012*. Ofqual/12/5225. Coventry: Office of Qualifications and Examinations Regulation.

Ollin, R. & Tucker, J. (2012). *The Vocational Assessor Handbook* (5th edition). London: Kogan Page.

Pellegrino, J.W., Chudowsky, N. & Glaser, R. (2001). *Knowing What Students Know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Perie, M. & Huff, K. (2016). Determining content and cognitive demand for achievement tests. In S. Lane, M.R. Raymond, & T.M. Haladyna (Eds.). *Handbook of Test Development* (2nd edition) (pp.119–143). New York, NY: Routledge.

Piper, P., Zuilkowski, S.S., Dubeck, M., Jepkemei, E. & King, S.J. (2018). Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers' guides. *World Development*, 106 (C), 324–336.

Pollitt, A. Ahmed, A. & Crisp, V. (2008). The demands of examination syllabuses and question papers. In P.E. Newton, J. Baird, H. Goldstein, H. Patrick, & P.

Tymms (Eds.). *Techniques for Monitoring the Comparability of Examination Standards* (pp.166–206). London: Qualifications and Curriculum Authority.

Priestley, M. (2017). *Approaches to Specifying Curriculum Areas of Learning. Paper prepared for the groups developing the Welsh Successful Futures AoLEs*. Stirling: The Stirling Network for Curriculum Studies.

Pring, R.A. (1995). *Closing the Gap: Liberal education and vocational preparation.* London: Hodder and Stoughton Educational.

QCA (Qualifications and Curriculum Authority) (2001). *NVQ Code of Practice*. QCA/02/875. London: Qualifications and Curriculum Authority.

QCDA (Qualifications and Curriculum Development Agency) (2010). *Key Questions about Assessment in the QCF*. QCDA/10/5202. Coventry: Qualifications and Curriculum Development Agency.

Raffe, D. (2015). First count to five: Some principles for the reform of vocational qualifications in England. *Journal of Education and Work*, 28(2), 147–164.

Read, H. (2012). *The Best Quality Assurer's Guide*. Bideford: Read On Publications.

Reeve, W.D. (1925). Objectives in the teaching of mathematics. *The Mathematics Teacher*, 18(7), 385–405.

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

Sadler, D.R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13(2), 191–209.

Sanchez, J.I. & Levine, E.L. (2012). The rise and fall of job analysis and the future of work analysis. *Annual Review of Psychology*, 63, 397–425.

Schmeiser, C.B. & Welch, C.J. (2006). Test development. In R.L. Brennan (Ed.). *Educational Measurement* (4th edition) (pp.307–353). Westport, CT: American Council on Education and Praeger Publishers.

Schools Council (1973). *Assessment of Attainment in Sixth-form Science*. Schools Council Examinations Bulletin 27. London: Evans Methuen Educational.

Shepard, L.A. (2018). Learning progressions as tools for assessment and learning. *Applied Measurement in Education*, 31(2), 165–174.

Slaney, K.A. & Garcia, D.A. (2015). Constructing psychological objects: The rhetoric of constructs. *Journal of Theoretical and Philosophical Psychology*, 35(4), 244–259.

Stanley, J. (2012). *Curriculum Reform in Europe: The impact of learning outcomes*. Cedefop Research Paper no. 29. Luxembourg: Publications Office of the European Union.

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.

Stanton, G. (2016). Further education and the case for vocational preparation. In. R. Pring & M. Roberts (Eds.). *A Generation of Radical Educational Change: Stories from the field* (pp.107–121). Oxford: Routledge.

Stevens, P. (1993). *City and Guilds of London Institute: A short history, 1878–1992*. London: City and Guilds of London Institute.

Suto, I., Greatorex, J., Vitello, S., & Child, S.F.J. (2020). A way of using taxonomies to demonstrate that applied qualifications and curricula cover multiple domains of knowledge. *Research Matters: A Cambridge Assessment publication*, 30, 26–34.

Tyler, R.W. (1931). A generalized technique for conducting achievement tests. *Educational Research Bulletin*, 10(8), 199–208.

Tyler, R.W. (1932a). Measuring the results of college instruction. *Educational Research Bulletin*, 11(10), 253–260.

Tyler, R.W. (1932b). Improving test materials in the social studies. *Educational Research Bulletin*, 11(14), 373–379.

van der Vleuten, C.P.M. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education*, 1(1), 41–67.

Vitello, S., Greatorex, J. & Shaw, S.D. (2021). *What is Competence? A shared interpretation of competence to support teaching, learning and assessment*. Cambridge: Cambridge University Press & Assessment.

Watson, A. (1991). Competency-based vocational education: Is this the answer? *The Vocational Aspect of Education*, 43(1), 133–145.

Wheelahan, L. & Moodie, G. (2018). What should vocational qualifications look like if the links between qualifications and jobs are so weak? In S. Allais & Y. Shalem (Eds.). *Knowledge, Curriculum, and Preparation for Work* (pp.127–146). E-Book: Brill.

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.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal for Research in Science Teaching*, 46(6), 716–730.

Winch, C. (2012). Vocational and civic education: Whither British policy? *Journal of Philosophy of Education*, 46(4), 603–618.

Winch, C. (2013). Curriculum design and epistemic ascent. *Journal of Philosophy of Education*, 47(1), 128–146.

Winch, C. (2014). Know-how and knowledge in the professional curriculum. In M. Young & J. Muller (Eds.). *Knowledge, Expertise and the Professions* (pp.103–128). London: Routledge.

Winch, C. (2015). Towards a framework for professional curriculum design. *Journal of Education and Work*, 28(2), 165–186.

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. (1995). *Competence-Based Assessment*. Maidenhead: Open University Press.

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

Wood, R. & Power, C. (1987). Aspects of the competence–performance distinction: Educational, psychological and measurement issues. *Journal of Curriculum Studies*, 19(5), 409–424.

Wood, R. (1968). Objectives in the teaching of mathematics. *Educational Research*, 10(2), 83–98.

Wood, R. (1993). *Assessment and Testing: A survey of research*. Cambridge: Cambridge University Press.

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/24/7176**