

Research and Analysis

Grading Vocational & Technical Qualifications

Recent policies and current practices

Paul E. Newton from Ofqual's Strategy, Risk and Research
directorates

Acknowledgements

I am very grateful to everyone who has played a role in supporting this project. I would particularly like to thank colleagues from awarding organisations who generously agreed to quality assure this report, and who provided important insights into the workings of their qualifications: Denise Sedgwick (ABC Awards), Lena Gray (AQA), Colleen Chater (ATHE), Kerry Lawlor (CIBTAC), Vanessa Lee and Jason Embery (IMI), Eleanor Maughan and Paul Kelly (NCFE), Paul Johnson and Sara Furness (NOCN), Rosanne Hill (and colleagues from OCR), Lizzie Watts (and colleagues from Pearson), Tim Bennett-Hart and Daniel Latham (RSL), Richard Burton (RSPH), Julia Martin (TCL), Lizzie Moore (UWLQ), Carina Fagan (VTCT) Patrick Craven (and colleagues from City & Guilds), and Rachael Meech. I would also like to thank colleagues from Ofqual, particularly Phil Carr and Michelle Meadows, and Ofqual's Standards Advisory Group.

Contents

Executive summary	1
Background	5
Recent policy context	8
Current practice in qualification regulation	10
<i>The QCF</i>	10
<i>The GCR</i>	19
Current practice in qualification design	23
<i>Selecting a sample of qualifications</i>	24
<i>Results from the exploratory survey</i>	26
<i>Theme 1: Measurement models</i>	29
<i>Theme 2: Measurement standards</i>	52
Discussion	63
<i>Grading CBAs</i>	63
<i>Technical issues in grading VTQs</i>	65
<i>Conceptual issues in grading VTQs</i>	73
<i>Further research</i>	78
References	80

Executive summary

This report presents outcomes from a small-scale, exploratory research project into grading practices within regulated Vocational and Technical Qualifications (VTQs) in England. Its aim was to explore the variety of approaches in operation, and to consider issues arising. It was based upon 18 qualifications, from 15 Awarding Organisations (AOs), sampled from the 2167 VTQs that were available to be studied by learners, as at 28/11/17, which were classified on the Ofqual database as awarding higher grades (ie grades above the passing grade). The variety of grading approaches observed, across just these 18 qualifications, was striking. It seems reasonable to conclude, on the basis of this evidence alone, that current practice in grading VTQs in England is not underpinned by a straightforward, generally accepted, set of principles governing good practice. Conversations with colleagues from each of the AOs represented in the sample supported this conclusion. That does not imply that any of the observed grading approaches reflects anything less than good practice. A fairer conclusion is that it remains far from obvious what set of principles ought to underpin grading practices for VTQs in England; and, in particular, whether the same principles ought to underpin grading practices across such a wide variety of qualifications. The present report represents a first step in a programme of work that will be necessary to understand in more depth what good practice in grading VTQs looks like.

There are plenty of precedents for grading VTQs, both historically and internationally. However, in England, under the influence of Competence-Based Assessment (CBA), such practices waned. CBA took root in England, in the 1990s, with the introduction of National Vocational Qualifications (NVQs), as a direct result of dissatisfaction with the traditional General Qualification (GQ) assessment approach when applied in vocational settings. The hallmarks of CBA are very distinctive, including design requirements such as:

- the atomistic specification of measurement standards in terms of learning outcomes and assessment criteria;
- a mastery measurement model, meaning that a certificate of competence could be interpreted to mean competent across each and every learning outcome and assessment criterion;
- assessment based on the exhaustive sampling of learning outcomes and assessment criteria.

Qualifications designed according to CBA will articulate, for each of the units that comprise the qualification, a list of discrete Learning Outcomes (LOs), and a list of discrete Assessment Criteria (AC) for each LO – hence atomistic specification. They will also assume that overall competence implies competence in each one of these AC, for each LO, for each unit – hence mastery. Therefore, to confirm competence, candidates are generally assessed in relation to every single AC, for each LO, for each unit – hence exhaustive sampling.

During the 2000s, the Qualifications and Curriculum Authority (QCA) designed and developed a new approach to regulating VTQs, based on the Qualifications and Credit

Framework (QCF). Although QCA hoped that all qualifications would eventually become part of the QCF, its 'entry requirements' were very restrictive, in the sense that it was fundamentally premised upon a one-size-fits-all model of qualification design. Indeed, its design requirements included the three hallmarks of CBA presented above. Although GQs were never recognised under the QCF, many VTQs were. The QCF was withdrawn in 2015, but even today it seems likely that the core characteristics of CBA continue to operate as something of a default template for designing (many although not all) VTQs in England.

As originally conceived, CBA has no role for grading; because it focuses purely upon whether or not a learner has reached a minimum competence threshold of proficiency. When they have reached that threshold, they can be certificated. The idea of certificating higher levels of proficiency, above the minimum competence threshold, simply does not arise. Having said that, many VTQs were graded prior to the introduction of the QCF, and the QCF did recognise the possibility of grading, even in relation to qualifications built upon the CBA model.

In recent years, following the publication of the Wolf Review, the Richard Review, and the Whitehead Review, grading has become a political imperative; particularly for VTQs that are to be recognised in school and college performance tables; as well as for new Apprenticeships, and for new Technical Qualifications. This underscores the importance of reaching an in-depth understanding of good practice in relation to grading VTQs. Of the 17728 regulated qualifications that were available for learners to study, as at 28/11/2017, 15852 were not GQs, thereby counting as VTQs for the purpose of the present research. Of these, 13.7% awarded higher grades, above the passing grade, thereby constituting the population of interest for the purpose of the present research. Aiming to identify as wide a range of grading practices as possible – bearing in mind the limitations of a small-scale, exploratory research project – documentary evidence for 18 qualifications was downloaded from the internet, to understand in detail the approaches that were being operated.

A wide variety of practices was identified. Yet, important similarities were also identified, which made it possible to propose two distinct themes according to which the qualifications could be classified:

1. measurement model – the approach that a qualification adopts to aggregating assessment information; where that information might exist in the form of task marks/scores or criterion judgements (which are aggregated within units to determine the unit grade), or in the form of unit marks or unit grade points or unit grade profiles (which are aggregated across units to determine the qualification grade); and
2. measurement standard(s) – the ultimate criterion (or criteria) against which candidates are to be judged; as specified for the passing grade and for higher grades.

Not surprisingly, given the influence of the QCF on qualification design in England, many of the sampled qualifications incorporated features and processes that resonated with the core characteristics of CBA; including atomistic specification, domain mastery, and exhaustive sampling. Not all, though. More importantly, those

qualifications that did superficially resemble CBAs also incorporated a variety of adaptations; apparently designed to preserve their own particular ethos, purpose and value, as well as simply to make grading work.

It was clear, for instance, that mastery was not the only measurement model in operation. In fact, four different aggregatory principles were observed:

1. mastery – overall result represents (or tends towards) the lowest level of proficiency across a specified domain, or subdomain;
2. compensation – overall result represents an average level of proficiency across a specified domain, or subdomain;
3. configuration – overall result represents a particular pattern, or configuration, of proficiencies across a specified domain, or subdomain; and
4. charity – overall result represents (or tends towards) the highest level of proficiency across a specified domain, or subdomain.

All of the sampled qualifications operated aspects of both mastery and compensation. Many of the qualifications operated at least three of these principles simultaneously, and some all four.

Similarly, a variety of different approaches to specifying measurement standards was observed. Criterion-related approaches tended to dominate, which were subdivided into 'performance quality' and 'performance complexity' approaches. Criterion-related approaches might be thought of as grading-for-CBAs. As noted earlier, CBA is characterised by the atomistic specification of measurement standards – which are defined in terms of lists of AC, nested within LOs, nested within units – where those standards are specified only for the passing grade. Transforming this model into a grading approach involves specifying higher-level criteria; either for each and every AC, or for a restricted set of AC. In effect, this transforms a list of AC for the passing grade into a two-dimensional grading grid; with criteria for higher-level grades, typically Merit and Distinction, appended to the passing criteria. Thus, measurement standards are defined, on a criterion-by-criterion basis, in terms of a simple set of statements; typically, three statements – one for Pass, one for Merit, and one for Distinction. Often, although not always, candidates needed to achieve all of the Merit criteria within a unit to be awarded Merit overall; and likewise for Distinction.

The difference between performance complexity and performance quality approaches relates to the content of the three statements for each criterion. Performance quality approaches tended to frame criterion statements in terms of increasing quality of performance, typically holding complexity constant across the statements/grades. For instance, if the Pass criterion statement for a particular LO specified complexity in terms of a command verb like 'explain' (eg explain the marketing strategy...), then the Merit criterion statement might specify that same command verb at a higher level of quality (eg explain with clarity and detail...), and the Distinction criterion statement might specify the same command verb at an even higher level of quality (eg comprehensively explain...). Differentiation between grades is therefore achieved on the basis of the quality of candidate performances in relation to criterion statements.

Performance complexity approaches, conversely, tended to frame criterion statements in terms of increasing complexity, indicated by the use of increasingly challenging command verbs, derived from Bloom's Taxonomy. For instance, if the Pass criterion statement for a particular LO specified that it must be 'described' (eg describe the marketing strategy), then the Merit criterion statement might specify a higher-level command verb (eg explain the marketing strategy), and the Distinction criterion statement might specify an even higher-level command verb (eg critically evaluate the marketing strategy). In this case, differentiation between grades is achieved on the basis of the cognitive complexity of candidate performances in relation to criterion statements.

As an exploratory research project, the primary aim was simply to provide insight into the variety of grading practices in operation; and this was achieved by classifying the 18 sampled qualifications in relation to two key themes – measurement standards, and measurement models. The project also aimed to consider what issues might arise in relation to the different approaches; and these were divided into two categories – short-term technical issues, and longer-term conceptual ones. Technical issues related to the effective operation of the kinds of practices observed across the 18 sampled qualifications. Questions were raised concerning standardisation, grading and levelling, comparability, weighting, burden and backwash, and transparency. Conceptual issues related to underpinning assessment models. In particular, a fundamental question was raised concerning the appropriateness and/or adequacy of CBA as a default template upon which to design and develop graded VTQs. The present research simply raises such questions. It remains for further research, analysis, and dialogue within the VTQ community to explore answers.

Background

It is standard practice for General Qualifications (GQs) in England to award **higher grades**, above the **passing grade**, to recognise higher levels of proficiency in a domain. Vocational and Technical Qualifications (VTQs), on the other hand, tend not to award higher grades, particularly where those qualifications are deemed to certify competence, ie the ability to practise competently in a domain. In such contexts, it is often assumed that VTQ candidates ought simply to be classified as having achieved (or not yet having achieved) a real-world minimum competence standard.¹

Increasingly, however, stakeholders have questioned whether grading ought to feature more prominently within VTQs. From one perspective, higher grades have the potential to provide qualification result users (eg employers, higher education selectors) with additional measurement information, enabling them to make better decisions (eg selection decisions). From another perspective, higher grades have the potential to motivate learners to achieve higher levels of proficiency, by raising their aspirations. Indeed, government policy now requires that certain kinds of qualification must be graded, eg where those qualifications are to be included in school and college performance table calculations.

The present report provides a survey of recent policies and current practices related to the grading of VTQs in England, identifying and exploring a variety of technical and conceptual issues. It adopts a very broad definition of 'vocational and technical' to include all Ofqual-regulated qualifications that are not classified as GQs.

The motivation for this work arose from pilot research within a broader programme, entitled *Developing Proficiency Specifications*. At the heart of this programme is the idea that qualification design is most effective when it is referenced to a proficiency specification (see Newton, 2017). Qualification design in England has traditionally been understood as a process of 'assessing the curriculum' and has therefore traditionally been referenced to a **curriculum specification**, which emphasises the centrality of curriculum coverage to effective assessment. The passing grade, from this perspective, distinguishes those who have learned 'enough' of the curriculum from those who have not; while higher grades distinguish between those who have learned 'enough' and those who have learned increasingly 'more' of the curriculum. However, exactly what characterises candidates at different grades is a secondary concern, from this perspective. Indeed, candidates might even be classified into grades fairly arbitrarily, eg by awarding a grade A to the top 10%, B to the next 10%, and so on, providing little indication of what candidates in each grade actually know, understand and can do.

¹ Even in England, this was not always so. Prior to the 1990s, it was not uncommon for VTQs to be graded. The practice of grading began to wane somewhat with the introduction of National Vocational Qualifications.

A more recent perspective treats qualification design as a process of ‘measuring the learner’ with (more or less explicit) reference to a **proficiency specification**. Developing a proficiency specification is not simply a matter of characterising the proficiency **construct**, ie the boundaries of the content domain, and the elements of knowledge and skill that it includes. It is also a matter of characterising the proficiency **scale**, ie the features that distinguish different levels of proficiency. Consequently, the **proficiency construct-scale** explains what it means for one candidate to have achieved a higher level of proficiency than another. It is therefore obviously relevant to qualifications that award higher grades. Yet, it is equally relevant to qualifications that only award a passing grade, because that still requires us to specify what distinguishes a candidate who is minimally competent from a candidate who is not yet competent.

The introduction of National Vocational Qualifications (NVQs) in England, during the 1990s, reflected a shift in perspective from ‘assessing the curriculum’ to ‘measuring the learner’. NVQs were conceptualised *purely* in terms of measuring learner competence, regardless of whether or not that learner had followed a formal course of learning, bounded by a curriculum. Instead of a curriculum, NVQs specified the **elements of competence** that comprised a domain of practice (thereby articulating the construct), alongside **performance criteria** that were used to judge whether or not each element had been achieved (thereby articulating the scale). The model underlying this shift in perspective – which was known as **Competence-Based Assessment** (CBA) – generalised beyond the NVQ context to influence the design of VTQs (and qualification frameworks) more generally; not just in England, but in countries across Europe and further afield, eg Australia.²

Nowadays, standards for many, if not most, VTQs in England are specified, typically at the unit level, via lists of elements of competence and performance criteria – now generally referred to as **Learning Outcomes** (LOs) and **Assessment Criteria** (AC) – and a Pass is typically contingent upon all of these criteria having been satisfied, ie it requires **mastery** of all LOs. This is often facilitated by a heavy (often exclusive) reliance upon internal assessment; whereby, as the learner gradually accumulates evidence to satisfy those criteria, so each criterion can be ‘ticked off’ the list.

In contrast, the GQ model is quite different. The passing grade (and higher grades) are specified far less precisely. Indeed, the primary specification is essentially by implication: the passing standard this year represents the same level of proficiency as it represented last year. Nowadays, grade descriptions are produced to give a rough indication of the nature of grading standards, and therefore of the proficiency scale. They are provided more for qualification users, to help them to interpret qualification outcomes accurately, than to support grading. Importantly, they are ‘best

² A glossary has been prepared, as a point of reference for certain of the novel, or less familiar, technical terms used in the present report, such as Competence-Based Assessment. The glossary is published as a supplementary document.

fit' descriptions, since there is no expectation within GQs that precisely defined criteria need to be satisfied for the award of any grade, ie they are **compensatory**, meaning that a high level of proficiency on one outcome can compensate for a low level of proficiency on another.³

As awarding organisations (AOs) respond to government policy on incorporating higher grades into VTQs, as well as to government policy on incorporating more external assessment, this raises important questions concerning the specification of proficiency construct-scales, and concerning the effectiveness of grading approaches under different measurement models. In particular, it raises a major question concerning the extent to which grading is compatible with a mastery model of domain competence.

The present report is not an evaluation of the arguments for and against incorporating higher grades into VTQs. It is an exploratory survey of grading practices, focused upon the small proportion of regulated VTQs in England that do award higher grades. It aims to explore the variety of grading practices in operation, and to consider issues arising.

As this survey was being undertaken, it became clear that **grading** and **levelling** are closely related practices; at least, as grading is practised by many VTQ AOs. Levelling is the process of associating a qualification with a particular level on (what is nowadays called) the Regulated Qualifications Framework (RQF).⁴ Both grading and levelling are ways of differentiating levels of proficiency. Where necessary, the present report considers issues of levelling and grading in parallel.

Finally, it is important to note that this is a research report (ie not an audit report) and that it is exploratory in nature (ie neither comprehensive nor definitive). Although it identifies practices related to specific, regulated qualifications – on the basis of information publicly available on the internet – it is not concerned with issues of compliance.⁵ It is intended to foreground more general issues related to VTQ design, on the basis of an 'opportunity sample' of qualifications, drawn in order to identify as wide a variety of practices as possible, within the limitations of an exploratory survey. These issues are examined through the general lens of validity, rather than in relation

³ The present report uses three terms – proficiency, competence, and attainment – in essentially the same way, to refer the general attribute that a qualification aims to measure. Note that any particular qualification will measure a specific attribute, eg proficiency in nursing, competence in bricklaying, attainment in mathematics. The term 'competence' is more common in VTQ contexts (where its implications have been heavily debated, incidentally). The term 'attainment' is more common in GQ contexts. The term 'proficiency' is used, in the present report, to imply that we are talking about essentially the same thing, regardless of context.

⁴ See <https://www.gov.uk/what-different-qualification-levels-mean/list-of-qualification-levels>

⁵ The decision to base the research on information that was publicly available on the internet was intended to minimise the regulatory burden of this project. There was no reason to conclude that this strategy might reduce the quality of the research. More importantly, the AOs whose qualifications were sampled were able to confirm the accuracy of reported information, and to provide additional information where necessary.

to specific regulatory requirements from Ofqual's *General Conditions of Recognition* or related documents.

Recent policy context

Recent debate over assessment practices within VTQs in England has been influenced by a number of high-profile reports, including:

1. Wolf (2011) *Review of Vocational Education*;
2. Richard (2012) *Review of Apprenticeships*; and
3. Whitehead (2013) *Review of Adult Vocational Qualifications in England*.

Each of these reports identified grading as a growing concern:

One fundamental problem for educational institutions is that they are competence-based awards which are meant to attest that someone has reached a particular threshold or level of workplace competence. [...] However, any candidate for educational progression needs to demonstrate not only a specific level of competence, but relative performance, otherwise the qualification is of little use to selectors. While a considerable number of QCF awards do allow for grading, awarding bodies told the review that this had been very difficult to achieve.

(Wolf, 2011, pp.86-7)

Also, similar to a University degree, I believe that the test at the end of an apprenticeship should be graded. Prospective employers should be able to use the grade in the test as evidence of the apprentice's ability and potential.

(Richard, 2012, p.56)

Vocational qualifications provided for adults should be designed to use a pass, merit, distinction structure or a more detailed scale, where such differentiation will increase the qualifications' value to employers and individuals [...] while acknowledging that for some qualifications, including some licences to practise, the costs of introducing grading may outweigh the benefits.

(Whitehead, 2013, p.35)

A subsequent Report from the Independent Panel on Technical Education (Sainsbury, 2016) also mentioned grading, albeit in passing.

Government has responded to these reports by promoting the increased use of grading within VTQs, Apprenticeships, and Tech level (T level) Technical Qualifications (eg, BIS, 2014; DfE, 2017a). Grading has also been specified as a blanket requirement of qualifications submitted to the Department for Education (DfE)

for consideration as Technical Awards, Technical Certificates, Tech Levels or Applied General qualifications, for inclusion in the key stage 4 and 16-to-18 performance tables from 2020 (DfE, 2017b).

Although various regulated VTQs have awarded higher grades for some time now, the recent promotion of grading has already led to greater numbers of VTQs awarding higher grades, and this trend is likely to continue.

Current practice in qualification regulation

Current practice in regulating VTQs needs to be seen in relation to a change in regulatory approach, circa 2014, which led to the Qualifications and Credit Framework (QCF) being withdrawn. Practices that ensured compliance with the detailed requirements of the QCF are likely still to be widespread, despite its withdrawal.

The QCF

The QCF lay at the heart of the approach to regulating VTQs that was inherited by Ofqual from the Qualifications and Curriculum Authority (QCA) in 2008. The aim of the QCF was to rationalise the qualifications systems of England, Wales, and Northern Ireland, by requiring that all qualifications should satisfy a set of regulations that would secure their effective design. These regulations ensured that all qualifications could be located within a common structure – the Framework – which recognised units of specified sizes and levels, that could be combined, according to rules-of-combination, to form qualifications. Most regulated qualifications became part of the QCF, although some remained outside, regulated via alternative mechanisms (eg some NVQs, and all GCSEs and A levels).

The mix-and-match approach to combining units into qualifications meant that units had to be produced according to detailed **design requirements**, which ensured their mutual compatibility and the coherence of the overall Framework. These requirements were specified, by the new regulator, via a single regulatory document – *Regulatory Arrangements for the Qualifications and Credit Framework* (Ofqual, 2008) – which was supplemented by a number of separately published operating rules. The requirements were further explicated within a series of guidance documents, which were published by the organisation that succeeded the QCA, the Qualifications and Curriculum Development Agency (QCDA). These included:

- *Guidance for using unit level descriptors within the Qualifications and Credit Framework*. Version 3. (QCDA, 2008);
- *Guidance for developing rules of combination for the Qualifications and Credit Framework*. Version 4. (QCDA, 2010a);
- *Guidelines for writing credit-based units of assessment for the Qualifications and Credit Framework*. Version 4. (QCDA, 2010b); and
- *Grading Qualifications in the QCF: Guidance for awarding organisations*. (QCDA, 2010c).

Of particular relevance to the present report are the regulations and guidelines that related to levels and grading.

Levels

The *Regulatory Arrangements* document defined the concept of qualification level as an: “indication of the relative demand, complexity and/or depth of achievement, and/or the autonomy of the learner, represented by a qualification” (Ofqual, 2008 p.41). It specified that:

All units must identify a single level for the unit that:

- a represents the complexity, autonomy and/or range of achievement expressed within the unit
- b is determined by comparing the learning outcomes and assessment criteria against the QCF level descriptors (see Annex E)
- c is a constant property of the unit, irrespective of the qualification in which it is located.

(Ofqual, 2008, p.12)

An extract from its Annex E, which presents QCF level descriptors for Level 2 and Level 3, is reproduced in Annex 1 of the present report. Note how each level was described in terms of three constructs (which were prefigured in level indicators for the National Qualifications Framework (NQF), which preceded the QCF, and also in NVQ level indicators):

1. knowledge and understanding;
2. application and action; and
3. autonomy and accountability.

The concept of qualification level was further explicated in the *Guidance for using unit level descriptors within the Qualifications and Credit Framework*, which explained that the descriptors are indicative of achievement at a particular level, but that they do not distinguish performance within a level (QCDA, 2008, p.3). This document recommended the following heuristic for checking the level of a unit (although it emphasised the importance of professional judgement in establishing the level of a unit, rather than strict application of rules or heuristics):

- *if it is principally concerned with knowledge and understanding, then the indicators in this category will be the most important in deciding level – although the way in which the knowledge and understanding is demonstrated and the level of self-direction expected of the learner should be in line with the indicators in the application and action category, and the autonomy and accountability category respectively*
- *if it is principally concerned with occupational competence, then the indicators in the application and action category, and possibly the autonomy and accountability category, are likely to be the most important – although*

the level of knowledge and understanding used should also show a broad match with the relevant indicators, even if knowledge is demonstrated tacitly. (QCDA, 2008, p.4; emphasis added)

The process of writing units to reflect qualification levels was explicated in the *Guidelines for writing credit-based units of assessment for the Qualifications and Credit Framework*. Of particular relevance to the present report is guidance on how to write LOs and AC across levels, for units of the same title that share LOs in common. In such circumstances, the guidance document explained that the burden of distinguishing between the requirements of units at adjacent levels might have to be carried almost exclusively by AC. The document provided an example of a unit titled Customer Service Skills, which was to be offered at both Level 1 and Level 2. Its LOs and AC are reproduced in Annex 2 of the present report. For this (hypothetical) unit, the six LOs were almost exactly the same, except for a minor wording change in the sixth outcome. Similarly, the AC were structurally very similar, covering essentially the same ground; although the Level 2 unit had an additional four criteria (15 criteria at L2, cf 11 criteria at L1). Notice how the distinction between levels is drawn almost exclusively in terms of **behavioural complexity**, ie in terms of the degree of challenge associated with the kind of behaviour required by corresponding AC.⁶ Thus, L1 criteria were specified in terms of less challenging **command verbs**, such as 'identify', 'outline', 'give examples', and 'list'. Whereas, L2 criteria were specified in terms of more challenging command verbs, such as 'describe', 'explain', and 'illustrate'.

The idea of a hierarchy of behavioural complexity derives from Benjamin Bloom's *Taxonomy of Educational Objectives*. Comprising a taxonomy of objectives for the cognitive domain, Bloom, et al (1956) identified six major classes of behaviour, ordered as follows: knowledge (lowest level of complexity), comprehension, application, analysis, synthesis, and evaluation (highest level of complexity).⁷ They explained that these behaviours were integrated within the hierarchy:

Our attempt to arrange educational behaviors from simple to complex was based on the idea that a particular simple behaviour may become integrated with other equally simple behaviors to form a more complex behavior. Thus our classifications may be said to be in the form where behaviors of type A form one class, behaviors of type AB form another class, while behaviors of type ABC form still another class.

(Bloom, et al, 1956, p.18).

⁶ Nowadays, we would tend to describe this as 'cognitive' rather than 'behavioural' complexity. Strictly speaking, we only ever observe behaviour, and mid-twentieth-century educationists, under the influence of behaviourism, tended to respect that technicality. As the influence of behaviourism waned, educationists became happy (once more) to characterise intelligent behaviour cognitively.

⁷ This represented one of three domains identified by Bloom, et al (1956); the other two being the affective domain and the psychomotor domain.

This approach to distinguishing between attainments at adjacent qualification levels was further illustrated by QCDA using a number of examples, such as that presented in Figure 1, below (reproduced from QCDA, 2010b, p.31). In this instance, the L1 criterion started with the command verb 'list' while the L2 criterion started with the verb 'identify' and the L3 criterion started with the verb 'explain' – thus creating a hierarchy of complexity of acquired proficiency. Figure 2 (reproduced from QCDA, 2010b, p.34) explicitly illustrates the (loose) association of command verbs with qualification levels. Note that there is no one-to-one mapping between command verbs and levels (eg 'demonstrate' appears in all four columns; 'define' appears in three columns, from Entry to L2). Importantly, though, certain command verbs do not appear until higher levels (eg 'estimate' in the L2 and L3 columns; 'analyse' and 'evaluate' in the L3 column).

Figure 1. A hierarchy of complexity of acquired proficiency

Unit title: Understanding health and well-being

Learning outcome The learner will: Understand the political and social context of health and well-being			
Level	1	2	3
Assessment criteria	List the government priorities for health promotion and health education	Identify the main points in government policies to improve the effectiveness of the NHS, especially in relation to preventative health care and health education	Explain the government thinking on how to improve the effectiveness of the NHS, especially in relation to preventative health care and health education

Figure 2. The (loose) association of command verbs with qualification levels

Entry level 3	Level 1	Level 2	Level 3
Define	Define	Apply	Analyse
Demonstrate	Demonstrate	Assess	Apply
Give examples	Give examples	Classify	Clarify
Identify	Identify	Compare	Classify
Indicate	Indicate	Define	Critically compare
Locate	Locate	Demonstrate	Demonstrate
Outline	Outline	Describe	Develop plan/idea
State	State	Differentiate	Diagnose
Use	Use	Distinguish	Differentiate
		Estimate	Distinguish
		Give (+/-points)	Draw conclusions
		Illustrate	Estimate
		Perform	Evaluate
		Select	Explain
		Use (a range of...)	Extrapolate
			Implement
			Interpret
			Judge
			Justify
			Perform
			Review and revise
			Summarise

The guidance document was careful to emphasise that command verbs should not be considered definitive of qualification level:

Some verbs are more commonly used in assessment criteria for units at particular levels than at other levels. However, this does not mean that use of one of these verbs in an assessment criterion identifies the unit as being at a particular level, nor does it mean that the use of one of these verbs will necessarily be inappropriate in assessment criteria for units at another level. The meaning of an assessment criterion depends on all of the words in the criterion, not just on the verb. It is the meanings of the assessment criteria in a unit that need to be considered in determining the unit's level, not just the verbs. For example, the word 'identify' could be used at low or high levels, depending on the nature of what has to be 'identified' and the level of cognitive skills required.

(QCDA, 2010b, p.33)

Having said that, it was also very clear from the guidance document that, when LOs are common across levels, there may be little other than command verbs to distinguish between attainments at adjacent qualification levels.

Toward the end of the guidance document, the principle underlying this approach to distinguishing qualification levels was clarified (with reference to Bloom's Taxonomy) and the limitations of this approach at higher levels of the QCF were acknowledged:

Within QCF level descriptors it is possible to detect a gradual shift away from practical skills based on physical actions at lower levels towards more cerebral knowledge based on mental actions as one moves to the higher levels of the QCF. As the basis of the level descriptors reflects a hierarchy of skills and knowledge (Bloom's Taxonomy of Learning is a source for these and other level descriptors) this gradual shift in focus is reflected as you move up the levels of the QCF.

One possible impact of this shifting focus across different levels is that it becomes more difficult to develop precise and easily measurable learning outcomes and assessment criteria at higher levels of achievement. There is a danger that assessment criteria at higher levels either become repetitive, or that they fail to establish an explicit assessment standard for the unit. There are now examples of higher level units in the QCF databank that use a variety of different assessment criteria to establish a clear assessment standard for the unit.

(QCDA, 2010b, p.58)

Grading

As far as the QCF was concerned, addressing the need to differentiate between learners with differing levels of proficiency was primarily a matter of developing units (and qualifications) at different levels. Thus, competent at L3 would represent a higher level of proficiency than competent at L2, where only the passing grade was defined at each level. However, the *Regulatory Arrangements* did allow for differentiation within qualification levels, via the award of higher grades. The *Arrangements* specified that:

Qualifications in the QCF may be graded. Any grading arrangements developed for a qualification or for the components of a qualification must:

- a specify explicitly the criteria against which individual performance within the qualification or component is to be differentiated
- b identify the grading scale to be used to signify performance differentiation within the qualification and/or components.

In specifying grading criteria and identifying a grading scale, the integrity of all the other specifications of the QCF must be maintained. In particular:

- a all grading scales must include a 'pass' grade based exclusively on the assessment criteria in the unit or units on which the component or qualification is based and for which credit must be awarded
- b any additional grading criteria must build explicitly on the assessment criteria of a unit or units within the rules of combination for the qualification
- c any additional grading criteria must be consistent with the overall level of the qualification or components to be graded and must be clearly distinguished from achievements at the next level of the QCF

(Ofqual, 2008, p.16, para. 1.28 & 1.29)

These requirements concerning higher grades need to be understood in the context of requirements concerning the passing grade. A learner who passed a unit would be awarded a certain number of credits, corresponding to the size of the unit, that is, to the number of LOs deemed achievable in 10 hours of learning. The *Regulatory Arrangements* specified that:

Credits must be awarded to learners for the successful achievement of the learning outcomes of a unit. The number of credits awarded must be the same as the credit value of the unit. It is not possible for some credits to be achieved for partial completion of a unit or learners to be awarded credit when all the learning outcomes are not achieved by virtue of any 'compensation' for stronger performance in other areas of learning.

(Ofqual, 2008, p.17, para. 1.33)

In effect, this meant that all QCF units were required to conform to a mastery measurement model, meaning that all LOs – and, by implication, all AC within each LO – would need to be achieved for the award of the passing grade (and corresponding credits). The requirement that “any additional grading criteria must build explicitly on the AC of a unit or units” seems to imply that higher grades ought also to be awarded for satisfying explicit criteria, where those criteria directly correspond to, but exceed, those specified for the award of the passing grade.

Requirements in the *Regulatory Arrangements* were further explicated in a QCDA guidance document, from 2010, entitled *Grading Qualifications in the QCF*. In retrospect, the narrative within this document sounds a little odd, from a regulatory perspective, giving the impression that VTQ grading is problematic, throws up a variety of technical challenges, and is best operationalised as simply as possible.⁸

Grading Qualifications introduced itself purely as a guide to “how the technical features of the QCF influence the development of graded qualifications” (QCDA, 2010c, p.4). These technical implications included, for instance, the fact that (because the QCF operated on the principle of combining credit-based units) grading criteria had to be understood as “*additional* to the unit specification, not *part* of the unit specification” (p.7), and that AOs were only obliged to mutually recognise the credits awarded for a unit, not additional grading criteria. This, incidentally, had the unusual implication – unusual from a QCF perspective, that is – that it allowed “awarding organisations to develop their own grading arrangements” (p.7).

Grading Qualifications explained that, although: “it would be technically feasible to develop grading criteria related to more than one unit, this guidance assumes that awarding organisations will establish grading criteria within components that relate to a single unit” (p.8). It also identified:

two simple and consistent principles that can be applied uniformly across all graded qualifications in the QCF:

- The differentiation of achievement in a component must be based on assessment.
- The differentiation of achievement in a qualification must be based on aggregation.

⁸ The guidance identified a variety of simplifying heuristics (suggestions, rather than requirements or recommendations) to mitigate inevitable validity threats. These included: the fewer additional sets of grading criteria that exist within a component, the more reliable grades will be; base all components on units at the same level; develop graded components based on units with the same credit value; report only component grades and not qualification grades; base component grades on the mandatory, rather than the optional, units of a qualification; a strong case could be made that it is not appropriate to differentiate individual achievement within qualifications at Entry Level (or Level 1).

(QCDA, 2010c, p.8)

In other words, grading criteria were to be applied via an assessment process, at the component (eg unit) level; whereas, qualification grades were to be derived via an aggregation process, across components (eg units). No additional grading criteria were to be applied at the qualification level.

Grading Qualifications explicitly clarified the *Regulatory Arrangements* on various issues; for example, that there was no requirement for graded components to be based on all the mandatory and/or optional units of a qualification. On other issues, it implicitly clarified the *Regulatory Arrangements*; for example, the following passage seems to imply that a Pass at one level ought to represent a higher level of proficiency than a Distinction at the level below.

It would be technically feasible to aggregate component grades based on units at more than one level, providing it was not feasible to raise the grade of the qualification above a 'pass' grade through performance based on components at lower levels. In other words, even if a learner recorded 'Distinctions' on all components based on lower-level units in a qualification, the grading algorithm would 'weight' these grades at less than a 'pass' at a higher level in aggregating component grades towards a qualification grade.

(QCDA, 2010c, p.13)

Although it explicitly refrained from promoting any particular approach, an appendix to *Grading Qualifications* contrasted four examples of grading models (see Table 1). For each of these models, the final qualification grade would be determined on the basis of a lookup table (eg Model 1, a five-component qualification, in which 3 Distinctions and 2 Passes was deemed worthy of an overall Distinction grade; or Model 3, a six-component qualification, in which 4 'Graded's and 2 Passes was deemed worthy of an overall Merit grade).

A number of points are worth highlighting, here. First, the discussion emphasised that it was better to specify fewer sets of grading standards – to minimise reliability threats – which is why only Model 2 has two sets of grading criteria above the AC for the component (the passing criteria). Second, there is an implication that grading criteria, for all four models, map directly onto AC; such that any higher grade criterion ought to map directly onto a passing grade assessment criterion. (The mapping between grading criteria and unit grades is unambiguous for Models 1 and 2, which is why they are described simply as Merit and Distinction criteria.) Third, Model 4 is interesting in permitting higher grades to diverge from the mastery model that is required for the award of the passing grade; in the sense that Merit is awarded on the basis of the 'majority' of the criteria, ie not all of them, which is reserved for Distinction.

Table 1. Four grading models

	Qualification grades	Component grades	Basis for deriving higher component grades
Model 1	P, D	P, D [Separate criteria for P and D]	Distinction = all P criteria, plus all D criteria.
Model 2	P, M, D	P, M, D [Separate criteria for P, M and D]	Merit = all P criteria, plus all M criteria; Distinction = all P criteria, all M criteria, plus all D criteria.
Model 3	P, M, D	P, Graded [Separate criteria for P and Graded]	Graded = all P criteria, plus all Grading criteria.
Model 4	P, M, D	P, M, D [Separate criteria for P; common Grading criteria for M and D]	Merit = all P criteria, plus majority of Grading criteria; Distinction = all P criteria, plus all Grading criteria.

The GCR

QCF regulations, which were specific to QCF qualifications, were superseded by the General Conditions of Recognition (GCR), which now apply to all regulated qualifications in England, whether GQs or VTQs. The GCR regulations do not specify design requirements in anywhere near the same level of detail as the QCF regulations did.⁹ This permits greater variation in the design of VTQs – thereby supporting innovation – and assumes that validity ought to be the primary consideration when designing qualifications. In relation to the present report, with its focus on grading and levelling, the key regulatory documents are the:

1. *General Conditions of Recognition* (Ofqual, 2016); and the
2. *Qualification and Component Levels Requirements and Guidance for All Awarding Organisations and All Qualifications* (Ofqual, 2015a).

⁹ Not only do the many details contained within QCF guidance documents no longer apply, but there are no longer any guidance documents specific to the vast majority of VTQs.

Levels

Although the QCF was withdrawn, and the focus for both design and regulation shifted back from units to complete qualifications, the concept of qualification levels (which had been associated with the NQF, prior to the QCF, and with NVQs too) was retained. The Regulated Qualifications Framework (RQF) was introduced to help people to understand the range of qualifications that Ofqual regulates, and to support consistency in how AOs describe the size and challenge, or demand, of the qualifications they offer. The RQF adopted the existing QCF levels (Entry 1 to 3; Levels 1 to 8) and mapped onto the Framework for Higher Education Qualifications, as well as to the European Qualifications Framework. Although level descriptors were updated, the demand of those levels was not changed. The descriptors were revised to become more outcomes-focused, to cover both GQs and VTQs, setting out the knowledge and skills that might typically be expected of someone with a qualification at that level.

In terms of levelling, the *General Conditions* does not specify a great deal other than that an “awarding organisation must assign one or more levels to each qualification which it makes available or proposes to make available.” (Ofqual, 2016, p.55). The *Levels Requirements and Guidance* document goes into more detail, including the specification of the descriptors according to which qualifications and components must be levelled. Unlike the QCF level descriptors, the RQF ones are described in terms of only two constructs – knowledge and skills.

Reporting on the outcome of a consultation on the nature of regulation ‘*After the QCF*’ the simplification of level descriptors was explained as follows:

Some respondents suggested that the level descriptors should also include a category for autonomy and accountability.

We considered whether to add this category but given the difficulty associated with aligning knowledge and/or skills with autonomy and accountability in some job roles, we have decided to confirm our original proposal and introduce level descriptors only in the knowledge and skills categories.

This will not prevent awarding organisations from assessing autonomy and accountability in their qualifications but this should not be factored into a determination of level.

(Ofqual, 2015b, p.2)

The new level descriptors are presumed to: “set out the generic knowledge and skills associated with the typical holder of a qualification at that level.” (Ofqual, 2015a, p.5). The descriptors for Levels 2 and 3 are reproduced for illustrative purposes in Annex 3.

To establish the level of any particular qualification or component, an awarding organisation:

should look at the range of level descriptors and identify the descriptor (or where the qualification will have more than one level, the descriptors) which provides the best match with the intended knowledge and skills outcomes for their qualification. [...] The fit does not have to be perfect; qualifications might naturally have a more knowledge- or skills-based focus and so will be a better fit with the knowledge or skills descriptor as appropriate.

(Ofqual, 2015a, p.10)

In contrast to the detailed design requirements of the QCF *Regulatory Arrangements* and associated guidance documents, the process of assigning levels under the *General Conditions* is essentially post hoc, intended to link the qualification to the level that best describes it. In particular, there is no longer any explicit indication that levels ought somehow to map onto Bloom's Taxonomy, either conceptually or in terms of how LOs and AC are written.

Grading

Unlike the QCF *Regulatory Arrangements*, the *General Conditions* does not explicitly refer to 'grading'. Instead, grading requirements are implied when the *Conditions* refers to differentiating between 'specified levels' of attainment, for instance:

An awarding organisation must ensure that the specification for a qualification sets out – [...]

- (f) the knowledge, skills and understanding which will be assessed as part of the qualification (giving a clear indication of their coverage and depth), [...]
- (h) the criteria against which Learners' levels of attainment will be measured (such as assessment criteria or exemplars), [...]
- (j) any specified levels of attainment

(Ofqual, 2016, p.46, condition E3.2)

In designing such an assessment, an awarding organisation must in addition ensure that the assessment – [...]

- (f) allows each specified level of attainment detailed in the specification to be reached by a Learner who has attained the required level of knowledge, skills and understanding, and

(Ofqual, 2016, p.49, condition E4.2)

An awarding organisation must produce a written document in relation to an assessment which sets out clear and unambiguous criteria against which Learners' levels of attainment will be differentiated.

(Ofqual, 2016, p.49, condition G1.3)

Note that the *General Conditions* does not require that qualifications must be designed according to a specific measurement model, be that mastery or compensatory. Indeed, there is no regulatory requirement for any VTQ to be based upon either a mastery model or a compensatory model; neither for the passing grade nor for any higher grade.

Finally, it is worth noting that certain kinds of regulated VTQ now have bespoke Conditions. For Functional Skills Qualifications, requirements and expectations are specified for the pass/fail grade only (eg Ofqual, 2018a; 2018b); whereas, for Technical Qualifications, requirements and expectations will also be specified for higher grades too (eg Ofqual, 2018c).

Current practice in qualification design

The main substance of the present report concerns an exploratory survey of grading practices within VTQs regulated by Ofqual.

Annex 4 represents the 17728 regulated qualifications that were available for learners to study on 28/11/2017, classified according to grading structure and qualification type. Qualification type is an informal classification that is useful for distinguishing between familiar groupings. The dataset included 1876 conventional GQs (ie GCSEs, A levels, and Other General qualifications) and 15852 remaining qualifications, most of which were conventional VTQs of one sort or another.

The majority of these remaining qualifications were designated (former) QCF ones (71%), while another substantial chunk of these remaining qualifications was designated VRQ, ie Vocationally-Related Qualifications (21%). VRQs include qualifications that were regulated separately from QCF qualifications, not having been built to conform to QCF design requirements. Of the 15852 qualifications in Annex 4 that were treated as VTQs, 2167 (13.7%) awarded higher grades beyond the passing grade.

The table in Annex 4 highlights, in yellow, the 10 cross-classifications that contained more than 40 available qualifications (excluding cells in the Missing/Unknown row which were assumed to contain qualifications that award only the passing grade, not higher grades).¹⁰ They comprised:

Pass/Merit

46 QCF Ranging across various sectors, including health and safety, customer service, horticulture, music, etc. (10 AOs).

Pass/Merit/Distinction

54 ESOL English for Speakers of Other Languages (8 AOs).

48 Higher Level All at Levels 4 to 6, across various sectors, including business, marketing, policing, engineering, etc. (6 AOs).

609 QCF Typically involving performing arts (eg dance, music, performance) but also including other sectors such as health and safety, management, travel and tourism, sports, hairdressing, employability, etc. (43 AOs).

¹⁰ Technically, in this analysis, the Missing/Unknown row contains both Pass/Fail qualifications and qualifications for which the grading structure is genuinely unknown, ie unrecorded on the Ofqual database. A higher-level analysis indicated that 76.5% of available qualifications were classified as P/F, 22.5% as having a different grading structure, and for only 1% was the grading structure genuinely unknown. In other words, it is safe to assume that almost all of the 13895 qualifications in the M/U row award only the passing grade.

465 VRQ Ranging across various sectors, including construction, engineering, maintenance, performance, treatments, communication, management, etc. (39 AOs).

Pass/Merit/Distinction/Distinction*

250 QCF Ranging across various sectors, including blacksmithing, floristry, horticulture, IT, music technology, sport, science, law, travel and tourism, agriculture, childcare, etc. (8 AOs).

261 VRQ Ranging across various sectors, including bricklaying, cyber security, barbering, engineering, equine care, site carpentry, nail technology, interactive media, business, law, massage therapies, etc. (12 AOs).

PP/PM/MM/MD/DD/DD*/D*D*

77 QCF Level 3 Diplomas: OCR Cambridge Technical; Pearson BTEC (2 AOs).

57 VRQ Level 3 Diplomas: AQA Technical; OCR Cambridge Technical; Pearson BTEC; VTCT (4 AOs).

PPP/PPM/PMM/MMM/MMD/MDD/DDD/DDD*/DD*D*/D*D*D*

45 QCF Level 3 Extended Diplomas: OCR Cambridge Technical; Pearson BTEC (2 AOs).

Selecting a sample of qualifications

Qualification titles within each of these 10 categories were scanned visually to identify a sample of qualifications that seemed likely to reflect a variety of grading practices. The intention underlying this informal approach was to identify as wide a variety of grading approaches as possible – within the limitations of a small-scale, exploratory survey – whilst also attempting to target common varieties of qualification within each category (eg Graded Exams in the 609 QCF category), as well as AOs that offered a lot of qualifications within each category (eg VTCT in the 261 VRQ category). The sampling was intended to be more theoretical than representative; however, given the scale and informality of the study, there was no attempt to reach ‘theoretical saturation’ in terms of the variety of grading approaches encountered. As an exploratory survey, the intention was to identify a reasonable range of relevant issues, to provide a foundation for a wider conversation concerning good practice in grading VTQs, rather than to identify a full set of issues. Consequently, this was very much an ‘opportunity sample’ and qualifications were only included if sufficient information could be located on the internet concerning their approach to grading. Specifications and related materials for the following 18 qualifications (identified

below in terms of Qualification Number and Qualification Title) were downloaded and scrutinised:

Pass/Merit – 46 QCF

500/5317/6 RSPH Level 4 Award in Food Safety Management for Manufacturing

Pass/Merit/Distinction – 54 ESOL

601/1931/7 NOCN Level 1 Certificate in ESOL International (B2)

Pass/Merit/Distinction – 48 Higher Level

603/2341/3 Pearson BTEC Level 4 Higher National Certificate in Policing

Pass/Merit/Distinction – 609 QCF

500/8255/3 Pearson BTEC Level 5 HND Diploma in Electrical Engineering

501/2073/6 TCL Level 3 Certificate in Graded Examination in Speech and Drama (Grade 8)

501/2030/X UWLQ Level 2 Certificate in Graded Examination in Oral Communication (Grade 4)

500/8477/X ABC Level 4 Foundation Diploma in Art, Design and Media

601/6170/X ATHE Level 3 Diploma In Business

Pass/Merit/Distinction – 465 VRQ

600/8603/8 Cskills Awards Level 2 Diploma in Site Carpentry (Construction)¹¹

601/8423/1 CIBTAC Level 3 Diploma in Beauty Therapy

601/7324/5 IMI Level 3 Diploma in Light Vehicle Maintenance (VRQ)

Pass/Merit/Distinction/Distinction* – 250 QCF

601/2624/3 NCFE Level 2 Diploma in Skills for Business

600/6609/X RSL Level 3 Diploma for Music Practitioners

Pass/Merit/Distinction/Distinction* – 261 VRQ

601/6993/X VTCT Level 3 Extended Diploma in Barbering

601/7145/5 AQA Level 3 Certificate in Applied Business

PP/PM/MM/MD/DD/DD*/D*D* – 77 QCF

600/4229/1 OCR Level 3 Cambridge Technical Diploma in Business (QCF)

¹¹ Just prior to the commencement of this study, 01/08/17, Cskills Awards – then an operating division of the Construction Industry Training Board – was acquired by NOCN. This particular qualification is now rebranded as the NOCN_Cskills Awards Level 2 Diploma in Site Carpentry (Construction). The present report focuses on the original Cskills specification, and therefore retains the original branding.

PP/PM/MM/MD/DD/DD*/D*D* – 57 VRQ

601/7101/7 OCR Level 3 Cambridge Technical Diploma in IT

PPP/PPM/PMM/MMM/MMD/MDD/DDD/DDD*/DD*D*/D*D*D* – 45 QCF

601/3493/8 Pearson BTEC Level 3 Extended Diploma in Dental Technology

Again, there was no intention to select a representative sample, and certain AOs were sampled more frequently than others, eg Pearson. This was partly due to certain AOs dominating certain categories, eg Pearson and OCR dominating the last three categories. It was also partly due to the lack of available information on grading approach for certain AOs, eg City & Guilds, where critical information on unit grading tended to be available only to registered centres.

Results from the exploratory survey

One of the most interesting observations, having selected a sample of qualifications, was how easy it was to identify a wide variety of grading practices. Almost every qualification within the initial selection of 'likely' qualifications presented something about its grading approach that rendered it at least slightly different from the others. Table 2, below, summarises key information on the structure of the qualifications chosen for the final sample.

The sample ranges from single-unit qualifications that are 100% externally assessed via a single exam to multi-unit qualifications that are 100% internally assessed via multiple methods. It includes qualifications with units at a single level, and qualifications with units ranging across levels. Although these qualifications award different qualification grades (eg P/M/D vs PP/PM/MM/MD/DD/DD*/D*D*), they almost all award the same unit grades, ie P/M/D. Note that **RSPH L4 Food Award**¹² was classified on the Ofqual register as a P/M qualification, although it actually awards P/D.

For the majority of qualifications in the sample, all component units/assessments are graded, and these grades are converted into an overall qualification grade. However, for 4 of the 18 qualifications, only certain units/assessments are graded. For 1 of these qualifications (**ABC L4 Art F.Dipl.**), only 1 of its 8 units is graded, and the grade on this unit determines the overall qualification grade.

¹² For the sake of space and legibility, the present report will refer to particular qualifications using this abbreviated form, rather than the full qualification title.

Table 2. Qualification structure

Qual. abbreviation	Qual. grade	Qual. type	No. units	Unit level(s)	External assess.	Internal assess.	Unit/Ass. grading	All U/A graded?
RSPH L4 Food Award	P/D	QCF	1	4	100%		P/D	Y
NOCN L1 ESOL Cert.B2	P/M/D	ESOL	4	1	100%		P/M/D	Y
BTEC L4 Policing HNC	P/M/D	HIGHER	8	4		100%	P/M/D	Y
BTEC L5 Electrical HND	P/M/D	QCF	~ 16	4 & 5		100%	P/M/D	Y
TCL L3 Speech Grade 8	P/M/D	QCF	1	3	100%		P/M/D	Y
UWLQ L2 Oral Grade 4	P/M/D	QCF	1	2	100%		P/M/D	Y
ABC L4 Art F.Dipl.	P/M/D	QCF	8	3 & 4		100%	P/M/D	N
ATHE L3 Business Dipl.	P/M/D	QCF	4	3		100%	P/M/D	Y
Cskills L2 Carpentry Dipl.	P/M/D	VRQ	8	1 & 2	8 units	5 units	P/M/D	N
CIBTAC L3 Beauty Dipl.	P/M/D	VRQ	6	3	100%		P/M/D	Y
IMI L3 Vehicle Dipl.	P/M/D	VRQ	10	2 & 3	8 units	10 units	P/M/D	N
NCFE L2 Bus.Skills Dipl.	P/M/D/D*	QCF	6	2		100%	P/M/D	Y
RSL L3 Music Dipl.	P/M/D/D*	QCF	~ 11	3		100%	P/M/D	Y
VTCT L3 Barbering E.Dipl.	P/M/D/D*	VRQ	10	3 (mainly)	6 units	10 units	P/M/D	N
AQA L3 App.Bus. Cert.	P/M/D/D*	VRQ	3	3	2 units	1 unit	P/M/D	Y
OCR L3 Business T.Dipl.	P/M/D/D* (x2)	QCF	12	3		100%	P/M/D	Y
OCR L3 IT T.Dipl.	P/M/D/D* (x2)	VRQ	11	3	3 units	8 units	P/M/D	Y
BTEC L3 Dental E.Dipl.	P/M/D/D* (x3)	QCF	16	3		100%	P/M/D	Y

For each of the 18 qualifications, in turn, Annex 5 presents a brief summary of its grading approach, preceded by a very brief summary of its purpose and assessment methods. Sufficiently comprehensive information was available for most qualifications, although a small amount of information was unavailable for a few. Where necessary, supplementary information was kindly provided by the relevant AO.

The information in Annex 5 focuses specifically on approaches to grading. Although grading is clearly linked to issues of comparability of standards, comparability was not a specific focus for the present project, and no attempt was made to identify approaches to linking standards (to achieve comparability) for any qualification. Not only was this beyond the remit of the present study, the relevant information is generally not publicly available and would therefore have required a different methodological approach. Consequently, although Annex 5 occasionally begs questions of comparability – eg when grade boundaries are established using fixed mark thresholds – it contains no information on such matters. (Comparability is considered as a technical issue arising, in the Discussion section of this report.)

In addition to information concerning the 18 qualifications, the summaries in Annex 5 sometimes contain additional information arising from cross-level analyses. When one of the focal qualifications was offered at multiple levels (eg L2 Certificate, L3 Certificate, and L4 Certificate), scrutiny of corresponding specifications permitted an element of comparison between approaches across levels. Because (unsurprisingly) grading approaches tended to be structurally very similar, if not identical, across levels, the principal focus for such comparisons turned out to be the use of command verbs within the AC for units at different levels. This cross-level analysis was opportunistic, and was only undertaken when direct comparisons seemed legitimate, eg when the focal qualification comprised units at different levels, or when the qualification (or a very close relative) was offered at multiple levels.

Annex 5 presents information concerning how (and, to some extent, why) grading approaches differed across the 18 qualifications. Key insights related to these approaches are summarised below.

The following two sections – concerned with measurement models and measurement standards, respectively – discuss the two major themes that emerged from the comparison of grading approaches. The term ‘measurement model’ is used, here, to describe the approach that a qualification adopts to aggregating assessment information; where that information might exist in the form of task marks/scores or criterion judgements (which are aggregated within units to determine the unit grade), or in the form of unit marks or unit grade points or unit grade profiles (which are aggregated across units to determine the qualification grade). The term ‘measurement standard(s)’ is used to describe the ultimate criterion (or criteria) against which candidates are to be judged; as specified for the passing grade and for higher grades.

Theme 1: Measurement models

As noted earlier, the (recently withdrawn) QCF regulations specified that all QCF units and qualifications should conform to a mastery measurement model, meaning that all learning outcomes (LOs) – and, by implication, all assessment criteria (AC) within each LO – would need to be achieved for the award of the passing grade. An additional requirement that “any additional grading criteria must build explicitly on the assessment criteria of a unit or units” seemed to imply that higher grades ought also to be awarded for satisfying explicit criteria, where those criteria directly correspond to, but exceed, those specified for the award of the passing grade.

Table 3 demonstrates that none of the 18 qualifications in the sample – whether former QCF qualifications or not – operates on the basis of a *pure* mastery measurement model. All of the sampled qualifications operate by combining elements of mastery and compensation; and, as we shall see, sometimes also by operating other aggregatory principles, including configuration and charity.

Table 3. Evidence of mastery versus compensation

Abbreviation	Basis for awarding higher qual. grades	Evidence of operating a mastery measurement model	Evidence of operating a compensatory measurement model
RSPH L4 Food Award	Performance across both exams.	Qualification P requires P on both exams (same for D).	Simple mark addition aggregation within both exams. “In order to be awarded a Pass, candidates must be able to recall and apply relevant knowledge and facts from some parts of the specification” (Detailed AC are not specified for either exam; only for the qualification overall.)
NOCN L1 ESOL Cert.B2	Performance across all units.	Qualification P requires P on all units.	Simple mark addition aggregation within all 4 exams. Weighted mark-total addition aggregation across (all) units allows compensation for higher grades. (Detailed AC are not specified for any unit.)
BTEC L4 Policing HNC	Performance across all units.	Each unit AC/grading criterion needs to be satisfied for award of unit P, M, or D.	Qualification P does not require P on all units. Unit grade point score addition across (all) units allows compensation for higher grades.
BTEC L5 Electrical HND	Performance in (~ 5) highest graded (L5) units only.	Qualification P requires P on all units. Each unit AC needs to be satisfied for award of unit P. Each generic grading criterion needs to be satisfied for award of unit M, or D (although	Unit grade point score addition across (highest graded) units allows compensation for higher grades. Higher proficiency on units that affect qualification grade may compensate for lower proficiency on other units.

Abbreviation	Basis for awarding higher qual. grades	Evidence of operating a mastery measurement model	Evidence of operating a compensatory measurement model
		this is intended with some leeway, which suggests a weaker notion of mastery).	
TCL L3 Speech Grade 8	Performance in single (two component) exam.	[Exam P = qualification P.]	Simple mark addition aggregation within the exam. (Detailed AC are specified, but they do not relate directly to marking schemes.)
UWLQ L2 Oral Grade 4	Performance in single exam.	[Exam P = qualification P.]	Simple mark addition aggregation within the exam. (Detailed AC are not specified.)
ABC L4 Art F.Dipl.	Performance in 1 unit only (totalling 1/3 of credits for qualification).	Qualification P requires P on all units. Each unit AC needs to be satisfied for award of unit P. Each unit grading criterion needs to be satisfied (on the single grading unit) for award of unit M, or D (which means no compensation for higher qualification grades).	In theory, higher proficiency on unit that determines qualification grade could compensate for lower proficiency on other units. However, in practice, staged approach to progression through units, and synoptic nature of the final unit, counters that logic.
ATHE L3 Business Dipl.	Performance across all units.	Qualification P requires P on all units. Each unit AC/grading criterion needs to be satisfied for award of unit P, M, or D.	Unit grade point score addition across (all) units allows compensation for higher grades.
Cskills L2 Carpentry Dipl.	Performance across all units; but only in MCT knowledge	Qualification P requires P on all units (for units with both MCT exam and practical assignment, unit P requires P on both). Each unit AC needs to be satisfied for award	Mark addition aggregation within MCT exams (for each unit). Unit grade aggregation method allows compensation for higher grades.

Abbreviation	Basis for awarding higher qual. grades	Evidence of operating a mastery measurement model	Evidence of operating a compensatory measurement model
CIBTAC L3 Beauty Dipl.	<p>exams (not practical assignments).</p> <p>Performance across all units.</p>	<p>of unit P on the (Occupational) practical assignments.</p> <p>Qualification P requires P on all unit-elements/subtests.</p> <p>Qualification grade corresponds to lowest grade achieved across all externally assessed unit-elements/subtests.</p> <p>Certain marks within the Practical Assignment must be achieved in order to pass, acting as hurdle-marks.</p>	<p>Higher proficiency on assessments that affect qualification grade may compensate for lower proficiency on other assessments.</p> <p>Simple mark addition aggregation within the MCT exam (for each unit-subtest).</p> <p>Mark addition aggregation within the Practical Assignment (but alongside hurdle-marks).</p> <p>(Unit AC are specified, but not assessed discretely.)</p>
IMI L3 Vehicle Dipl.	<p>Performance across 3 MCT exams and 1 synoptic assessment Practical Task Job Report (excludes performance in 2 optional units).</p>	<p>Qualification P requires P on all units. Each unit AC needs to be satisfied for award of P on the Task-Portfolios.</p> <p>Each AC/grading criterion needs to be satisfied for award of unit P, M, or D on the synoptic assessment Practical Task Job Report.</p>	<p>Simple mark addition aggregation within the 3 MCT exams (covering 8 mandatory units). Unit grade aggregation method allows compensation for higher grades.</p> <p>Higher proficiency on assessments that affect qualification grade may compensate for lower proficiency on other assessments.</p>
NCFE L2 Bus.Skills Dipl.	<p>Performance across all units.</p>	<p>Qualification P requires P on all units. Each unit AC/grading criterion needs to be satisfied for award of unit P, M, or D.</p>	<p>Lookup table aggregation across (all) units allows compensation for higher grades.</p>

Abbreviation	Basis for awarding higher qual. grades	Evidence of operating a mastery measurement model	Evidence of operating a compensatory measurement model
RSL L3 Music Dipl.	Performance across all units.	Qualification P requires P on all units. Each unit AC, for each LO, needs to be satisfied for award of unit P.	Within units, tasks/AC are graded, and unit grade is 'average' task/AC grade (P, M, or D). Unit grade point score addition across (all) units allows compensation for higher grades.
VTCT L3 Barbering E.Dipl.	Performance across 5 practical assessments, 2 exams and 1 synoptic assessment (excludes performance in 4 optional units and 1 mandatory unit).	Qualification P requires P on all units. Each unit AC needs to be satisfied for award of unit P across all units (via practical assessment on mandatory units and portfolio assessment on optional ones). Each unit grading criterion needs to be satisfied (via practical assessments) for award of unit M, or D. For these units, grading criteria and AC are quite distinct.	Simple mark addition aggregation within both terminal exams. Grade point aggregation method allows compensation for higher grades. Synoptic assessment graded using a best-fit rubric. Higher proficiency on assessments that affect qualification grade may compensate for lower proficiency on other assessments.
AQA L3 App.Bus. Cert.	Performance across all units.	Qualification P requires P on all units.	Simple mark addition aggregation within the unit 1 exam. Even though detailed AC/grading criteria are specified for practical assignment units (unit 2 – internal, unit 3 – external), these do not all have to be satisfied, for award of any grade. Even at P, it is possible to compensate for not satisfying an AC by achieving another at a higher grade. Unit grade point score addition across (all)

Abbreviation	Basis for awarding higher qual. grades	Evidence of operating a mastery measurement model	Evidence of operating a compensatory measurement model
OCR L3 Business T.Dipl.	Performance across all units.	Qualification P requires P on all units. Each unit AC/grading criterion needs to be satisfied for award of unit P, M, or D.	units allows compensation for higher grades. (Detailed AC are not specified for the exam only unit.) Unit grade point score addition across (all) units allows compensation for higher grades.
OCR L3 IT T.Dipl.	Performance across all units.	Qualification P requires P on all units. Each unit AC/grading criterion needs to be satisfied for award of unit P, M, or D on the 8 internally assessed units.	Simple mark addition aggregation within the 3 exams (units 1-3). Lookup table aggregation across (all) units allows compensation for higher grades. (Detailed AC are not specified for the 3 exam-only units.)
BTEC L3 Dental E.Dipl.	Performance across all units.	Qualification P requires P on all units. Each unit AC/grading criterion needs to be satisfied for award of unit P, M, or D.	Unit grade point score addition across (all) units allows compensation for higher grades.

1a. Mastery within qualification grading

Almost all of the sampled qualifications operate what might be considered a fairly basic mastery requirement, at least as far as competence-based qualifications are concerned: that a qualification Pass cannot be achieved without having achieved a Pass on all units.

However, one of the sampled qualification, **BTEC L4 Policing HNC**, does not operate this requirement. Instead, it permits candidates to fail a unit (but only one unit) and still pass the qualification. On the other hand, the same qualification operates a strong mastery requirement at the unit level, specifying not only that each AC needs to be satisfied at Pass to pass the unit overall, but that all Merit criteria need to be satisfied for Merit overall, and all Distinction criteria need to be satisfied for Distinction overall.

1b. Mastery within unit grading

This strong mastery requirement – satisfy all criteria for the award of the unit grade – was fairly common amongst the sampled qualifications with internally assessed units (see also **ATHE L3 Business Dipl.**, **NCFE L2 Bus.Skills Dipl.**, **OCR L3 Business T.Dipl.**, **OCR L3 IT T.Dipl.**, **BTEC L3 Dental E.Dipl.**). For qualifications like these, each internally assessed unit is specified in terms of LOs and associated AC. Some qualifications specify Pass criteria for each AC – essentially the same AC, just worded slightly differently – while, for others, the AC simply are the Pass criteria. Additional, higher grade criteria are also specified, for Merit and Distinction. Typically, these higher grade criteria map directly onto passing grade criteria; such that satisfying an AC at a higher grade (Merit or Distinction) means also having satisfied the same criterion at a lower grade (Merit or Pass). Often, though, there are fewer Merit or Distinction criteria for a unit than Pass criteria.

Figure 3 illustrates this approach, using **OCR L3 Business T.Dipl.** (Unit 5, 60 GLH, 10 credits).¹³ In this example, all 5 AC need to be satisfied at Pass for the overall unit Pass. Similarly, all of the corresponding grading criteria need also to be satisfied at both Merit and then Distinction. However, only 3 are actually specified for Merit, and only 2 for Distinction.

¹³ Reproduced from page 3 of undated OCR document titled: OCR Level 3 Cambridge Technical Certificate/Diploma In Business. Human Resource Management In Business. K/502/5445. Level 3 Unit 5. Guided Learning Hours: 60. Unit Credit Value: 10.

Figure 3. Example of grading criteria (strong mastery requirement # 1)

Learning Outcome (LO)	Pass	Merit	Distinction
The learner will:	The learner can:		
1 Know the factors that are involved in human resource planning in organisations	P1 describe the internal and external factors to consider when planning the human resources requirements of an organisation		
	P2 describe how the skills that employees require to carry out jobs in an organisation are identified	M1 explain the benefits to an organisation of using a skills audit for human resource planning	D1 evaluate the steps that an organisation could take to improve the skills of its employees
2 Know how organisations motivate employees	P3 outline how an organisation motivates its employees	M2 analyse different methods of motivating employees in an organisation	D2 recommend ways for a specific organisation to improve employee motivation
3 Understand how to gain committed employee cooperation	P4 explain how organisations obtain the cooperation of their employees		
4 Understand the importance of managing employee performance at work	P5 explain how employee performance is measured and managed.	M3 analyse the advantages and limitations of measuring and managing employee performance	

Although this approach was common across qualifications with internally assessed units, there was quite a lot of variation – both within qualifications across units, and across qualifications – in the proportions of AC for which Merit and Distinction criteria were specified (in other words, for each unit, how many of the Merit and Distinction boxes were filled with criteria). The following two examples illustrate this point:

NCFE L2 Bus.Skills Dipl.

	Unit 1 (8 credit, optional)	Unit 8 (8 credit, optional)
LOs	3	3
Pass criteria	12	9
Merit criteria	12	9
Distinction criteria	3	4

BTEC L3 Dental E.Dipl.

	Unit 3 (15 credit, mandatory)	Unit 15 (10 credit, mandatory)
LOs	5	6
Pass criteria	17	8
Merit criteria	13	2
Distinction criteria	5	2

Although it was typical (within the sampled qualifications) for the strong, unit-level mastery requirement to be framed in terms of a hierarchical set of (up to 3) criteria for each criterion, other approaches were evident. For instance, Figure 4, below, presents grading criteria for the practical assessment from the **VTCT L3 Barbering E.Dipl.** qualification (Unit UHB130M).¹⁴

¹⁴ Reproduced from pages 36-7 of undated VTCT document titled: Assessment Record. VTCT Level 3 Extended Diploma in Barbering. 601/6993/X. HB3ED2F_v2.

Figure 4. Example of grading criteria (strong mastery requirement # 2)

Learners must meet all pass criteria to achieve a pass grade.

Pass Criteria	✓
L04 Be able to prepare the client, self and work area for the creation of facial hair shapes	○
P1 - Prepare and check the area, products, tools and equipment prior to the service	○
P2 - Prepare the client for services	○
P3 - Explain and agree the service outcome with the client	○
L05 Be able to provide the service of creating facial hair shapes	○
P4 - Select products, tools and equipment to create facial hair shapes	○
P5 - Use safe and hygienic working methods throughout the service	○
P6 - Use products, tools and equipment to produce the desired finish	○
P7 - Use a combination of techniques to achieve the desired outcome considering influencing factors	○
P8 - Provide in-service checks with the client	○
P9 - Provide advice and recommendations to the client	○

Learners must meet all pass criteria and all merit criteria to achieve a merit grade.

Merit Criteria	✓
M1 - Adapt and tailor interpersonal communication to the client throughout the service	○
M2 - Use effective strategies to promote and sell products and additional services throughout the facial hair shaping service	○
M3 - Use the tools and equipment in a methodical and controlled way to create facial hair shapes	○

Learners must meet all pass criteria, all merit criteria and all distinction criteria to achieve a distinction grade.

Distinction Criteria	✓
D1 - Justify the selection of products, tools, equipment and techniques used throughout the creating facial hair shaping service	○
D2 - Create a finished facial hair shape that reflects the mastery of a professional barbering service	○

Notice how there is still a strong mastery requirement, here, at the assessment level, since all criteria need to be satisfied for the award of each grade. And notice how the criteria are also hierarchical, in a sense, because a unit Merit (or Distinction) grade cannot be achieved without also having satisfied all of the Pass (or Merit) criteria. However, in this example, the higher grade criteria do not map directly onto the passing grade criteria. Instead, the higher grade criteria are of a somewhat different kind; indicating that the candidate has achieved the unit with a greater degree of professional sophistication. Indeed, the unit specification explicitly contrasts:

- Pass grade criteria – “proficient technical skills and safe performance [...] minimum industry standards of practice”; with
- Merit grade criteria – “excellent technical skills and techniques, organisational skills or communication skills”; and
- Distinction grade criteria – “mastery of professional techniques [...] an ability to reflect upon and evaluate one’s own performance and to justify the methods and techniques chosen”.

Other variants of the unit-level mastery requirement included **BTEC L5 Electrical HND** which awards Merit and Distinction grades, across units, on the basis of generic grading criteria – entirely distinct from the unit-specific AC – which are used to judge the full complement of work produced for the unit. The Merit and Distinction criteria are specified as follows:

Merit

1. identify and apply strategies to find appropriate solutions;
2. select/design and apply appropriate methods/techniques; and
3. present and communicate appropriate findings.

Distinction

1. use critical reflection to evaluate own work and justify valid conclusions;
2. take responsibility for managing and organising activities; and
3. demonstrate convergent/lateral/creative thinking.

Although these statements for both Merit and Distinction refer to criteria that “must” be demonstrated for the award of the grade, they – in particular, their ‘indicative characteristics’ (see Annex 5) – are intended to be applied with some leeway, which suggests that they embody a somewhat looser mastery requirement.

Finally, it is worth noting the example of **ABC L4 Art F.Dipl.** which operates a strong, unit-level mastery requirement, at all grades, but that only awards higher grades on a single unit. This is unusual, in the sense that it requires mastery of all specified criteria, but only specifies criteria for Merit and Distinction on one of the units. The rationale for grading only a single unit, the final unit, comprising 1/3 of the available credits for the qualification, stems from the stage-based model upon which it is based. The first three units constitute the Exploratory Stage of the qualification, where learners get to grips with its aims and purposes. The next four units constitute the Pathway Stage, in which learners identify their own areas of interest, and develop

a more personal programme of study. The final unit constitutes the Confirmatory Stage, which leads to the production of a Personal Confirmatory Study, and which is ultimately graded. This brings together all of the knowledge, skill and understanding developed throughout the course.

1c. Partial grading approaches

In fact, this is not the only qualification from the sample to determine higher grades on the basis of partial evidence, although different approaches appeared to be based upon differing rationales. Two of the qualifications exclude optional units from grading, requiring simply a Pass on those units (**VTCT L3 Barbering E.Dipl.**, **IMI L3 Vehicle Dipl.**). One L5 qualification, with units at both L4 and L5, determines higher qualification grades on the basis of a relatively small subsample of those L5 units with the highest unit grades (**BTEC L5 Electrical HND**). Another qualification determines higher qualification grades on the basis of information from all 8 units, but determines each unit grade purely on the basis of a multiple-choice 'knowledge test' exam, excluding evidence from practical assignments (**Cskills L2 Carpentry Dipl.**).

Grading on the basis of partial evidence means that higher grades are based on less than complete evidence of proficiency. This is not necessarily inappropriate. Indeed, it could be argued that limiting the award of higher grades to especially significant evidence is highly appropriate. On the other hand, the risk with partial grading is that decisions on what to include/exclude are made inappropriately, opportunistically, or arbitrarily, leading to the award of higher grades on inappropriate bases; and potentially also having a negative backwash on teaching and learning.

1d. Compensation when aggregating AC within units/assessments

Although the strong, unit-level mastery requirement was fairly common across the sampled qualifications with internally assessed units, there were a number of variants. For example, one qualification operates a strong mastery requirement for the unit Pass grade, but awards unit Merit and Distinction grades via compensation (**RSL L3 Music Dipl.**). For each LO, within each unit, AC are assessed via discrete tasks, and performance on each task/AC is graded. The LO grade corresponds to the 'average' or 'majority' grade (P, M, or D) across the tasks/AC associated with each LO. For units with multiple LOs, the unit grade corresponds to the 'average' or 'majority' grade (P, M, or D) across LOs. (If there were two LOs in a unit – graded Pass and Merit, respectively – then the unit grade would be the higher of the two.)

For another qualification, even though detailed AC/grading criteria are specified for practical assignment units, these do not all have to be satisfied for award of *any* grade (**AQA L3 App.Bus. Cert.**). Even at P, it is possible to compensate for not satisfying a particular AC by achieving another at a higher grade. For yet another qualification, the synoptic assessment is not graded according to a detailed set of AC/grading criteria, but according to a 'best-fit' rubric (**VTCT L3 Barbering E.Dipl.**).

1e. Compensation when aggregating marks within units/assessments

The sampled qualifications that include external exams typically adopt the conventional GQ approach of awarding marks, and then using simple mark addition aggregation to determine the corresponding unit/exam mark total. Consequently, these externally assessed units/exams all operate on the principle of compensation, whereby high performance on certain elements of the LO(s) assessed by the unit/exam can compensate for low performance on others.

A notable exception was the **CIBTAC L3 Beauty Dipl.** qualification, which has an externally assessed Practical Assignment that is marked according to a mark scheme, with (unit-element) marks totalled and converted into grades via thresholds. By adopting this approach, it does permit an element of compensation. However, the mark scheme also specifies that candidates need to achieve specific marks on the Practical Assignment – for the demonstration of specific skills that need to have been mastered – in order to pass each unit-element. These marks therefore function as hurdles for each unit-element; and each unit-element has a moderately large number of hurdle-marks.

For some of the exam-incorporating qualifications, certain units are assessed purely by external exam, rendering the specification of detailed AC somewhat redundant. For some of these units, LOs are published without AC (eg **OCR L3 IT T.Dipl.**). For others, AC are published, although their primary use relates to item writing, and to facilitating representative content sampling (eg **Cskills L2 Carpentry Dipl.**).

1f. Compensation when aggregating across units/assessments

With the exception of the 2 single-exam qualifications, and the qualification that awards higher grades from a single unit, and the double-exam qualification, almost all other qualifications in the sample determine their overall qualification grade by aggregating component marks, points, or grades in a manner that allows compensation for higher qualification grades; such that, for instance, a Distinction on one unit would be allowed to compensate for a Pass on another, resulting in an overall Merit grade.

The one exception to this trend was the **CIBTAC L3 Beauty Dipl.** qualification, which operates a very strict mastery principle, requiring at least Merit (Distinction) across all external assessments for an overall qualification Merit (Distinction).

1g. Charity when aggregating across units/assessments

The principle of compensation is generally assumed to imply an 'averaging' of proficiency judgements across subdomains (or lower level elements) within a domain. That might correspond to an 'averaging' of marks (eg the **RSPH L4 Food Award** where stronger and weaker performances within each of the two exams 'average out' in their respective mark totals); or an 'averaging' of AC judgements (eg the **RSL L3 Music Dipl.** where the unit grade corresponds to the 'average' grade across the LOs/tasks/AC that comprise it). Equally, it might correspond to an

'averaging' of marks/points/grades across units within a qualification, as described above.

Sometimes, though, proficiency judgements are aggregated more generously. As noted above, the **CIBTAC L3 Beauty Dipl.** qualification operates a very strict mastery principle, whereby the qualification grade is determined by the lowest of any component grades. Clearly, this is the opposite of being 'generous' to candidates. Yet, the information that it provides to qualification users – indicating the lowest level of performance demonstrated across subdomains – can still be useful in certain contexts (eg when certificating or licensing professionals who need to be at least minimally competent across all subdomains).

In theory, this principle could be flipped on its head, meaning that the qualification grade would be determined by the highest of any component grades. This illustrates the idea of being 'generous' to candidates. The information that this would provide to qualification users – indicating the highest level of performance demonstrated across subdomains – could also be useful in certain contexts (eg when identifying learners who have shown a certain amount of flair for the subject, even if only patchily). Under this aggregatory principle – which I will refer to as the principle of **charity** – many candidates will end up with a qualification grade that exceeds the 'average' of their component grades. The three aggregatory principles so far discussed can be contrasted as follows:

- mastery – overall result represents (or tends towards) lowest level of proficiency across specified subdomains of a domain;
- compensatory – overall result represents average level of proficiency across specified subdomains of a domain;
- charity – overall result represents (or tends towards) highest level of proficiency across specified subdomains of a domain.

Although stronger implementations of the charity principle are rarely encountered in educational settings, weaker implementations do occur. Amongst the sampled qualifications, this principle was observed most clearly in the approach that certain AOs adopt to aggregating unit points/grades to determine higher qualification grades.

Table 4 indicates where the principle of charity operates across the sampled qualifications, in relation to higher qualification grades. Whereas almost all of the (multi-unit) qualifications apply a strict mastery principle when determining the qualification passing grade – requiring a Pass on all units, and some also requiring a Pass on all assessed components within each unit – almost all of these (multi-unit) qualifications apply at least a compensatory principle when determining higher grades; and some apply a charity principle, particularly for Distinction and Distinction* grades. Table 4 indicates, via blue font, where the aggregatory principle is best described as charity rather than either compensatory or mastery. For instance, for both the 8-unit **BTEC L4 Policing HNC** qualification and the 6-unit **NCFE L2 Bus.Skills Dipl.**, a Merit grade is awarded with only half of the units at Merit and the other half at Pass, meaning that the overall result tends towards the higher grade. This is the weakest version of charity, within Table 4, which nudges what might

otherwise be a compensatory decision upwards; that is, when the notional 'average' grade falls halfway between Pass and Merit, and the grading decision tends towards the higher grade (charity), rather than the lower one (mastery).

This contrasts with the **ATHE L3 Business Dipl.** qualification and the **VTCT L3 Barbering E.Dipl.** qualification, both of which require three-quarters of the units to be achieved at Merit for the award of Merit overall. In other words, these qualifications aggregate half-Pass/half-Merit down to a Pass, thereby nudging the overall grading decision towards the lower of two possible grades (mastery).

Interestingly, though, whereas **ATHE L3 Business Dipl.** retains its three-quarters requirement for the Distinction grade, **VTCT L3 Barbering E.Dipl.** lowers it to half; in effect, switching to a different aggregatory principle. Note that the **VTCT L3 Barbering E.Dipl.** qualification has an additional D* grade to award (cf **ATHE L3 Business Dipl.**). When the highest grade awarded to a unit is a D, and unit points/grades are aggregated directly into the qualification grade, it is simply not possible to award a D* without adopting a version of the charity principle. If the aggregation model relied purely upon either mastery or compensation, then a profile of straight Ds across all units would correspond to grade D at the qualification level, leaving no 'headroom' for D*. The fact that **VTCT L3 Barbering E.Dipl.** has to make room for a D* helps to explain why it applies more charity at D than at M. This also helps to explain why the last 7 rows of the table, which correspond to the qualifications that award D*, present far more text in blue font than the first 11.

The only multi-unit qualification from the sample that applies compensation but not charity is the **AQA L3 App.Bus. Cert.** qualification. It adopts an assessment approach that might be classified as halfway between GQ and VTQ; for example, unit grading requires judgements at the level of AC, but those judgements are aggregated as though they were marks. In addition, it adopts a Uniform Mark Scale (UMS) approach to scaling unit marks; the scaled scores are then aggregated to determine the qualification grade. For this qualification, the qualification grade thresholds are equal to the sum of the unit grade UMS thresholds. In other words, if you score at the unit UMS Merit threshold on each of the units, then your UMS mark total will be equivalent to the qualification UMS Merit grade threshold. The use of a UMS model is significant, here, because it allows unit achievement beyond the D grade (threshold) to be rewarded with additional UMS marks. In other words, the UMS marks distinguish clearly between a high D on a unit and a low D. (Whereas, most of the other qualifications make no such distinction.) This use of the UMS makes it possible to differentiate between D and D* at the qualification level, despite not adopting the charity principle.

The biggest qualifications, within Table 4, tend also to operate the most charity, particularly for the double-/triple-Distinction grade (DD or DDD). **OCR L3 Business T.Dipl.** (DD), **OCR L3 IT T.Dipl.** (DD), and **BTEC L3 Dental E.Dipl.** (DDD) all award their double-/triple-Distinction grade with only a minority of unit grades (around 1/3) at Distinction.

Table 4. Evidence of charity in awarding higher qualification grades

Abbreviation	D*	Degree to which charity principle operated across higher grade units (blue = charity)
RSPH L4 Food Award	N	N/A (single unit).
NOCN L1 ESOL Cert.B2	N	Based purely on exam marks.
BTEC L4 Policing HNC	N	M (threshold) = 4 P + 4 M (ie 1/2 at M) D (threshold) = 4 M + 4 D (ie 1/2 at D)
BTEC L5 Electrical HND	N	Graded on basis of best 5 L5 units: M (threshold) = 5 M (from L5 units worth at least 15 credits) D (threshold) = 5 D (from L5 units worth at least 15 credits)
TCL L3 Speech Grade 8	N	N/A (single unit).
UWLQ L2 Oral Grade 4	N	N/A (single unit).
ABC L4 Art F.Dipl.	N	N/A (single unit specified as the grading unit).
ATHE L3 Business Dipl.	N	M (threshold) = 1 P + 3 M (ie 3/4 at M) D (threshold) = 1 M + 3 D (ie 3/4 at D)
Cskills L2 Carpentry Dipl.	N	Based purely on exam marks.
CIBTAC L3 Beauty Dipl.	N	Charity principle not applied. Qualification grade corresponds to lowest unit grade, which corresponds to lowest component grade.
IMI L3 Vehicle Dipl.	N	Not possible to determine from available information.

Abbreviation	D*	Degree to which charity principle operated across higher grade units (blue = charity)
NCFE L2 Bus.Skills Dipl.	Y	M = 3 P + 3 M (ie 1/2 at M) D = 3 M + 3 D (ie 1/2 at D) D* = 6 D (ie all at D)
RSL L3 Music Dipl.	Y	Roughly speaking: M (threshold) = 13/24 5-credits at M, rest at P (ie > 1/2 at M) D (threshold) = 12/24 5-credits at D, rest at M (ie 1/2 at D) D* (threshold) = 19/24 5-credits at D, rest at M (ie > 3/4 at D)
VTCT L3 Barbering E.Dipl.	Y	Roughly speaking: M (threshold) = 2 P + 6 M (ie 3/4 at M) D (threshold) = 4 M + 4 D (ie 1/2 at D) D* (threshold) = 2 M + 6 D (ie 3/4 at D)
AQA L3 App.Bus. Cert.	Y	Charity principle not applied. Qualification thresholds are simply 3 x unit thresholds. (D* is made possible, without charity at other grades, by use of UMS model, which allows unit achievement beyond the D grade (threshold) to be rewarded with UMS marks.)
OCR L3 Business T.Dipl.	Y	MM (threshold) = 4 P + 8 M (ie 2/3 at M) DD (threshold) = 8 M + 4 D (ie 1/3 at D) D*/D* (threshold) = 2 M + 10 D (ie 4/5 at D) (Straight unit Ms = 960 = qualification DM.)
OCR L3 IT T.Dipl.	Y	Roughly speaking: MM (threshold) = 3 P + 8 M (ie > 2/3 at M) DD (threshold) = 7 M + 4 D (ie > 1/3 at D) D*/D* (threshold) = 3 M + 8 D (ie > 2/3 at D) (Straight unit Ms = 192 = qualification DM.)
BTEC L3 Dental E.Dipl.	Y	MMM (threshold) = 12/36 5-credits at P, rest at M (ie 2/3 at M) DDD (threshold) = 24/36 5-credits at M, rest at D (ie 1/3 at D) D*D*D* (threshold) = 6/36 5-credits at M, rest at D (ie 5/6 at D) (Straight unit Ms = 1440 = qualification DMM.)

The **BTEC L5 Electrical HND** qualification presents an interesting case. It seems to be almost inherently based upon a charity principle, because only the best 5 (L5) qualifications are used to compute the overall qualification grade. (In theory, the remaining L4 and L5 units could all be achieved at no higher than a Pass.) However, having selected these 5 units, the overall qualification Distinction grade is based on a strict mastery principle, ie Distinction overall requires 5 unit Distinctions.

1h. High mark thresholds and (degrees of) mastery

When VTQs are assessed via external exams, mastery is sometimes given meaning via high mark thresholds, which are often specified as a percentage of the mark total. This is particularly true for pass-fail competence-based exams, which are intended to certify full competence within a domain, or subdomain, of practice. Having said that, even when it is assumed that all competent candidates ought, in theory, to be able to answer all questions correctly, competence-based exams often specify a less-than-perfect-pass-mark, eg 90%, to give candidates the benefit-of-the-doubt. This accommodates the occasional, understandable, lapse on the candidate's behalf.¹⁵

Table 5. Unit mark thresholds

	P	M	D
RSPH L4 Food Award	50%		80%
NOCN L1 ESOL Cert.B2		Unit boundaries not fixed.	
TCL L3 Speech Grade 8	65%	75%	85%
UWLQ L2 Oral Grade 4	65%	75%	85%
Cskills L2 Carpentry Dipl.	~60%	~80%	~90%
CIBTAC L3 Beauty Dipl.	60%	75%	85%
IMI L3 Vehicle Dipl.	60%	75%	85%
VTCT L3 Barbering E.Dipl.		Exam boundaries not fixed.	
AQA L3 App.Bus. Cert.		Unit boundaries not fixed.	
OCR L3 IT T.Dipl.		Exam boundaries not fixed.	

Table 5 provides information concerning the 10 sampled qualifications that included external exams. For some of these qualifications, unit grade threshold marks are not fixed; which may suggest that the qualification does not use threshold marks specifically to convey information concerning (degree of) mastery. For the remaining qualifications, the Pass threshold is specified as at least 50% and, in most instances, at least 60%.

Although 60% might be thought of as loosely indicative of mastery of the LOs associated with a unit – in the sense of ‘more mastered than not mastered’ perhaps – it is not really mastery in the sense of internally-assessed units that require all AC

¹⁵ Similarly, it accommodates the occasional lapse on the AO's behalf, should certain questions fail to function as well as they ought to.

to be satisfied for the overall unit Pass; particularly so for exams that embrace compensation via simple mark aggregation. For the units presented in Table 5, even the Distinction thresholds, which range from 80% to ~90%, would have quite a lot of accommodation built-in for lapses; if they were to be understood as providing a strong indication of mastery of the content assessed by the exam.

1i. Aggregating units of differing sizes and levels

A final issue related to aggregation, which arises with qualifications that comprise units of different sizes, and at multiple levels, is whether, and if so then how, to take this into account when grading. Once again, different qualifications within the sample adopt different approaches.

When aggregating across units of different sizes – expressed in terms of either credit or Guided Learning Hours – the typical approach, amongst the sampled qualifications, is to weight unit points according to size before aggregating them (eg **RSL L3 Music Dipl.**, **VTCT L3 Barbering E.Dipl.**, **OCR L3 IT T.Dipl.**, **BTEC L3 Dental E.Dipl.**).

Most of the sampled qualifications comprise units at the same level, obviating questions of how to aggregate across them. One exception is the **BTEC L5 Electrical HND** qualification – which comprises units at both L4 and L5 – and which determines the qualification grade purely on the basis of the highest-awarded L5 units (around 5 units from a total of around 16). This neatly controls for qualification level when awarding grades.

The **Cskills L2 Carpentry Dipl.** qualification comprises units of differing sizes, at both L1 and L2, and grades purely on the basis of 8 MCT exams. Its specification states that all units are weighted equally in the overall qualification grade.

The **ABC L4 Art F.Dipl.** qualification comprises units of differing credit sizes, at both L3 and L4. However, it sidesteps the question of whether/how to weight unit credit and level in the overall qualification grade by grading only the final, synoptic unit.

Table 6. Use of command verb progression

Abbreviation	Use of command verb progression to differentiate between candidates at different proficiency levels	Additional comments on standards
RSPH L4 Food Award	Not used to differentiate between unit grades. Some use to differentiate between L3 and L4 qualification (via AC).	Attainment descriptions identify characteristics associated with qualification grades, but these are not used to grade performances.
NOCN L1 ESOL Cert.B2	Not used.	Skills standards, derived from CEFR framework, inform item and mark scheme writing, but are not used directly for grading.
BTEC L4 Policing HNC	Strong use to differentiate between unit grades. No obvious use to differentiate between qualification levels (via AC).	
BTEC L5 Electrical HND	Strong use to differentiate between M and D unit grades; but apparently not between P and M. No obvious use to differentiate between qualification levels (via AC).	Grades determined on the basis of the highest-awarded L5 units.
TCL L3 Speech Grade 8	Not used.	Exam-Grade-specific marking schemes are the principal mechanism for differentiation, in conjunction with fixed grade thresholds. Attainment descriptions identify characteristics associated with qualification grades, but these are not used to grade performances.
UWLQ L2 Oral Grade 4	Not used.	Best-fit descriptions are provided, which are purely for illustrative purposes.

Abbreviation	Use of command verb progression to differentiate between candidates at different proficiency levels	Additional comments on standards
ABC L4 Art F.Dipl.	Not used to differentiate between unit grades (on the single graded unit). Some use to differentiate between qualification levels (via AC).	For graded unit, standards match characteristics of task performances to unit grades. Grades determined by performance on same L4 unit, whether qualification awarded at L3 or L4.
ATHE L3 Business Dipl.	Strong use to differentiate between unit grades. Some use to differentiate between qualification levels (via AC).	
Cskills L2 Carpentry Dipl.	Not used to differentiate between grades on the MCT exams (only exams graded). Strong use to differentiate between L1, L2, and L3 units (via AC).	Although all 8 MCT exams are equally weighted in the qualification grade, only 6 are at L2, and 2 are at L1.
CIBTAC L3 Beauty Dipl.	Not used to differentiate between grades on either the MCT tests or the Practical Assignment. AC within the knowledge LOs tend to be framed in terms of fairly low-level command verbs (eg describe).	Grading descriptions are provided for the Practical Assignment, but they are not used directly to grade.
IMI L3 Vehicle Dipl.	Not used to differentiate between grades on the MCT exams. Some use to differentiate between grades on the synoptic assessment Practical Task Job Report. Some use to differentiate between L2 and L3 units (via AC).	Most, but not all, of the units upon which the grading is based are at L3.
NCFE L2 Bus.Skills Dipl.	Not used to differentiate between unit grades in the primary command verb, but some use in secondary command verbs. Strong use to differentiate between qualification levels (via AC).	Standards match characteristics of task performances to unit grades (sometimes supplemented by command verb differentiation). Additional standards specified for qualification grades, but for information rather than for grading.

Abbreviation	Use of command verb progression to differentiate between candidates at different proficiency levels	Additional comments on standards
RSL L3 Music Dipl.	Not used to differentiate between unit grades. Some use to differentiate between qualification levels (via AC).	Analysis of an L2 unit with a direct L3 analogue seemed to suggest that the L2 D grading criterion was similarly demanding to the L3 M grading criterion. Standards match characteristics of task performances to task/AC grades. They identify similar grade-related distinctions across levels.
VTCT L3 Barbering E.Dipl.	Not directly used to differentiate between grades on the terminal exams; although they are specified in guidance for item writers, to ensure that items within each test target a range of task demands. Some use to differentiate between grades on practical assessments. No apparent use to differentiate between grades on the synoptic assessment. Some use to differentiate between qualification levels (via AC).	For the practical assessments, additional generic standards are specified for unit grades, but these are not used directly to grade (they are used to ensure that the grading criteria are appropriately targeted, as well as to illustrate each standard at a high level). For the synoptic assessment, best-fit standards link characteristics of task performances to grades.
AQA L3 App.Bus. Cert.	Not used to differentiate between unit grades on the unit 1 exam. Strong use to differentiate between unit grades on the unit 2 and 3 assessments.	
OCR L3 Business T.Dipl.	Strong use to differentiate between unit grades. Not used to differentiate between qualification levels (via AC/grading criteria). (Essentially the same command verbs used to differentiate between unit grades at both L2 and L3.)	

Abbreviation	Use of command verb progression to differentiate between candidates at different proficiency levels	Additional comments on standards
OCR L3 IT T.Dipl.	<p>Not used to differentiate between unit grades on the 3 exams (units 1-3).</p> <p>Some use to differentiate between unit grades on the 8 internally assessed units.</p> <p>Not used to differentiate between qualification levels (via AC/grading criteria). (Essentially the same command verbs used to differentiate between unit grades at both L2 and L3.)</p>	<p>Additional generic standards specified for qualification grades. These are not used for unit grading; but, they are used for developmental purposes, as well as to inform grade awarding on the exams.</p>
BTEC L3 Dental E.Dipl.	<p>Strong use to differentiate between unit grades.</p>	<p>Additional generic standards specified (at the domain level, across 4 domains), but these are not used as a basis for grading.</p>

Theme 2: Measurement standards

The second major theme to arise from the comparison of grading approaches concerned measurement standards – the substantive basis for differentiating levels of proficiency. Information related to this theme is summarised above, in Table 6, for each of the sampled qualifications.

Qualification standards identify the groups into which candidates are classified on the basis of qualification results, eg Pass, Merit, Distinction. They explain what it means for one candidate to have reached a significantly higher level of proficiency in the assessed domain than another candidate. The more explicitly qualification standards are specified, the greater the potential for qualification users to reach a common understanding of the standards, enabling them to interpret assessment results in the same way. Often, it is necessary to articulate qualification standards explicitly, to enable assessors to reach a common understanding of those standards, to ensure that candidates are assessed in the same way.

Qualification standards are not always specified explicitly. As noted earlier, GQ standards are often defined more by implication than by explication; in particular, when the primary aim of awarding is to ensure that grade standards this year represent the same levels of proficiency as they represented last year. Standards are also defined more by implication than by explication when candidates are classified by judging their performances relative to the range of performances observed from a known population of candidates; that is, via pure norm-referencing.¹⁶ Within the sample of qualifications scrutinised for the present study, some standards were articulated far more explicitly than others.

Qualifications that included written exam components, often multiple-choice tests, tended not to articulate standards explicitly for these components; even when standards were articulated explicitly for other components, or for the qualification overall. Comprising two written exam components, the **RSPH L4 Food Award** qualification provides little explicit articulation of measurement standards, defined simply as having reached (or exceeded) a threshold of 50% of available marks on both exams for the passing grade, or 80% for the higher grade. These standards are further explicated as follows:

“In order to be awarded a Pass, candidates must be able to recall and apply relevant knowledge and facts from some parts of the specification and demonstrate a satisfactory level of understanding of the principles and concepts used in food safety management such that the candidate will be able to satisfactorily work in the food manufacturing or related industries. The majority of answers to examination questions will contain some information of relevance.”

“In order to be awarded a Distinction, candidates must be able to recall and apply relevant knowledge and facts from the entire specification with few significant omissions and demonstrate a high level of understanding of the principles and

¹⁶ Although it is often assumed that GQs in England are norm-referenced, this view is mistaken (see Newton, 2011).

concepts used in food safety management. The majority of answers to examination questions will be correct and relevant.”

Although very general descriptions of this sort might well be of some use to qualification users, to help them to reach a common understanding of the qualification standards, they do not actually figure within the grading process for this qualification, which is a matter of simple mark addition aggregation and fixed thresholds. In essentially the same way, the appendix to the **OCR L3 IT T.Dipl.** qualification specification provides detailed qualification grade descriptions, despite all grading being conducted at the unit level, on the basis of unit-specific grading criteria. Once again, these descriptions are provided for illustrative purposes; to support qualification development, and to support grade awarding on the exam components.

The two Graded Exams demonstrate a more substantive link between grades and published grade descriptions; albeit a relatively weak link, because the standards are not used directly to award marks or grades. The **TCL L3 Speech Grade 8** qualification is based upon a single, two component (performance plus discussion) external exam. Figure 5 presents the generic grade descriptions for this qualification.¹⁷

¹⁷ Reproduced from page 39 of the 2017 Trinity College London document titled: Acting and Speaking Qualification Specifications from 2017.

Figure 5. Graded Exam grade descriptions

INTERMEDIATE (GRADES 4–5, RQF LEVEL 2)	ADVANCED (GRADES 6–8, RQF LEVEL 3)
<p>DISTINCTION (85 OR MORE MARKS) Work that demonstrates a secure, accurate and sustained response to the chosen material. A sense of spontaneity and personal involvement will be achieved through the employment of a wide range of performance skills that engage an audience effectively.</p>	<p>DISTINCTION (85 OR MORE MARKS) Work that achieves richness and a sense of total performance through a synthesis of advanced performance skills at a level of sustained excellence. Sophistication of interpretation will be demonstrated through a sense of originality and a wholly independent response to the material performed.</p>
<p>MERIT (75–84 MARKS) Work that demonstrates a considerable level of control of the material and a relatively wide range of performance skills. There will be appropriate establishment of mood and character and a level of sensitivity to the needs of an audience.</p>	<p>MERIT (75–84 MARKS) Work that demonstrates understanding through a mature and imaginative commitment to the material. This will result in a performance of some complexity, using a wide range of advanced performance skills effectively.</p>
<p>PASS (65–74 MARKS) Work that demonstrates a reasonable control of the chosen material and a range of appropriate performance skills. There will be some attempt to engage an audience and convey a sense of personal involvement in the ideas communicated.</p>	<p>PASS (65–74 MARKS) Work that demonstrates some evidence of mature understanding presented with an element of identification with the material. Performance skills are integrated and used with some assurance.</p>
<p>BELOW PASS (64 AND BELOW MARKS) Work that may show lack of preparation and control of performance skills (even though some may be evident). At best, the performance may be sporadic in its attempt to communicate effectively.</p>	<p>BELOW PASS (64 AND BELOW MARKS) Work in which some skills may be evident but they are insufficiently integrated or contain significant lapses in technical achievement. There may be inadequate preparation and inappropriate response to the material.</p>

Within these descriptions, differentiation between grades at L3 is specified in terms of qualities such as:

- richness, sophistication, originality (Distinction);
- maturity, imagination, complexity (Merit);
- identification, assurance (Pass); and
- lapses, inappropriateness (Below Pass).

Although these descriptions are not used to judge exam performances directly, the distinctions that they draw are operationalised via detailed marking schemes for each component. Notice that these Graded Exams only explicitly specify standards at the Grade Band level (eg L3, covering Grade 6, 7 and 8). The standards are further delineated (at the Grade Band level) via expectations concerning the language, context, duration and complexity of exam performances.

2a. Description-related grading approaches

The best example of a description-related grading approach, within the sample, is provided by the graded synoptic assessment component of the **VTCT L3 Barbering E.Dipl.** qualification. This component is graded directly using an holistic – and explicitly compensatory – grading rubric, which specifies characteristics for Pass, Merit, and Distinction. The first sentence of the rubric, for each grade, provides a sense of how this qualification understands differing levels of proficiency in barbering:

- The learner carried out a complete barbering service which satisfied the essential requirements of the client and mostly aligned with the consultation. (P)
- The learner systematically carried out the agreed services which fully aligned with the consultation and employer expectations. (M)
- The learner systematically and seamlessly integrated services, which fully aligned with the consultation and exceeded employer expectations. (D)

The qualification specification states that the characteristics mentioned in the grading rubric focus specifically on the grading of those elements most valued by employers.

The defining characteristic of a description-related grading approach, like this, is that grades are awarded to candidates on a ‘best-fit’ basis; that is, the grade awarded to a candidate corresponds to the generic description – for below Pass, Pass, Merit, or Distinction – that best describes the overall quality of their task performance.

Although best-fit grading approaches are inherently compensatory – allowing better performances in one area to compensate for worse performances in another – they are perhaps better described as operating on the aggregatory principle of **configuration**; whereby grading relates to the overall pattern of knowledge and skill

demonstrated – its configuration – rather than to the ‘average’ or ‘sum total’ of that knowledge and skill (Sadler, 1987).

2b. Criterion-related grading approaches

Most of the sampled qualifications adopt a criterion-related approach to grading internally assessed units/assessments. Although a variety of different approaches were observed, they all linked the award of grades to detailed criteria, set out as a list for Merit and a list for Distinction. Often, these grading criteria for Merit and Distinction mapped directly onto the AC that had been specified for a unit, thus characterising greater proficiency at the level of individual AC (see Figure 3, above). However, a number of qualifications specified different kinds of criteria for Merit and Distinction, distinguishing them clearly from the AC that had been specified for the unit in question (see Figure 4, above). Two quite different criterion-related grading approaches were identified: those that differentiated primarily in terms of performance quality; and those that differentiated primarily in terms of performance complexity.

2bi. Performance quality criterion-related grading

The clearest example of a performance quality criterion-related grading approach came from the **RSL L3 Music Dipl.** qualification. The full set of grading criteria for Unit 344 can be found in Annex 5. Key extracts are tabulated below, in Table 7.

Table 7. Extract from grading criteria chart for Unit 344

	Pass	Merit	Distinction
AC 1.1	Explain the marketing strategy...	Explain with clarity and detail...	Comprehensively explain...
AC 1.2	Implement a promotional campaign...	Implement a clear and detailed...	Implement a rigorous...
AC 1.3	Produce a Promotion E-Portfolio...	Produce a clear and detailed...	Produce a comprehensive...
AC 1.4	Evaluate [...] by creating a career focused action plan...	Evaluate [...] by creating a clear and detailed...	Evaluate [...]by creating a perceptive and rigorous...

For each unit, within this qualification, the Pass criteria simply repeat the AC. Unit 344 has 4 AC, corresponding to a single LO. As mentioned earlier, the AC for this qualification tend to be associated with specific tasks, such that each task/AC can be graded in terms of performance quality. Note, from Table 7, how simple explanation is all that is required to Pass 1.1, whereas explanation with clarity and detail is required for Merit, while comprehensive explanation is required for Distinction. The remaining AC are differentiated in essentially the same way.

The unit-level aggregation procedure for this qualification is somewhat unusual – when compared with similar qualifications within the sample – as the LO/unit grade corresponds to the ‘average’ of the task/AC grades. The fact that this grading approach is designed to reward general performance quality, in relation to each task/AC, may help to provide a rationale for the ‘averaging’ approach.

Other sampled qualifications that adopt a performance quality criterion-related grading approach include the **ABC L4 Art F.Dipl.** qualification and the **NCFE L2 Bus.Skills Dipl.** qualification; although each operates their approach slightly differently. Figure 6, reproduces the first 8 AC from **NCFE L2 Bus.Skills Dipl.** Unit 5.¹⁸

From Figure 6, it is clear that differentiation between grades is achieved on the basis of performance quality criteria such as detail, thoroughness, and confidence. In contrast to Unit 344 (**RSL L3 Music Dipl.**) from Table 7, Unit 5 (**NCFE L2 Bus.Skills Dipl.**) does not specify criteria for all AC at all grades, although units for this qualification typically do specify criteria for all AC at all grades.

A general challenge for this approach to grading is that it provides no detail on what distinguishes performance quality at one level from performance quality at another. For instance, from Table 7, what is it that distinguishes a comprehensive explanation (D) from a clear explanation (M) or from a mere explanation (P)? Similarly, from Figure 6, what is it that distinguishes a thorough explanation (Distinction) from a detailed explanation (Merit) or from a mere explanation (P)?

Finally, notice one significant feature of both of the examples provided (Figure 6 and Table 7). For each assessment criterion, exactly the same command verb is used to preface the grading criterion for each grade, eg ‘describe’ is used at all grades in criterion 1.2 (Figure 6), and ‘evaluate’ is used at all grades in criterion 1.4 (Table 7). This is completely the opposite strategy from that adopted in performance complexity criterion-related grading, which we shall now turn to.

¹⁸ Reproduced from page 47 of the 2017 NCFE document titled: Qualification Specification. NCFE Level 2 Diploma in Skills for Business (601/2624/3). Issue 2 November 2017.

Figure 6. Part of grading criteria chart for Unit 5 (highlighting added)

Assessment criteria	Pass	Merit	Distinction
1.1 Outline own job role	The candidate will outline own job role	The candidate will outline own job role in detail	No Distinction for this AC
1.2 Describe how own role fits into their team's structure	The candidate will describe how own role fits into their team's structure	The candidate will describe in detail how own role fits into their team's structure	The candidate will comprehensively describe how their own role fits into their team's structure
1.3 Outline why it is important to work with others to achieve objectives	The candidate will outline why it is important to work with others to achieve objectives	The candidate will outline, in detail , why it is important to work with others to achieve objectives	No Distinction for this AC
1.4 Agree own tasks for a range of team objectives	The candidate will agree own tasks for a range of team objectives	No Merit for this AC	No Distinction for this AC
1.5 Carry out own tasks in line with agreed systems and procedures	The candidate will carry out own tasks in line with agreed systems and procedures	The candidate will confidently carry out own tasks in line with agreed systems and procedures	The candidate will carry out own tasks in line with agreed systems and procedures, with confidence and demonstrating advanced skills
1.6 Seek guidance from others where appropriate	The candidate will seek guidance from others where appropriate	The candidate will seek timely guidance from others where appropriate, showing initiative	No Distinction for this AC
2.1 Explain why it is important to recognise individual strengths within a team	The candidate will explain why it is important to recognise individual strengths within a team	The candidate will explain in detail why it is important to recognise individual strengths within a team	The candidate will give a thorough explanation of why it is important to recognise individual strengths within a team
2.2 Identify individual strengths within their own team	The candidate will identify individual strengths within their own team	The candidate will identify individual strengths within their own team, demonstrating critical understanding	No Distinction for this AC

2bii. Performance complexity criterion-related grading

A good example of a performance complexity criterion-related grading approach came from the **BTEC L3 Dental E.Dipl.** qualification. Figure 7 reproduces the first 4 AC from Unit 4.¹⁹

Figure 7. Part of grading criteria chart for Unit 4 (highlighting added)

Assessment and grading criteria		
To achieve a pass grade the evidence must show that the learner is able to:	To achieve a merit grade the evidence must show that, in addition to the pass criteria, the learner is able to:	To achieve a distinction grade the evidence must show that, in addition to the pass and merit criteria, the learner is able to:
P1 explain how elements used in dental materials can be categorised using the periodic table, including electronic structure and bonding [IE1, SM2,3]	M1 discuss how a dental technician can use the periodic table to assist in the determination of properties for dental materials used in the fabrication of dental appliances	
P2 explain the importance of the chemical, biological, and mechanical properties in the selection of materials used in dental technology	M2 discuss the key points relating to chemical, biological, physical and mechanical properties of dental materials	D1 analyse the properties of a range of dental biomaterials, justifying their selection
P3 explain the use of dental waxes found within the dental laboratory	M3 review the properties of dental waxes	D2 evaluate waxes used in the dental laboratory
P4 explain the selection criteria applied to assess the suitability of using either gypsum or synthetic stone materials in the dental laboratory [IE1]	M4 discuss the use of gypsum and synthetic stone materials used in the dental laboratory	D3 critically evaluate gypsum and synthetic stone materials used in the dental laboratory

Contrary to the approach discussed above, grades for units from this qualification are not differentiated primarily on the basis of performance quality, but on the basis of performance complexity, expressed in terms of a progression in command verb challenge.

Notice, from Figure 7, how higher grades are more challenging than lower grades in terms of the cognitive demands that are made of candidates, in relation to essentially the same criterion:

¹⁹ Reproduced from page 94 of the 2016 Pearson Education Limited document titled: Pearson BTEC Level 3 Diploma and BTEC Level 3 Extended Diploma in Dental Technology. Specification. First teaching September 2014. Issue 2.

- 'explain' (P);
- 'discuss', 'review' (M); and
- 'analyse', 'evaluate', 'critically evaluate' (D).

This progression in command verb challenge derives from Bloom's *Taxonomy of Educational Objectives*, discussed earlier.²⁰ It indicates that, for the award of higher grades, candidates must be operating at a qualitatively higher cognitive level, on essentially the same content. This is particularly true for qualifications, like **BTEC L3 Dental E.Dipl.**, that implement a strong version of this approach, where command verbs are the primary mechanism of differentiation across grades, and where particular command verbs are associated consistently with particular grades (see also **BTEC L4 Policing HNC**, **ATHE L3 Business Dipl.**, **AQA L3 App.Bus. Cert.**, **OCR L3 Business T.Dipl.**).

2c. Performance complexity criterion-related grading vs. levelling

A general challenge for this approach to grading is that it re-purposes an approach that has traditionally been used to 'level' qualifications. This is challenging because these levels are still a defining characteristic of all regulated qualifications. The qualifications that adopted performance complexity criterion-related grading appeared to navigate this challenge in different ways (or not to navigate it explicitly).

For some qualifications, the dominant use of command verb differentiation is in relation to grading. The **OCR L3 Business T.Dipl.** qualification provides a good example of this. This diploma is also provided at Level 2, although the L2 units contain different content from the L3 units, so cannot be compared directly (eg at the individual AC level). Just as for the L3 diploma, the L2 diploma grading criteria operationalise a strong version of this approach. In fact, as illustrated in more detail in Annex 5, the command verbs used to differentiate grades at L2 are essentially the same as the command verbs used to differentiate grades at L3, which include:

- 'describe', 'interpret', 'explain' (P);
- 'analyse' (M); and
- 'evaluate' (D).

This would seem to imply that 'grade-ness' and 'level-ness' are somewhat independent of each other, and that: 'level-ness' is primarily a matter of content demand; while 'grade-ness' is primarily a matter of cognitive demand (interacting with content demand).

For other qualifications, the dominant use of differentiation via a progression in command verb challenge appeared to be in relation to levelling. The **Cskills L2 Carpentry Dipl.** provides a good example of this. Note that this qualification awards

²⁰ Although the criteria for this unit are framed primarily in terms of cognitive demands, Bloom also identified the affective domain and the psychomotor domain, both of which can be expressed (via command verbs) in terms of hierarchical demands.

grades purely on the basis of performance across the suite of MCT 'knowledge test' exams; so it does not actually specify grading criteria at all. It does, however, specify LOs and AC for all 8 units. Because this qualification contains units at L1 and L2 – and the L3 Diploma contains units at L1, L2 and L3 – it is fairly straightforward to compare the use of command verbs across levels. As illustrated in more detail in Annex 5, it appears that this qualification relies strongly upon command verbs to differentiate units (and therefore AC) at different levels, for example:

- 'identify', 'state', 'list' (L1);
- 'describe', 'interpret', 'state', 'define', 'compare' (L2); and
- 'detail', 'explain', 'compare', 'define', 'evaluate', 'analyse' (L3).

The **ATHE L3 Business Dipl.** qualification appeared to use command verb differentiation both in relation to grading and in relation to levelling. For instance, across its mandatory units:

- 'evaluate' is only ever used within grade criteria for Distinction;
- 'analyse' is mainly used within grade criteria for Distinction, although sometimes also for Merit; whereas the
- grading criteria for Pass (ie the unit AC) tend to involve verbs like 'state', 'describe', 'explain'.

Additionally, when compared with AC from the ATHE Level 4/5/6 Diplomas in Business and Administrative Management, it appeared that the L3 AC (from **ATHE L3 Business Dipl.**) were pitched at a slightly lower level than L4 and L5 AC (from the higher diplomas) which went beyond 'describe' and 'explain' to include 'analyse' and 'evaluate'. The L6 criteria included even more sophisticated command verbs, such as 'critically evaluate' and 'critically appraise'.

Although, some qualifications consistently associate certain command verbs with particular grades, and some with particular levels, this consistency within qualifications is not necessarily evident across qualifications. This hints at an element of arbitrariness in the use of command verbs within the sector. For instance, 'evaluate' tends to be reserved for use as a Distinction grade L3 criterion amongst the sampled qualifications that operate a strong, performance complexity approach to grading (eg **BTEC L3 Dental E.Dipl.**, **OCR L3 Business T.Dipl.**, **ATHE L3 Business Dipl.**). Yet, for qualifications that operate a strong, performance quality approach to grading (eg **RSL L3 Music Dipl.**), or that do not use grading descriptions or criteria at all (eg the L3 version of the **Cskills L2 Carpentry Dipl.**), 'evaluate' seems as likely to be used as a Pass grade L3 criterion (ie to feature within the AC for any particular unit).

Finally, it is worth consulting Annex 5 for details of how the **NCFE L2 Bus.Skills Dipl.** qualification combines a performance quality approach with a performance complexity approach, for certain of its units. This is an interesting example, because the L3 diploma is structurally equivalent to the L2 diploma, making cross-unit comparisons straightforward. For Unit 10, the hybridised approach to grading

appears almost to link the criterion for L2 Distinction to the criterion for L3 Merit (on this unit, at least).

Discussion

The most interesting finding from this exploratory survey of grading practices across regulated VTQs in England was the striking variety of approaches observed. This seems all the more striking given the small scale of the research, which considered just 18 qualifications, across 15 awarding organisations (cf the 2167 available VTQs that were classified as awarding higher grades, as at 28/11/17). It seems reasonable to conclude, on the basis of this evidence alone, that current practice in grading VTQs in England is not underpinned by a straightforward, generally accepted, set of principles governing good practice. Conversations with colleagues from each of the AOs represented in the sample supported this conclusion. That does not imply that any of the observed grading approaches reflects anything less than good practice. A fairer conclusion is that it remains far from obvious what set of principles ought to underpin grading practices for VTQs in England; and, in particular, whether the same principles ought to underpin grading practices across such a wide variety of qualifications.

Grading CBAs

Although grading practices differed widely across the 18 sampled qualifications, various characteristics were clearly recognisable across many of them; in particular, characteristics that appeared to derive from the model of Competence-Based Assessment (CBA) that was originally associated with National Vocational Qualifications (NVQs). As well as being grounded in the idea of competence, which implies a clear link to professional or occupational standards, hallmarks of CBA include:

- the atomistic specification of measurement standards in terms of learning outcomes and assessment criteria;
- a mastery measurement model, meaning that a certificate of competence could be interpreted to mean competent across each and every learning outcome and assessment criterion; and
- assessment based on the exhaustive sampling of learning outcomes and assessment criteria.

Over the course of the past three decades, since the introduction of NVQs, characteristics such as these have become increasingly associated with VTQs in England. Indeed, during the 2000s, they were deemed so fundamental to high quality assessment that they were specified as regulatory requirements for any qualification within the Qualifications and Credit Framework (QCF); alongside a plethora of related requirements and expectations concerning qualification levels, differentiation of attainment, and so on. Some of these requirements and expectations went far beyond the core characteristics associated with CBA; such as the very specific expectation that differentiation across qualification levels could be understood and managed in terms of the 'hierarchy' of behavioural complexity elaborated in Bloom's *Taxonomy of Educational Objectives*.

As indicated in Annex 4, 11291 of the qualifications that were available to candidates on 28/11/2017 were classified as (former) QCF qualifications. This represents 64% of all available qualifications at that point in time. Furthermore, many of the non-QCF qualifications from the present study also bore some, or all, of the hallmarks of CBA. This suggests that the influence of CBA on regulated VTQs in England may be extremely widespread. Indeed, it does not seem unreasonable to conclude that the set of characteristics described above – atomistic specification, domain mastery, and exhaustive sampling – may have become something of a ‘default template’ for designing VTQs in England, ie a model that VTQ designers will often default to nowadays, unless a qualification obviously demands an alternative approach, or unless an alternative approach is specifically mandated by government or the regulator (eg performance table qualifications, Functional Skills Qualifications).

More interestingly, though, while the core characteristics of CBA have been appropriated within many regulated VTQs in England – as designers have responded to the requirements and expectations established by a variety of social, political, and regulatory forces – insights from the present study suggest that this appropriation may also have been highly pragmatic at times. It became apparent that these characteristics were sometimes appropriated only partially, and sometimes only superficially, as qualification designers prioritised the need to ensure that their qualifications actually worked in specific contexts, for specific populations, and for specific purposes. The challenges presented by grading are particularly significant in this respect, since grading directly contradicts what Alison Wolf described as one of six key ‘requirements’ for CBA in the NVQ system:

Competent/not yet competent judgements only. Only two judgements can be made: either the person has consistently demonstrated workplace performance which meets the specified standards or they are not yet able to do so – ‘competent’ or ‘not yet competent’. Grading is rejected – the idea being that someone either has or has not reached the level required by a holistic model of competence. How individuals perform in comparison to others is irrelevant.

(Wolf, 1995, p.22)

In order to respond to contexts, populations, and purposes that require grading, qualification designers have had to find ways of making their qualifications work; including ways that directly challenge certain of the principles associated with CBA, especially the idea of domain mastery. Thus, grading foregrounds issues related to the adequacy and appropriateness of CBA, at least in its purest form, as a general model for qualification design in England. Consequently, the present report raises a general question concerning the influence of CBA upon qualification design in England; as well as more specific questions concerning VTQ grading practices.

The following two sections consider a number of issues arising from the study. The first has a shorter-term perspective, highlighting technical issues, and raising questions concerning present practices. The second has a longer-term perspective, highlighting conceptual issues, and raising questions concerning future practices.

Technical issues in grading VTQs

This section focuses on how grading is currently practiced, based on evidence from a small sample of regulated qualifications. It identifies a number of technical issues related to their effective operation.

Standardisation

Standardisation is the process of ensuring that assessor judgements are effectively calibrated, ie that assessor judgements always embody the same measurement standard(s). This is a challenge whenever assessors are required to judge task performances – whether the assessment is internal or external. With external assessment, the challenge is typically to ensure that all assessors apply a common mark scheme, in the same way, when judging responses to a common task. With internal assessment, it is more likely that there will be neither a common task nor a common mark scheme. In the absence of an externally-set task, internal assessment faces the additional challenge of ensuring that centre-devised tasks are effective in eliciting evidence of the proficiency that needs to be assessed. Yet, even when an AO provides centres with common tasks, the challenge of standardising assessor judgements remains. The three main approaches to grading identified earlier – description-related, performance quality criterion-related, and performance complexity criterion-related – each present different standardisation challenges.

The challenges presented by description-related approaches (as adopted within the **VTCT L3 Barbering E.Dipl.** synoptic assessment, for instance) are reasonably well understood. They are essentially the same as those presented by what Ahmed and Pollitt (2011) described as ‘unconstrained levels-based mark schemes’. Ahmed and Pollitt argued that mark schemes like this are most effective: when they describe levels of response to the task that was actually set, rather than to a generic one; and when they make reference to the essential ‘trait’ that the assessor is supposed to be judging. In other words, they are most effective when they explain how an assessor is likely to see differing levels of proficiency being manifested in candidates’ responses to the task in hand.

The challenges presented by performance quality criterion-related approaches are quite similar, in principle. Typically, for each criterion, the assessor’s task is to allocate each candidate to one of four levels: Below Pass, Pass, Merit, or Distinction. However, in practice, the ‘grade scheme’ is likely to be extremely limited, even when referenced to a common, externally-set task. That is, rather than there being a detailed, levels-based ‘grade scheme’ for each criterion – providing a thorough explanation of how differing levels of proficiency are likely to be manifested – there will typically be a simple set of quality statements for each criterion. For instance, referring back to Table 7, the four levels of proficiency on AC 1.1, from **RSL L3 Music Dipl.** unit 344 – ‘Explain the marketing strategy for an agreed product including a range of physical and digital promotional devices’ – are characterised by the following four statements:

1. provide insufficient evidence to explain the marketing strategy for an agreed product, giving an insufficient range of physical and digital promotional devices (Unclassified);

2. explain the marketing strategy for an agreed product including a range of physical and digital promotional devices (Pass);
3. explain with clarity and detail the marketing strategy for an agreed product including a range of physical and digital promotional devices (Merit); and
4. comprehensively explain the marketing strategy for an agreed product including a range of physical and digital promotional devices (Distinction).

If all that is available to an assessor, in relation to each AC, is a minimally detailed set of performance quality statements, then this presents a substantial challenge to standardising criterion-level grading judgements across assessors, or within assessors from one candidate to the next. In the above example, for instance, what is it that distinguishes a clear and detailed explanation from a simple explanation? And, more importantly, how will assessors come to judge this criterion in the same way?

Incidentally, the underlying similarity between description-related approaches and performance quality criterion-related approaches raises the question of whether performance quality criterion statements – such as the four provided above – are better understood as best-fit descriptions, rather than as clear-cut criteria. This, incidentally, raises a more general question concerning the nature and function of criteria within CBA; that is, whether (as a set) they can ever be said to precisely define the measurement standard, or whether their nature and function is simply to exemplify that standard. The issue, here, relates to the limits of precision of linguistic specification (see Wolf, 1995; Oates, 2004; Lum, 2013; and many other scholars).

The challenges presented by performance complexity criterion-related approaches are somewhat different. Although they are likely also to be presented as grading grids (in which the grading criteria for each AC are specified as a simple set of statements) the standards are framed in terms of complexity statements rather than quality statements. This time referring back to Figure 7, the four levels of proficiency on AC 1.1, from **BTEC L3 Dental E.Dipl.** unit 4, are characterised by the following three statements (the first level of proficiency remaining unspecified in this example):

1. explain the use of dental waxes found within the dental laboratory (Pass);
2. review the properties of dental waxes (Merit);
3. evaluate waxes used in the dental laboratory (Distinction).

In this example, the distinction between levels refers to a qualitatively different kind of performance, in which the candidate has responded successfully to a qualitatively different command verb. Conceivably, this might make it easier for assessors to allocate candidates to grades consistently; on the assumption that these statements function somewhat more like clear-cut criteria than best-fit descriptions.

From this perspective, it would seem to be assessors' understandings of command verb meaning that need to be standardised. Many AOs publish official definitions of their command verbs, to support consistent understanding and application. These

are sometimes generic, and intended to be applied across a range of qualifications; other times they are published for specific qualifications. For instance, the following definitions of 'describe', 'analyse' and 'evaluate' are published (alongside others) for the Pearson LCCI Level 4 Certificate in Organisational Behaviour and Performance:²¹

- Describe – Giving an account of something including a series of features/points/trends/factors.
- Analyse – Examine in detail to discover the meaning or essential features of a theme, topic or situation. Break something down into its components, examine factors methodically and in detail to recognise patterns by applying concepts and making connections to predict consequences.
- Evaluate – Examine in detail the meaning or essential features of a theme, topic or situation; break something down into its components; examine factors methodically and in detail, identify separate factors, say how they are related and how each one contributes to the topic to make reasoned judgements and conclusions.

Similarly, City & Guilds has published a set of definitions for its suite of Technical Qualifications, which includes:²²

- Describe a... – Write what something is like – usually what it looks, tastes, feels, sounds like.
- Describe the process for... – Give the steps in a process.
- Describe the effect of (for example an event) on... – Write about what has changed/happened because of... (for example, an event).
- Analyse – Study or examine usually a complex issue in detail to identify essential elements, causes, characteristics.
- Evaluate – Make an analysis about the success/quality of for example end product/outcome – usually systematic, proposing improvements.

Alongside these definitions, City & Guilds notes that these command verbs will typically be indicated by particular qualities:

- Describe – Correctness, completeness, relevance.
- Analyse – Making links, cause and effect, drawing information together for a purpose, coherence, logic.

²¹ Accessed from: https://qualifications.pearson.com/content/dam/pdf/LCCI/lcci-online-training-events/accounting/organisational-behaviour-and-performance/Command_verbs.pdf

²² Accessed from: https://www.cityandguilds.com/~/_media/cityandguilds-site/documents/technical-qualifications/deliver/j392637cg1419teachinglearningassessmentguide%20pdf.ashx

- Evaluate – Compared against standards, systematic, evidenced, thorough, quality.

Yet, even at this level of specificity, there is still an important judgement that each assessor is required to make, concerning the *degree to which* any particular task response (or set of task responses) can be classified as a *successful* description, analysis, or evaluation. In other words, how correct, complete, and relevant does a response need to be, to be classified as a successful description? And how integrated, coherent, and logical does a response need to be, to be classified as a successful analysis? Clearly, there are significant standardisation challenges for any such judgement call.

An important question regarding the role of command verbs within performance complexity criterion-related grading approaches concerns how literally they are intended to be understood. Interpreted most literally, the assessor's role is to judge the evidence provided for each criterion in terms of whether it constitutes a genuine case of 'identification' or a genuine case of 'description' or a genuine case of 'evaluation' or suchlike. From this most-literal perspective, the overall unit grade can be interpreted as meaning that a candidate has demonstrated a certain level of cognitive complexity in the domain given by its LOs. That is, they are able to engage with the unit's LO(s): (purely) at the level of identification; or at the level of description (beyond identification); or at the level of evaluation (beyond description). From a standardisation perspective, the command verb is the primary focus, here, and the LOs are essentially contexts through which these command verbs are demonstrated.

Interpreted least literally, the assessor's role is to judge the evidence provided for each criterion in terms of an anticipated progression in learning for each LO. From a standardisation perspective, the anticipated progression in learning is the primary focus, and the command verbs are essentially illustrative 'markers' intended to capture something (but *only* something) important concerning the level of expertise that has been acquired in the micro-domain given by that criterion/LO. From this least-literal perspective, the overall unit grade can be interpreted to mean that a candidate has demonstrated a certain level of expertise in the domain given by its LO(s).

The standardisation challenge for the least-literal perspective is likely to be greater than for the most-literal perspective, assuming that its use of command verbs is more illustrative than definitive. Rather than being able to spot 'an evaluation' as distinct from 'a description' the assessor needs to be able to distinguish between gradations of expertise, for each criterion, which are illustrated (but not defined) by command verbs. As such, the standardisation challenge more closely resembles that which faces the performance quality criterion-related grading approach; where the transition from one grade to the next is more continuous and harder to capture in words. Indeed, once again, we might ask whether the performance complexity criterion statements are better understood as best-fit descriptions, rather than as clear-cut criteria, according to this least-literal perspective.

A final technical issue concerns the impact of a lack of standardisation of criterion-related grading judgements; which presents more of a threat when it is associated with a strong mastery approach to aggregating assessment information. For

qualifications that operate a strong mastery requirement at the unit level – such that, for example, all Merit criteria must be satisfied for the award of Merit overall – lack of standardisation can be particularly serious. This is because unsystematic errors of judgement do not ‘average out’ when aggregated, as tends to happen when aggregation is compensatory. To put it simply, under a strong mastery requirement, the unit grade is determined by the lowest grade achieved across all AC within a unit. So, if a learner is rated as Merit for 5/6 criteria, and Pass for 1/6, then their overall grade will be Pass. Yet, what if that Pass judgement happened to have been wrong, and should have been a Merit after all? Conversely, a compensatory approach would have awarded Merit overall, reflecting the modal average AC judgement; thereby allowing for the possibility of the odd erroneously low (or high) judgement. A strong mastery approach makes no such allowance, meaning that each individual AC judgement needs to be accurate in its own right.

Grading and levelling

As noted earlier, Annex 2 contains a hypothetical example (from the *Guidelines for writing credit-based units of assessment for the Qualifications and Credit Framework*) of a unit titled Customer Service Skills, which was to be offered at both Level 1 and Level 2. The important point to notice, from this example, is the implication that it might be entirely legitimate to differentiate standards, across qualifications at adjacent levels, purely in terms of performance complexity; with content remaining essentially the same.

When a suite of qualifications is designed at multiple, adjacent levels – and when standards are differentiated across levels more in terms of performance complexity (AC command verbs) than in terms of differential content (of LOs) – the idea of grading each of those qualifications raises a major technical challenge. The challenge of specifying sets of criteria that are sufficiently distinct from each other to support consistent and reliable assessment judgements, without ‘encroaching’ on the standard of units at a higher level of the QCF, was recognised in the QCF regulations and guidance on grading (see Ofqual, 2008, p.16; QCDA, 2010c, p.10). Yet, it is far from obvious how best to address this challenge, especially when qualifications at adjacent levels have similar LOs. When AOs rely heavily on command verbs to differentiate both across levels and across grades, there is a strong likelihood of encroachment; the more similar the LOs, from one level to the next, the more likely this encroachment will be.

An issue for further consideration is the apparent arbitrariness – when looking across qualifications and across AOs – with which particular command verbs are used to signify either the standard associated with particular levels or the standard associated with particular grades. For instance, there seemed to be a tendency for qualifications that made strong use of command verbs to differentiate across grades to use lower-level command verbs for their passing grade, when compared with qualifications that made strong use of command verbs to differentiate across levels. This raises questions of comparability, which are discussed in more detail below.

Comparability

Questions of comparability arise when identical grades (eg Pass, Merit, Distinction, Distinction*) are awarded from similar qualifications – either within or across AOs – that are located at the same qualification level. Such questions are most significant when qualifications with exactly the same title are awarded by different AOs. Any difference in grading approach, across those AOs, might constitute a potential threat to the comparability of standards; for instance:

- when adopting quite different approaches (eg performance quality criterion-related vs performance complexity criterion-related);
- when adopting essentially the same approach (eg performance complexity criterion-related), but specifying
 - quite different command verbs for each grade;
 - quite different numbers of Merit or Distinction criteria;
- when applying different aggregation models, both within and across units (mastery, compensatory, configural, charitable).

Different grading structures also raise questions of comparability. For instance, amongst the sampled qualifications that awarded a D* grade, there was a tendency to be quite charitable at qualification grade D, at least partly attributable to the need to make 'headroom' for D* at qualification level when only awarding D at unit level. This raises a question concerning the general meaning of qualification grade D in P-to-D structures, when compared with P-to-D* structures.

Unfortunately, the greater the differences between grading approaches, across any two qualifications, the harder it becomes to anticipate and/or investigate potential threats to the comparability of standards. This is particularly so when qualifications operate a variety of aggregation principles, but in differing ways. For instance, if a qualification were to operate a strict mastery principle for Merit and Distinction grades at the unit level, but was fairly charitable at the qualification level, then how might that be compared – from a comparability of standards perspective – with a similar qualification that operated a compensation principle for Merit and Distinction grades at the unit level, but that did not operate a charity principle at the qualification level? Is it even reasonable to raise questions of comparability, or lack of comparability, when qualifications have been designed according to very different aggregation models? Perhaps it is, for qualifications that award the same grades, under the same qualification title, at the same level? But, perhaps not, for qualifications that do not share exactly the same title? Answers to questions like these are not at all obvious.

Issues of comparability also arise within qualifications when candidates are able to choose from a range of optional units. If it is structurally easier to achieve a grade on one unit than on another, then this will threaten the comparability of standards for candidates who have taken different routes to the overall qualification grade. One

such structural difference might be the number of criteria specified for each grade for each LO (even when the number of LOs is constant across optional units).

Weighting

A specific technical challenge that arises when qualifications comprise units of different sizes, and at multiple levels, is how to account for this when grading, if at all. For CBAs that only award the passing grade, and that operate on a strict mastery principle, this challenge does not arise. Each unit simply operates as a hurdle, regardless of its size or level. However, when performances on units/components are differentiated via grades, this raises the question of whether performances on higher-level or larger-size units/components ought to receive a higher weighting in the aggregation model than performances on lower-level or smaller-size units/components. Amongst the sampled qualifications, a variety of approaches was observed: some weighting all units/components equally; others applying complex weighting formulas.

When a decision has been made not to grade all units/components within a qualification, this is tantamount to a decision concerning weighting. In other words, units/components that only award the passing grade – which are then treated simply as hurdles, ie prerequisites for passing the qualification – thereby receive a zero weighting when determining higher qualification grades. Technically, we would hope that the decision to zero-weight a unit/component would have some principled basis, rather than being purely pragmatic. In other words, we would hope that units/components with the potential to capture information concerning the most important differences between candidates should not be zero-weighted *simply* because they are less easy to grade, in comparison with other units/components within a qualification.

Similarly, we would hope that the distribution of Merit and Distinction criteria across LOs, for criterion-related grading approaches that do not specify Merit and Distinction criteria for all LOs/AC, would also have some principled basis. In other words, we would hope that higher unit grades are awarded on the basis of LOs/AC with the potential to capture information concerning the most important differences between candidates. A more arbitrary approach to distributing Merit and Distinction criteria across LOs might risk determining unit grades on the basis of information concerning the least important differences between candidates, eg differences on the smallest LOs, or differences on LOs that are least fundamental to the domain.

Burden and backwash

One reason for not writing Merit and Distinction criteria for all AC – within a criterion-related grading approach – is the potential impact that each higher grade criterion will have on the overall assessment burden of a unit, when higher grade criteria require additional assessment activities. This is not an issue for performance quality criterion-related approaches, since grades are awarded on the basis of how well a candidate responds to the demands of the task associated with each AC; that is, higher grades are associated with higher quality performances, in response to exactly the same task that was set for the passing grade. However, it is potentially an issue for performance complexity criterion-related approaches, where there is an expectation that evidence sufficient to demonstrate (say) a ‘description’ of X is qualitatively different from, and additional to, evidence sufficient to demonstrate

(say) an 'evaluation' of X; and, more specifically, where these correspond to qualitatively different tasks.

Across the sampled qualifications, AOs that operated performance complexity criterion-related grading approaches differed in relation to the assumption that higher grades might require additional assessment activities. ATHE, for instance, provides 'extension activities' for higher grades, and these higher grade tasks present additional demands to candidates. With a view to the potential impact upon the overall assessment burden of a unit, ATHE intentionally restricts the number of Merit and Distinction criteria specified. Pearson and OCR, by way of contrast, do not anticipate the provision of additional activities, presuming that higher grade criteria can be demonstrated within essentially the same activity that is set for the passing criteria. Having said that, their grading grids still tend to leave certain AC without corresponding criteria for Merit and Distinction, reflecting a recognition that certain AC are less suitable for reconfiguration at a higher level of cognitive complexity.

The flipside of the potential impact upon burden, which arises when additional assessment activities are specified for higher grades, is the potential for some learners to choose to tackle the qualification at a lower level of cognitive complexity than others. They would do so by choosing only to provide evidence sufficient to demonstrate (say) a 'description' of X, as opposed to (say) an 'evaluation' of X. By adopting the same strategy across all AC, all LOs, and all units, a learner would therefore be aiming to maximise their chances of passing the qualification, whilst sacrificing their chances of achieving a higher grade. Whether this potential impact upon teaching and learning is, on balance, more positive than negative is open to debate.

Transparency

A final technical issue concerns the transparency of the information that is provided to qualification users by candidates' results; that is, the information about a candidate's proficiency that is conveyed by their qualification result, which is then used to make a decision (eg whether or not to hire that candidate).

One of the principles underlying CBA is that qualification results ought to be perfectly transparent concerning a candidate's proficiency; in the sense that a candidate who passes should have met the competence threshold for every AC, on every LO, on every unit. As long as a qualification user can access documents that specify the LOs/AC for each unit, this should provide them with precise information concerning the minimum level of competence of any candidate with a Pass. In other words, under CBA, the qualification Pass, alone, is sufficient to convey all of this information.²³ Of course, it conveys no information about differences between candidates who have met or exceeded the minimum level of competence; which is where grading becomes useful.

Across the 18 sampled qualifications, only a couple came close to approximating this transparency principle; in particular, the **CIBTAC L3 Beauty Dipl.** qualification. Its overall qualification grade enables users to infer (from the overall grade alone) a certain minimum level of competence across all units/components; even extending

²³ This is true unless there are optional units; in which case the user would also need to know which combination of units was taken. Having said that, the theory of CBA would recommend against optional units, for the sake of transparency.

to the possession of certain critical skills. This is true for Pass, Merit, and Distinction grades, since this qualification requires that all units/components must be achieved at the grade in question (or higher) for the award of that grade overall. There is still an element of compensation in this qualification, through the aggregation of marks within each exam, but the mastery principle clearly dominates. This makes the qualification grade as transparent as possible. The **RSPH L4 Food Award** is similar, in requiring that both components must be achieved at the grade in question (or higher) for the award of that grade overall. But its use of two compensation-based written exams obscures the meaning of its grades somewhat.

For most of the sampled qualifications, the overall qualification grade lacked transparency, conveying only a very limited amount of information. This was particularly so for qualifications that operated a charity principle, when aggregating information across units/components, beyond a mere compensatory one. One way of responding to this challenge would be for an AO simply to provide information to candidates, and therefore to users, concerning attainment at the unit/component level.²⁴ However, the fact that most of the qualifications operated multiple aggregatory principles simultaneously – sometimes combining mastery with compensation and charity – makes the transparency challenge particularly tricky to overcome.

Grades from written exam units, within the sampled qualifications, also tended to lack transparency. These grades tended not to be determined by having met specific AC, which might otherwise have enabled them to convey information concerning the acquisition of specific competencies. In addition, the requirement to report multiple grades meant that even the passing grade – which typically corresponded to around 60% of available marks – could not really be said to provide information concerning ‘mastery’ of the LOs associated with a unit (ie 60% would seem to be substantially lower than mastery).

The foregoing discussion of transparency has focused on the potential for qualification users to access information concerning candidate proficiency at the unit/component level; either by being provided with that information directly, or by being able to infer it from the overall qualification grade. Yet, having accessed that information, the challenge of how to understand it – in terms of the LOs/AC that each unit/component assesses – still remains. This would seem to be a particular challenge for criterion-related grading approaches, where measurement standards are specified as disaggregated sets of criteria, rather than as holistic, best-fit descriptions. How can qualification users be supported in making sense of candidates’ unit/component grades, under such circumstances? In other words, what does a successfully met list of AC for a unit *actually mean* about a candidate’s overall level of proficiency?

Conceptual issues in grading VTQs

This second section considers deeper, more conceptual, issues concerning grading practices for regulated VTQs in England. It notes the widespread influence of CBA,

²⁴ An alternative approach, which has been adopted by certain AOs, at various points in time, would be to recognise achievement purely at the unit level, without aggregating to the qualification level; meaning that candidates (and users) would be provided with a profile of results, but not a single overarching result.

but questions its adequacy and appropriateness as a default template for designing VTQs – particularly graded VTQs – basing this analysis on evidence from the sampled qualifications. It begins by acknowledging how regulations and guidance related to the Qualifications and Credit Framework (QCF) helped to ensure the widespread influence of CBA.

One size does not fit all

The most important principle underlying qualification design is that it should be tailored to purpose(s), population(s), and context(s). Many different design considerations will ultimately shape the features and processes built into a qualification. But the guiding principle must always be fitness for purpose(s), population(s), and context(s).

The fact that the QCF was fundamentally premised upon a one-size-fits-all model of qualification design was always going to be problematic. Rachael Meech’s account of how AOs sought recognition for Graded Exams within the National Qualifications Framework (NQF) during the 1990s, and then within the QCF during the 2000s, illustrates this perfectly (Meech, 2018). She explained how many AOs “found elements of the new QCF regulatory criteria challenging to meet whilst simultaneously preserving the ethos and purpose of their graded examinations and associated processes [and] faced challenges with the conceptual framework of the QCF” (p.15). These challenges included, for example: requirements for centre approval when a centre-based model of assessment was not operated for Graded Exams; and requirements for reconfiguring nomenclature, processes, and rules for qualifications that were already well-established, well-understood, and well-articulated via detailed syllabuses. Meech explained how AOs “had to work on devising a format which both preserved the ethos and value of the graded examinations model but which would also be acceptable to QCA” (p.18), noting the particular problem “of producing assessment criteria in a criterion-referenced template for a compensatory method of assessment through a ‘one off’ external examination” (p.19).

Just like Graded Exams, many well-established qualifications were overhauled in order to bring them into line with QCF regulations and expectations, which included design requirements bearing the hallmarks of CBA, such as:

- atomistic specification of measurement standards in terms of LOs and AC;
- a mastery measurement model; and
- assessment based on the exhaustive sampling of AC.

Since the withdrawal of the QCF, in 2015, these requirements no longer apply to regulated VTQs in England. However, because qualification redesign is very expensive, time-consuming, and challenging for all stakeholders, it seems likely that many regulated VTQs in England will continue to bear these hallmarks for some time to come. Yet, even amongst the 18 qualifications sampled for the present study – some of which were former QCF qualifications and others not – it was interesting to note how these characteristics were not adopted in a single, uniform fashion. Instead, these core characteristics of CBA were accommodated, and adapted, in a

variety of more or less subtle ways, so as to preserve the ethos, purpose and value of each qualification.

Not necessarily mastery

None of the sampled qualifications operated a pure mastery measurement model. All of them incorporated compensation – a very different aggregatory principle – in one way or another. Some of them operated a variety of aggregatory principles simultaneously; also including configuration, and charity. Mastery and charity are, at their most extreme, polar opposites. This makes the frequency of their cooccurrence within the sampled qualifications all the more interesting.

Certain of the qualifications took on some of the trappings of mastery, whilst not necessarily delivering all of the goods. Qualifications that operated a performance quality criterion-related grading approach – including strong mastery for higher grades, ie all Merit/Distinction criteria for unit Merit/Distinction – provide an interesting case in point. For qualifications like these, the meaning of each unit grade is ‘sort of’ transparent, in the sense that a Merit means that the candidate has achieved all of the Merit criteria but not all of the Distinction criteria. Yet, in practice, it may not be especially transparent, if the meaning of each Merit criterion, by dint of the inevitable simplicity of the set of quality statements associated with it, amounts to little more than: higher-quality than a Pass, but lower-quality than a Distinction.

The one interpretation of mastery that was almost universally respected was the idea that candidates need to pass all units in order to pass the qualification overall. Even here, though, one qualification diverged; allowing candidates to fail a unit, yet still pass the qualification.

Requiring a Pass on all units (to pass the qualification overall) provides at least some transparency for the qualification Pass grade. Typically, though, there was no such transparency for the higher grades of the sampled qualifications. In effect, the only information conveyed by these higher grades was that a candidate with a Merit (or Distinction) had demonstrated a higher level of proficiency, overall, than a candidate with a Pass (or Merit).

Mastery, as an aggregatory principle, is crucial when it is important to be able to infer, from the unit/qualification grade, that a candidate has reached a certain minimum level of proficiency in all specified elements of a domain. The less important it is to be able to make this kind of inference, the less crucial it becomes to implement a mastery model.

Not necessarily atomistic nor exhaustive

Many of the sampled qualifications did adopt an atomistic approach to specifying LOs/AC, as well as an exhaustive approach to assessing them, leaving no criterion unassessed. There were certainly exceptions, though; including qualifications that were 100% externally examined via a compensatory principle, for which detailed AC were not even specified (eg **NOCN L1 ESOL Cert.B2, UWLQ L2 Oral Grade 4**), and qualifications for which detailed AC were specified but not used for grading (eg **CIBTAC L3 Beauty Dipl., TCL L3 Speech Grade 8**). For units/components assessed purely by written exam, with marks aggregated by simple addition, the need for detailed AC was lessened. Certain qualifications specified detailed AC only for

internally-assessed units, ie not for externally-assessed ones (eg **AQA L3 App.Bus. Cert.**).

Some of the sampled qualifications adopted different approaches for assessing the passing grade versus higher grades; embracing atomistic specification and exhaustive sampling for the passing grade, but not for higher grades. The **BTEC L5 Electrical HND** qualification diverged for the higher grades in two respects. First, it adopted generic grading criteria, which were neither atomistic nor unit-specific. Second, it awarded grades based on a restricted set of units; so, in terms of the higher grades, it was ultimately not exhaustive. Similarly, the **VTCT L3 Barbering E.Dipl.** qualification was less than exhaustive for the higher grades, excluding assessment information from 4 optional units and 1 mandatory unit (the **IMI L3 Vehicle Dipl.** qualification also excluded information from optional units). The **VTCT L3 Barbering E.Dipl.** qualification was unusual in that it adopted an atomistic approach to specifying AC, including for higher grades on certain units. Yet, these higher grade criteria, for each unit, were quite different to criteria specified for the passing grade; aiming to identify features of higher-level professional practice, in a similar way to the best-fit descriptions that were used as measurement standards for the synoptic assessment.

Atomisation goes hand in hand with mastery, when it is important to be able to infer, from the unit/qualification grade, that a candidate has reached a certain minimum level of proficiency across all specified elements of a domain. The greater the number of critical micro-competencies identified, the greater the atomisation of AC will need to be, and the more exhaustive the sampling process will also need to be. Conversely, the fewer the number of critical micro-competencies identified, the less important atomisation and exhaustive sampling become.

Differentiation via command verbs

The present study raises important questions concerning the role of command verbs in grading and levelling VTQs. In particular, as discussed in the section on standardisation, it will be important to consider the degree to which they are adequate or appropriate for bearing the weight of literal interpretation; or whether they are better understood as illustrative markers of measurement standards. The key issue, here, is how assessors understand the interaction between command verbs and the content of the LOs to which they relate; and how assessors can be supported to reach common understandings.

Although the idea of a hierarchy of knowledge and skills, derived from Bloom's Taxonomy, is still widely used to scaffold assessment practices, it has been widely critiqued over the years. Sockett (1971), for instance, referred to the "breath-taking naivety" (p.22) of Bloom's epistemology. Pring (1971) developed this critique, suggesting that "through lack of any epistemological analysis of cognitive processes, the taxonomy makes false distinctions and thereby provides a nonsensical classification" (p.89). Winter (1993; see also 1994) argued that Bloom's Taxonomy could legitimately be used to define measurement standards, for the purpose of grading work; but could not legitimately be used to define qualification levels:

┆ Bloom's taxonomy provides a strategy for *grading* work at *any* educational level: qualities of analysis, synthesis and evaluation will characterise a 'good'

piece of work in primary school, secondary school, years 1, 2 and 3 of a first degree, and a PhD.

(Winter, 1993, p.94)

But what these theories do *not* provide is a conceptual structure of successive (i.e. chronologically separable) hierarchically ordered educational levels, in which a satisfactory process at one level would be clearly differentiated in terms of intellectual functions from a satisfactory process at another level.

(Winter, 1993, p.95)

Bereiter and Scardamalia's (2005) critique of Bloom's Taxonomy echoed Winter's, but went substantially further, raising questions concerning its legitimacy for either levelling or grading.

Assessing and grading VTQs

Although this project began as an investigation into alternative approaches to grading VTQs, it has raised broader questions concerning assessment in VTQs more generally. The model that seems to have become something of a default template for designing VTQs in England, CBA, is clearly no panacea. Indeed, it may be an inappropriate template in certain circumstances; particularly for qualifications whose domains are not characterised by large sets of critical micro-competencies, or for qualifications that would be better served by a measurement model more akin to compensation or configuration. In other circumstances, CBA may not be inappropriate, *per se*; but somewhat inadequate in its basic format. Many of the sampled qualifications appropriated key ideas from CBA, but adapted them in a variety of ways, in order to preserve their ethos, purpose and value.

Even when CBA seems plausible, as a default design template for a particular qualification, it is important to consider whether it is required for the entire qualification, or perhaps only for certain units. In other words, ought the entire domain to be modelled in terms of multiple sets of critical micro-competencies? Or, assuming that sets of critical micro-competencies do need to be identified, might they be restricted to certain units, or to certain LOs within all units? Or, as operated by some of the sampled qualifications, might these sets be distributed across all LOs of all units, but only relevant to judging the passing grade?

These assessment-related questions lead to wider curriculum-related questions, concerning the degree to which certain VTQs are better conceptualised and designed on the model of Curriculum-Embedded Assessments (CEAs), ie designed to function more like traditional GQs than CBAs. In a regulatory context that no longer presumes that one size should fit all, approaches to assessment and grading need to be determined relative to anticipated purpose(s), population(s), and context(s). For certain VTQs, this might recommend an approach closer to CBA, with or without grading; for others, this might recommend an approach closer to CEA; for yet others, this might recommend a hybrid approach, or perhaps the adoption of an entirely different model. Simply defaulting, unthinkingly, to a design template that might be inadequate or even inappropriate is not a sensible approach to qualification design. Nor is it sensible to adopt an unprincipled pick-and-mix approach to qualification design; borrowing an aggregation model from one type of qualification (where it seems to work well), a grading process from another (where it seems to work well), and so on. The pick-and-mix approach is doomed to failure, owing to a

lack of consideration of how the parts of the system interact (or fail to interact) with each other. What is required is a principled, integrated approach, where each step in the design and development of a qualification is referenced back to a detailed specification of anticipated purpose(s), population(s), and context(s).

Principled design

For some of the sampled qualifications, it was clear how core features and processes within their grading practices had emerged from a principled approach to qualification design. In the case of the **ABC L4 Art F.Dipl.** qualification, for instance, certain features of its grading approach derived directly from an underlying curriculum model. The units were structured so as to reflect a developmental progression, from an Exploratory Stage of the qualification, to a Pathway Stage, and culminating in a Confirmatory Stage. As the Personal Confirmatory Study brought together all of the knowledge, skill and understanding developed throughout the course, there was a clear rationale for deriving grades from this unit alone.

A clear rationale was also evident for operating a strong mastery principle within the **CIBTAC L3 Beauty Dipl.** qualification. In order for qualification users to be confident in using results from this qualification as a basis for making hiring decisions, those users need to be provided with specific information concerning candidate competencies; including competence across the full breadth of the domain of learning, as well as competence in relation to very specific skills. Indeed, the approach adopted within the qualification was determined in the context of concerns from stakeholders that results from previous qualifications had failed to provide this kind of reassurance.

Finally, the design rationale underpinning the **VTCT L3 Barbering E.Dipl.** qualification was also very clear, owing to its publication in a document on their website (VTCT, undated), which set out a common set of design principles for VTCT's suite of Technical Qualifications.

Further research

The aim of this small-scale research project was to explore the variety of grading practices in operation, across the landscape of regulated VTQs in England, and to consider issues arising. A wide variety of practices was identified, which differed most significantly in relation to their approaches to specifying measurement standards, and to aggregating measurement information. Of particular interest was how many of the qualifications had adapted a default design template associated with CBA – that was premised upon atomistic specification, domain mastery, and exhaustive sampling – so as to incorporate grading. This typically involved the provision of grading grids; in which criteria were specified, at a micro-competency level, both for the passing grade and for higher grades. All sorts of approaches were taken to aggregating measurement information related to these criteria – both within units/components, and then across them – drawing upon principles as diverse as mastery, compensation, configuration, and charity.

This diversity of grading practices highlighted both technical and conceptual issues. Their detailed consideration lies beyond the scope of the present paper, although they clearly indicate avenues for further research. Technical issues related to the effective operation of the kinds of practices observed across the 18 sampled

qualifications. Questions were raised concerning standardisation, grading and levelling, comparability, weighting, burden and backwash, and transparency. Conceptual issues related to underpinning assessment models. In particular, a fundamental question was raised concerning the appropriateness and/or adequacy of CBA as a default template upon which to design and develop graded VTQs. The present research simply raises such questions. It remains for further research, analysis, and dialogue within the VTQ community to explore answers.

The present research left plenty of questions unasked, let alone unanswered.²⁵ In particular, it is important to remember that it was based upon an opportunistic sample of just 18 qualifications. Even within the landscape of regulated VTQs in England, there may well be quite different grading practices in operation, which will raise quite different questions. Likewise, the research is based upon the experience of regulated VTQs in England, which has been heavily influenced by a variety of social, political, and regulatory forces over the past few decades; including high profile reviews, government initiatives, and the establishment of qualifications frameworks (both the NQF and the QCF). Experiences from countries such as Australia and New Zealand, which have also leaned heavily upon the CBA model for vocational qualifications, may prove to be instructive; as may insights from countries where grading practices have not been so heavily influenced by CBA.

Although this research was supported by very useful conversations with AO colleagues, the most important caveat to bear in mind is that it was essentially a desk-based exercise, based upon the representation of measurement standards and grading practices within qualification documents. What remains far from clear is how these representations are appropriated and used by assessors, and moderators, when judging candidates' performances. There might well be significant differences between how grading is presumed to be practised – as recorded in qualification documents – and how it is actually practised. Such differences might either be to the detriment of assessment accuracy, or, conceivably, to its benefit. The possibility of significant differences between documented and actual practices seems all the more likely for qualifications that took on some of the trappings of CBA, largely to comply with social-political-regulatory requirements or expectations, without radically changing well-established assessment procedures. Even assuming no differences between documented and actual practices, it would still remain unclear exactly how documentary representations are appropriated and used by assessors and moderators; in particular, the nature and level of (professional, pedagogic, or assessment) expertise that is required in order to ensure the effectiveness of any particular grading approach (see Johnson, 2008, for insights at this level of grain size). In short, the present report represents just a first step in a programme of work that will be necessary to understand in more depth what good practice in grading VTQs looks like.

²⁵ For instance: structural questions, related to the number of grades awarded, or to how they are labelled; technical questions, related to variance attenuation (regression to the mean) when assessment information is aggregated, or to determining the reliability of grading judgements; pragmatic questions, related to the costs of grading, or to accommodating assessors with limited (professional, pedagogic, or assessment) expertise; consequential questions, related to how grades are used, or to their impacts on teaching and learning; and so on.

References

- Ahmed, A. and Pollitt, A. (2011). Improving marking quality through a taxonomy of mark schemes. *Assessment in Education: Principles, Policy & Practice*, 18 (3), 259-278.
- Bereiter, C. and Scardamalia, M. (2005). Beyond Bloom's Taxonomy: Rethinking knowledge for the knowledge age. In M. Fullan (Ed.). *Fundamental Change: International Handbook of Educational Change* (pp.5-22). Netherlands: Springer.
- 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.
- Department for Business, Innovation and Skills (2014). *Getting the job done: The government's reform plan for vocational qualifications*. Available online: <https://www.gov.uk/government/publications/vocational-qualification-reform-plan>
- Department for Education (2017a). *Implementation of T level programmes. Government consultation*. Available online: <https://consult.education.gov.uk/technical-education/implementation-of-t-level-programmes/>
- Department for Education (2017b). *Technical and applied qualifications for 14 to 19 year olds. Key stage 4 and 16 to 18 performance tables from 2020: technical guidance for awarding organisations*. Available online: <https://www.gov.uk/government/publications/14-to-19-technical-and-applied-qualifications-technical-guidance>
- Johnson, M. (2008). Assessing at the borderline: Judging a vocationally related portfolio holistically. *Issues in Educational Research*, 18 (1), 26-43.
- Lum, G. (2013). Competence: A tale of two constructs. *Educational Philosophy and Theory: Incorporating ACCESS*, 45 (12), 1193-1204.
- Meech, R., with McBirnie, A. and Jones, G. (2018). *Graded exams in Music, Dance & Drama and the educational regulatory landscape in the UK and internationally*. London: Council for Dance, Drama and Musical Theatre. Available online: <http://www.gradedexams.com/images/Sector-report-on-graded-examinations-2018.pdf>
- Newton, P.E. (2011). A level pass rates and the enduring myth of norm-referencing. *Research Matters*, Special Issue, 2, 20-26.
- Newton, P.E. (2017). *An Approach to Understanding Validation Arguments*. Coventry: Office of the Qualifications and Examinations Regulator. Available online: <https://www.gov.uk/government/publications/an-approach-to-understanding-validation-arguments>

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

Ofqual (2008). *Regulatory arrangements for the Qualifications and Credit Framework*. Coventry: Office of the Qualifications and Examinations Regulator.

Ofqual (2015a). *Qualification and Component Levels Requirements and Guidance for All Awarding Organisations and All Qualifications*. Ofqual/15/5774. Coventry: Office of Qualifications and Examinations Regulation.

Ofqual (2015b). *After the QCF: A New Qualifications Framework Decisions on Conditions and Guidance for the Regulated Qualifications Framework (RQF)*. Coventry: Office of Qualifications and Examinations Regulation.

Ofqual (2016). *General Conditions of Recognition*. June 2016. Ofqual/16/6023. Coventry: Office of Qualifications and Examinations Regulation.

Ofqual (2018a). *Functional Skills English Conditions and Requirements (Ofqual/18/6385/4)*. Coventry: Office of Qualifications and Examinations Regulation.

Ofqual (2018b). *Functional Skills Mathematics Conditions and Requirements (Ofqual/18/6385/6)*. Coventry: Office of Qualifications and Examinations Regulation.

Ofqual (2018c). *Implementation of Technical Qualifications: Consultation on rules and guidance for Technical Qualifications (Ofqual/18/6422/3)*. Coventry: Office of Qualifications and Examinations Regulation.

Pring, R. (1971). Bloom's Taxonomy: A philosophical critique (2). *Cambridge Journal of Education*, 1 (2), 83-91.

QCDA (2008). *Guidance for using unit level descriptors within the Qualifications and Credit Framework*. Version 3. Coventry: Qualifications and Curriculum Development Agency.

QCDA (2010a). *Guidance for developing rules of combination for the Qualifications and Credit Framework*. Version 4. Coventry: Qualifications and Curriculum Development Agency.

QCDA (2010b). *Guidelines for writing credit-based units of assessment for the Qualifications and Credit Framework*. Version 4. Coventry: Qualifications and Curriculum Development Agency.

QCDA (2010c). *Grading Qualifications in the QCF: Guidance for awarding organisations*. QCDA/10/5201. Coventry: Qualifications and Curriculum Development Agency.

- Richard, D. (2012). *Richard Review of Apprenticeships*. Available online: <https://www.gov.uk/government/publications/the-richard-review-of-apprenticeships>
- Sadler, D.R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13 (2), 191-209.
- Sainsbury, D. (2016). *Report of the Independent Panel on Technical Education*. Available online: <https://www.gov.uk/government/publications/post-16-skills-plan-and-independent-report-on-technical-education>
- Sockett, H. (1971). Bloom's Taxonomy: A philosophical critique (1). *Cambridge Journal of Education*, 1 (1), 16-25.
- VTCT (undated). *VTCT Technical Qualifications: Design Rationale (Version 10)*. Hampshire: Vocational Training Charitable Trust. Available online: <https://qualifications.vtct.org.uk/finder/docs/common/Technical%20Qualifications%20-%20Design%20rationale.pdf>
- Whitehead, N. (2013). *Review of Adult Vocational Qualifications in England*. London: UK Commission for Employment and Skills. Available online: <https://www.gov.uk/government/publications/review-of-adult-vocational-qualifications-in-england--2>
- Winter, R. (1993). The problem of education levels (part 1): Conceptualising a framework for Credit Accumulation and Transfer. *Journal of Further and Higher Education*, 17 (3), 90-104.
- Winter, R. (1994). The problem of educational levels, part II: A new framework for credit accumulation in Higher Education. *Journal of Further and Higher Education*, 18 (1), 92-106.
- Wolf, A. (1995). *Competence-Based Assessment*. Maidenhead, UK: Open University Press.
- Wolf, A. (2011). *Review of Vocational Education: The Wolf Report*. Available online: <https://www.gov.uk/government/publications/review-of-vocational-education-the-wolf-report>



© Crown Copyright 2018

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated.

To view this license, visit

www.nationalarchives.gov.uk/doc/open-government-licence/

or write to

Information Policy Team, The National Archives, Kew, London TW9 4DU

Published by:

ofqual

Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

0300 303 3344
public.enquiries@ofqual.gov.uk
www.gov.uk/ofqual