Internal assessment in existing national technical and vocational qualifications

ofqual

Authorship

This report was written by Charlotte Lockyer and Stuart Cadwallader of the Strategy, Risk and Research Directorate.

Acknowledgements

The authors would like to thank the teachers and assessors who participated in this research and provided us with the benefit of their insight and expertise. The authors are also grateful for feedback received from members of Ofqual's Research Advisory group.

Contents

1 Executive summary

Internal assessment is widely used in Vocational and Technical Qualifications (VTQs) to assess practical skills and competencies which are traditionally more difficult to assess through an external exam. Although internal assessment is widespread, there is little published research on how it works on the ground in VTQs. In our 2018-21 Corporate Plan, Ofqual committed to "evaluate the functioning of internal assessment in existing national VTQs". As a first stage in this work we undertook an exploratory qualitative study. This aimed to give us a contemporary overview of practice in the internal assessment of VTQs. We selected 6 regulated qualifications on which to base our study. These included Level 2 and 3 qualifications in Construction, Hair and Beauty, and Information Technology. They fell broadly into 2 groups: 'strongly vocational' and 'softly vocational' qualifications. Strongly vocational qualifications lead directly into specific employment, whilst softly vocational qualifications are more focussed on applied knowledge and skills and usually lead to further study.

As part of our study, we carried out interviews with 45 teacher-assessors in 21 centres. The scope of our interviews was wide, covering summative task setting; task taking; learner support; assessment judgements; and quality assurance. We deliberately adopted this holistic approach in order to develop a contextualised understanding of how the assessment system works as a whole. We should note that all views expressed in this report reflect the perceptions of the assessors interviewed. We do not suggest that practice described by assessors definitively reflects AO processes and qualification requirements or the views of other agents in the sector. Instead, we use these assessor accounts to build our understanding of the lived experience of internal assessment in VTQs.

One of the most striking features of this study is just how diverse, complex and dynamic the VTQ sector is. The 6 VTQs in our sample were delivered to a different learner base in different settings and had very different purposes. As such, they had differing assessment requirements, grading structures and quality assurance mechanisms. As well as the variation between VTQs, there was variation in how centres delivered the same VTQs. This is a highly complex space for regulation.

Interestingly, assessors rarely talked about this complexity. Instead many expressed frustrations about the frequency of reform in the sector, particularly in softly vocational qualifications. Assessors had experienced frequent turnover of courses in their centres and regular change in the content and structure of VTQs. This flux can make it harder for assessors to internalise assessment criteria and requirements. This can lead to a blurring of the lines between different courses as assessors draw on their experience of delivering other VTQs to inform their delivery of a new course. In contrast, the strongly vocational courses in our sample had experienced greater stability. Whilst this promoted assessor familiarity with assessment criteria, it had a trade-off in terms of the currency of qualification content. Some assessors believed these courses had not kept pace with practice in fast-moving industries.

Although awarding organisations (AOs) remain responsible for the standards of VTQs, this research confirms that many decisions underpinning these standards are devolved to centres. This reinforces the findings of our work over the past couple of years on Accountability for Awards, where assessment judgements are made by centres on behalf of AOs. Depending on the VTQ, assessors are trusted to make

important decisions about task setting, task taking, applying the assessment criteria and ensuring quality. The level of oversight AOs have (or that assessors believe that AOs have) over each aspect of this process currently varies significantly. In any case, AOs place considerable trust in the knowledge and integrity of assessors. This might be viewed by many as a strength of the system; the assessors we interviewed were generally highly experienced and confident in their ability to deliver assessments.

Linked to the devolution of control from AOs to centres is the suggestion that qualifications do not always work in the manner we may assume from reading specifications. There was a sense that the delivery of these VTQs rely on a level of tacit knowledge and understanding. This was particularly true of the application of assessment criteria where assessors may consciously or unconsciously apply an assessment standard which is subtly different from that set out by the AO. Although assessors found assessment criteria simple to use, they often drew on other sources of information when making assessment decisions. These may include their knowledge of industry standards, their experience of delivering other VTQs or a common understanding of the standard developed through communities of practice.

One aim of this study was to identify vulnerabilities in the delivery of internal assessment in VTQs. Although we have tried to do this, it does not lead us to a straightforward set of recommendations. If anything, the study reinforces the need for cautiousness and considerable thought when looking to make changes or improvements. The assessment system is a complex network of interconnecting parts; any attempts to strengthen one part could well undermine other elements of the system. Moreover, we found that often the points of vulnerability in these VTQs double as their strengths. For example, whilst devolution of decision-making to centres can introduce risks to reliability, it also affords assessors the flexibility to deliver VTQs which engage learners and best prepare them for the needs of local employers. Assessors saw this flexibility as one of the most valuable aspects of these VTQs. There is a perceived risk that additional regulatory controls could adversely affect the value of certain qualifications.

Nonetheless, there are aspects of the system we may wish to explore further. One of these is the variation in support given to learners by assessors as they undertake summative tasks. Assessors delivering the same course often had different views on what was permissible in terms of learner support, a source of variability which may have implications for standards. Assessors also identified issues with the clarity or sufficiency of AO guidance particularly in the early days of qualification delivery.

Another area where we expect to maintain a close focus is the quality assurance of centre-based assessment decisions. Assessors often had very different views on the intended aim and focus of quality assurance activity. In both internal and external quality assurance (IQA and EQA) there was a conflation of capacity building and standardisation, on one hand, and quality control on the other[1]. For many assessors, the EQA process was as much about supporting centres in their delivery of internal assessment as scrutinising them. Although assessors expressed confidence in the rigour of the EQA process, they had little experience of EQAs overturning or adjusting assessment judgements. Where AOs devolve control of internal assessment to centres, it is vital they can assure that standards in these centres are

---

[1] We should note that the external quality assurance (EQA) by AOs of centre-based assessment delivery referred to here has no connection with the EQA of apprenticeship end-point assessments.

being upheld. For this reason, we are setting out new requirements for AOs to put in place a process to scrutinise the standards of centre assessment decisions, known as a Centre Assessment Standards Scrutiny (CASS), and for all centre assessments to be subject to a form of such scrutiny. We will also be carrying out further research in this area.

Although we identified certain points of vulnerability in the delivery of internal assessment in these VTQs, we found the qualifications were highly valued by assessors. They were seen as valid, engaging and pitched at an appropriate level of demand, whilst providing learners with the skills employers need. In the strongly vocational qualifications in particular, the importance of preparing learners to meet the needs of local employers was paramount. This often went beyond upholding assessment standards and extended to assessors delivering extra technical and employability skills to ensure learners were good ambassadors for their centre. In this regard, centres are incentivised to maintain the qualification standard because they are judged not only on the grades received by their learners, but how those learners go on to perform in local industry.

Overall, this study has enhanced our understanding of how the delivery of internal assessment is influenced by both the nature of the specific qualification and the exact context in which that qualification operates. This nuanced information is important for directing further research into internal assessment in VTQs and informing future decisions around how best to regulate such qualifications. We will now use these findings as a basis for further study into internal assessment in the sector.

2 Introduction

2.1 Background

In many Vocational and Technical Qualifications (VTQs), awarding organisations rely on teachers and assessors in centres to make decisions about their candidates' performance against assessment criteria. This is known as internal assessment.

Internal assessment allows for the assessment of competencies which are not easily assessed by external exam. It might take a wide variety of forms. Perhaps traditionally in VTQs, internal assessment is associated with outcomes-based observations of practical skills. It may also comprise written assessments of practical knowledge. In England, assessment of VTQs usually takes the form of criterion-referenced assessment or competence-based assessment. These models of assessment are based around the specification of clear and detailed learning outcomes and assessment criteria. They provide learners and assessors with greater transparency as to what is expected of them, and give employers and end-users more information about what a person holding the qualification knows and can do.

In their comprehensive review of the Learning and Skills Sector (LSS), Torrance et al (2005) note this clear specification of learning outcomes and assessment criteria has "significantly benefited learners in the LSS in terms of the numbers of learners retained in the system and the awards which they achieve" (p1). Jessup (1991) identified this widening of access of learning as a key feature of competence-based assessment in the context of NVQs.

Proponents of criterion-referenced and competence-based assessment identify many other benefits. One of the foremost of these is that it promotes the validity of assessment. There are several ways by which validity can be enhanced. It enables assessors to devise tasks that reflect authentic workplace contexts. It also allows them to take into account the needs and interests of their learners. Learner-centred assessments may enhance performance and motivation (Ecclestone, 2000) and hence give a better representation of a learner's understanding. In review of assessment in upper secondary education for the OECD, Dufaux (2012) argued that lower-performing students may be particularly affected by the type of task, feeling more discouraged and anxious under the pressure of exams. Others suggest that clear and unambiguous assessment criteria take much of the subjectivity out of assessment decisions (Jessup, 1991). Internal assessment also allows centres to adapt tasks to their facilities and may minimise disruption to teaching (Vitello and Williamson, 2017).

Although criterion-referenced assessment has been a feature of internal assessment in England for at least the last 40 years, there is little published research into its implementation in VTQs. Certainly, there is no comprehensive exploration of current practice in the literature. What research there is suggests that, for its ostensible benefits, the advantages of criterion-referenced assessment have historically been difficult to realise. One of the key challenges is in defining clear, unambiguous assessment criteria. Wolf (amongst others) argues that the pursuit of perfect clarity in assessment is an unattainable goal. Even the most detailed criteria cannot convey every possible meaning and it follows that one can never remove the role of assessor judgement from the assessment process (Wolf, 1993 and 1995).

This view is supported by various small-scale studies in the field. Over the years researchers have found variation in assessor interpretation of apparently quite specific assessment criteria (Wolf and Silver, 1986; Black et al, 1989; Carter and Bathmaker, 2017) and task taking requirements (Wolf and Silver, 1986). In research for the Manpower Services Commission, Wolf and Silver (1986) set experienced assessors and workplace supervisors the task of devising assessments based on detailed 'standard tasks'. The assessments they produced turned out to be very diverse. Moreover, assessors all had different ideas of what 'competence' in these assessments would look like. More recently, Torrance et al (2005) found variation in assessor interpretation of assessment criteria and qualification requirements across a range of VTQs and settings. This is not to say internal assessment is inherently unreliable, however. In a study on the reliability of workplace-based assessments in 3 qualifications in hairdressing and electrotechnical engineering, Harth and Hemker (2013) found extremely high rates of inter-rater reliability.

Research suggests that the consistency of assessor judgements is influenced by a multiplicity of factors. Some authors discuss the idea that all assessors have a unique 'position' in the education and assessment system which is determined by their individual values; past experiences; views of the assessment; and views of the competencies required in a particular vocational field. Their interpretation of criteria is shaped by this unique combination of factors (Wolf, 1995; Shay, 2005; Johnson, 2008). Researchers found evidence of this at work across a range of vocational settings (Wahlberg and Gleeson, 2003; Torrance et al, 2005; Johnson, 2008).

Despite the atomised assessment criteria and the mastery model often associated with VTQs, there is evidence that assessors still make holistic, contextualised and compensatory assessment decisions (Torrance et al, 2005; Johnson, 2008; Carter and Bathmaker, 2017). For example, in 2 separate studies with engineering and health and social care assessors, researchers found that assessors have a notion of the 'worth' of a student, referring to a typical 'pass grade' or 'merit grade' student (Garland, 1998; Carter and Bathmaker, 2017). Assessors interpret criteria through this lens, in some cases disregarding assessment criteria when it may disadvantage a student they believe is 'worth' a certain grade (Garland, 1998).

One way in which assessors try to standardise their interpretation of assessment criteria is through the use of communities of practice (Black et al, 1989; Wolf, 1995; Torrance 2005; Johnson, 2008). Here formal or informal networks of assessors agree a common understanding of any subtleties in the assessment criteria. In a study of Scottish National Certificate assessors, Black et al (1989) found such networks highly effective in standardising decisions where assessment criteria were more ambiguous. Torrance et all (2005) suggests these local networks "constitute the context in which all meaningful judgements about standards are made" (p.3).

The published research has also pointed to evidence of instrumentalism in the assessment of VTQs. This is discussed in most depth in Torrance et al (2005). They found that clarity in assessment criteria and processes can lead to an undesirable focus on criteria compliance which they describe as "assessment *as* learning". As part of this they identified widespread support for students across the sector and a conflation of formative and summative assessment. A similarly assessment-driven approach was identified in other smaller scale studies with BTEC (Carter and Bathmaker, 2017) and GNVQ assessors (Garland, 1998; Ecclestone, 2002; Wahlberg and Gleeson, 2003). Researchers found assessors were focused on

"getting students through" (Carter and Bathmaker, 2017, p470) through extensive formative feedback and the use of numerous discrete and accessible tasks. Some authors suggest this can create a scenario where candidates are coached to comply rather than learn (Garland, 1998; Ecclestone, 2002).

The approaches described above have been linked to improvements in learner achievement and progression. However, they have also raised questions about the quality of the learning outcomes achieved (Ecclestone, 2002; Torrance et al, 2005). In some studies assessors questioned whether such assessment practices developed the proficiency required for students' progression into Higher Education or employment (Sutherland and Pozzi, 1995; Carter and Bathmaker, 2017). Other studies show evidence of unwarranted grade inflation in certain VTQs (Kelly, 2017; Cuff et al, 2018). These types of findings have raised concerns amongst some stakeholders about assessment standards in certain VTQs and set the context for major reviews of the sector. Both the Wolf Review (2011) and Sainsbury Review (2016) have emphasised the importance of external assessment in safeguarding assessment standards.

In recent years there has been little further research published in this area and we have little indication of whether the findings described above might still describe the current assessment system. It is in this context that we set out to improve our understanding of the operation of internal assessment in VTQs.

2.2 Our research approach

In our 2018-21 Corporate Plan, we committed to *"evaluate the functioning of internal assessment in existing national VTQs"*. The intention was to use our findings to help inform the design of internal assessment in reformed VTQs. As a first stage in this work we designed a study to *explore* the functioning of internal assessment in VTQs. This exploratory study aimed to improve our understanding of what constitutes current practice in the internal assessment of a selection of 6 high-profile regulated VTQs and to explore some of the strengths and vulnerabilities of this practice. Our intention was to carry out a holistic piece of research to build a contextualised understanding of how the system works as a whole in order to inform intelligent regulation of the sector.

The research drew upon the perspectives of teachers and assessors to explore the following research questions:

1. What do teachers and assessors perceive to be the strengths and vulnerabilities of internal assessment in the context of the qualification?

2. How do teachers and assessors go about setting tasks for the internal assessment, and how is this informed by wider internal and external quality assurance processes?

3. What is the level and nature of the support which learners receive from their assessors when completing internal assessment?

4. What processes do assessors go through to reach their assessment decisions?

5. How are assessment judgements standardised across assessors, both within the centre and between centres?

2.3 Methodology

Our research questions were open, nuanced and explorative, which lends them to a qualitative methodology that emphasises depth over breadth. For this reason, a semi-structured interview approach was selected for this study. This method, in this context, only allows very tentative generalisations beyond the sample (and not enough to, for example, ascertain the prevalence of the discussed practices and issues across *all* centres). But it does allow us to better understand the complex mechanisms through which the assessment arrangements may work in practice. This will help us to identify whether there are systemic vulnerabilities which might threaten the validity or reliability of internal assessment in VTQs.

Face-to-face interviews were conducted with teachers and assessors from a sample of 6 different qualifications. Each participant was interviewed on a one-to-one basis or as part of a small group (alongside colleagues from the same centre). Group interviews usually involved 2 or 3 assessors, at most including 5 assessors from a centre. The type of interview was selected pragmatically, based on the timing of the visit and the availability of the assessors. Five members of Ofqual's research team[2] were involved in conducting the interviews and in each case they travelled to meet the assessor(s) at their centre.

Our discussions were based on a semi-structured interview schedule (Annex A), which aimed to elicit information about the process and experience of delivering internal assessment in practice. Though the interview schedule provided a structure and focus for the interviews, the assessors were encouraged to raise anything they felt was pertinent to the discussion. All interviews were audio recorded and then transcribed verbatim by an external transcription organisation.

Prior to taking part in the research, participants gave their informed consent to the interview (Annex B). They were assured of the complete confidentiality of their responses and told that any information gathered during the interviews would be used for the purposes of our research only. Participants were able to withdraw from the research at any time.

2.3.1 Sample

Three sector subject areas were selected for the study: Construction, Hair and Beauty, and Information Technology (IT). These areas were chosen because they cover significantly different skills and knowledge, potentially leading to very different career destinations. Two qualifications were sampled from each of these 3 sector subject areas (Table 1), each of which was offered by a different Awarding Organisation (AO). The number of specifications is deliberately small to allow us to develop sufficiently detailed and contextualised understanding. The qualifications were selected on the basis that:

1. They included an internally assessed and an externally assessed component,

2. They were either Level 2 or Level 3[3],

---

[2] Each researcher conducted at least 2 of the interviews. Some of the earlier interviews were attended by 2 researchers, with one leading the interview and the other acting (primarily) as an observer. This was for the purpose of standardising researchers in their use of the interview schedule.

[3] Two of the Level 2 qualifications offered the option for awarding a pass at Level 1 if a learner did not meet the level required for a Level 2 pass.

3. They were funded for delivery in centres or contributed to performance tables.

4. A significant number of learners had certificated with the qualification within the last year.

Table 1. Qualifications in the study

Subject area	Qualification
Information Technology	Level 2 Certificate
	Level 3 Certificate / Diploma
Construction	Level 3 Diploma
	Level 2 Award
Hair and Beauty	Level 2 Diploma
	Level 2 Diploma

The awarding organisations for the selected qualifications provided us with a list of centres who had entered candidates for the qualification in the previous year. A random number generator was used to create a randomised list of these centres, with the aim of sampling a variety of centre types from across a range of geographical regions (though, for budgetary reasons, only centres within 100 miles of Ofqual's office in Coventry were included).

Letters and, where possible, emails were sent out to the Head Teacher or Principal for each of the top 20 centres on each of the qualification-specific lists. The letters explained the purpose and method for the study and invited the centre to take part. A small financial incentive was offered to provide centres with compensation for their time and ensure that enough assessors could be recruited to take part. It is possible that this incentive may have affected the sample composition, but we felt it would help encourage participation from centres operating under financial restrictions.

Twenty-one centres took part in the research. Participating centres varied in the number of assessors which were available for interview. A total of 45 assessors were interviewed, 11 on a one-to-one basis and 34 as part of a group interview. Interviews ranged between 41 and 122 minutes in duration, with an average duration of approximately 64 minutes. Table 2 shows a breakdown of the number of centres and interviews of each type by subject area.

Table 2. Interviews

Subject Area	Centres	1 to 1 Interviews	Group interviews
Information Technology	6	6	2
Construction	7	4	3
Hair and Beauty	8	1	9
Total	21	11	14

Where possible we tried to recruit a range of different centre types although we were ultimately constrained by the centres responding to our invitation. We visited 10

secondary schools and 10 further education (FE) colleges as well as one private training provider (PTP).

Table 3. Types of centre in the study

Subject Area	Secondary schools	FE colleges	Private training providers
Information Technology	6	0	0
Construction	4	3	0
Hair and Beauty	0	7	1
Total	10	10	1

2.4 Analysis

Data from the interviews was analysed thematically (for example, Aronson, 1995; Braun & Clarke, 2006) using the NVivo qualitative analysis software package. The process was completed by 2 researchers, one of whom was the lead. There were 3 stages to the analysis:

1. An initial coding scheme was prepared on the basis of the questions that had underpinned the interview schedule (Annex A). The 2 researchers each used this coding scheme to individually analyse a different sample of the transcripts. While analysing the data, they refined the initial codes and added new ones to represent emerging themes and information.

2. Following this primary coding process, the researchers worked collaboratively to add and refine codes on the basis of their individual analyses. They jointly developed and agreed upon a *final coding scheme*.

3. The lead researcher used the *final coding scheme* to code all of the transcripts (including those from the stage 1 sample) on a line by line basis.

4. The second researcher 'double coded' a sample of the transcripts (approximately 20%) to ensure consistency against the *final coding scheme*.

Throughout this process, the researchers sought to validate and refine their understanding of the data by searching for any evidence which elaborated or undermined the emerging themes. Passages of text were coded (rather than just individual sentences) to ensure that themes and quotes were captured in context. Researchers had access to both the written transcript and the original audio recording of each interview, allowing them to account for the tone of the discussion wherever there was any ambiguity.

2.5 Reporting

The findings in this report explore the perceptions, attitudes and behaviours of assessors (our collective term for interviewees) based on their lived experiences of delivering the 6 VTQs in our sample. We should not assume that any views are a factually accurate representation of the official documented content of and controls associated with each VTQ. Similarly, they may not reflect the views of other agents

or organisations in the sector. Instead, they provide a useful source of evidence on how internal assessment is delivered by assessors on the ground.

Although we use the term 'assessors' throughout the report, almost all assessors also taught the course they assessed. Many also held an Internal Quality Assurer (IQA) or a centre leadership role. Their insights into the system are likely to be influenced to at least some degree by these other roles.

Despite the complexity of the data we have tried to quantify the coding at points in the report in order to give a sense of the weight of responses. Any numbers should be used with caution. The broad and flexible nature of the interviews meant that we could not discuss every aspect of internal assessment consistently in each interview. Numbers are therefore not definitive and are included solely to give an indication of the prevalence of opinion.

In our discussion of findings, we present an extensive range of quotes to represent assessors' experiences in their own words. These are selected to illustrate themes that emerged from the data rather than to represent the sample or population as a whole. To preserve the anonymity of our participants we do not report centre names, instead each quote is attributed to a qualification and a centre type.

As we undertook our analysis we identified 2 distinct groupings of qualifications. We refer to these in the report as 'strongly vocational' and 'softly vocational' qualifications. The 'strongly vocational' qualifications are heavily focused on providing learners with practical skills which lead directly into specific employment, whilst the 'softly vocational' qualifications are more focussed on applied knowledge and skills and tend to lead to further general or vocational study. We categorised half of the qualifications in our sample as softly vocational and half as strongly vocational. We do not refer to individual qualification titles in the report, instead we refer to the Level of the qualification and the general field of study.

3 Findings

3.1 Overview of qualifications

The qualifications in our sample are extremely diverse. They have fundamentally different purposes which are reflected in both the nature of cohorts taking them, and their varying destinations on completion of the qualification.

This section provides an overview of the purposes of, progression from and profile of learners taking the VTQs in our sample. We also summarise assessors' views of the demand of these qualifications. Although these topics were not a specific focus of our interviews, they provide important context for the rest of the report. The target audience and their intended destination was an important part of how assessors thought about the qualification. It influenced their thinking about assessment demand, qualification validity and the role of internal assessment.

3.1.1 Qualification purpose and progression

The qualifications in our sample fall into 2 broad categories. The first is made up of qualifications we term 'strongly vocational'. These include both Level 2 Diplomas in hair and beauty and the Level 3 Diploma in construction. All 3 VTQs aim to prepare learners to directly enter employment. They were often taken by a diverse range of learners including school leavers and mature learners. In our sample, they were delivered in further education colleges (FE) or private training providers (PTPs). Learners completing these VTQs usually progressed into employment in the relevant industry. Some assessors took an active role in securing work for learners, whether through assessor informal links with industry or through work placement officers.

Not all of the VTQs in our sample aimed to prepare learners for specific jobs. The second grouping of VTQs included those we term 'softly vocational'. These include both IT qualifications and the Level 2 construction qualification. These courses tend to be delivered in secondary schools and sixth form colleges. They represent an alternative to academic qualifications for learners aged 14 to 18 and provide them with a basis for further general or vocational study. Learners progressed to a variety of destinations on completion of these courses. IT assessors reported that learners tended to progress to further academic or vocational study, often within the field of IT. Assessors on the Level 3 course reported learners progressing to "good courses" in Higher Education.

> *Like I said it recruits well, they go to university to do IT-related courses or they'll go and do apprenticeships; they use it.* Level 3 suite of qualifications (IT).

Of the 3 softly vocational qualifications, learners taking the Level 2 Award in construction appeared most likely to progress directly into employment. This VTQ was felt to give learners a good "taster" of roles available in the construction industry and some good basic skills to take to local construction employers.

> *I mean this area in particular, a lot of kids go onto apprenticeships and go into trades and things like that…the course does suit that quite well because it gives them real skills that you can go on and sort of get by, get a job with.* Level 2 Award (Construction).

3.1.2 Cohort attainment

A number of assessors discussed the level of attainment of the learners they taught. In some cases this was determined by course entry requirements. In the strongly vocational courses, centres usually insisted on learners having achieved relevant Level 1 or 2 qualifications or holding previous industry experience. Nevertheless, assessors found cohorts still varied in terms of attainment. This was often a product of prior industry experience; learners who had previously worked in industry were often able to "race through" through the course. Other learners had an aptitude for a certain vocation, compared to others for whom practical skills did not come so easily.

> *We've got a different level of students coming in. So I sometimes think that can be I suppose one of the challenges. If you get a student coming in who's a natural hairdresser, can work with their hands quite well etc. And then you get others that come in and they're really struggling.* Level 2 Diploma (Hair and Beauty).

Entry requirements for the softly vocational qualifications were more varied. The Level 2 Award in construction appeared to have no entry requirements to the course. In contrast, many of the centres delivering the IT qualifications required learners to have a certain level of prior attainment.

The stringency of course entry requirements was often reflected in the attainment of learners. In centres delivering the Level 2 softly vocational qualifications (in IT and construction), the level of achievement within each cohort was broad. Assessors felt these VTQs catered for a *"huge range"* of learners. In the Level 3 IT qualification, course entry requirements varied considerably between centres in our sample. As such, assessors in different centres variously described learners as being "the weaker students", a mixture of attainment levels, or academically able learners.

Notwithstanding this, most assessors delivering softly vocational courses believed their cohort was "less academic" than those taking other courses in the centre. They might struggle with exams or with expressing themselves in writing. Assessors felt the practically focused internally assessed tasks were well suited to these learners.

> *The students that take this sort of course are not very exam friendly. They take it because they think it's a practical, hands-on course, which I've found it to be very much so, actually a really kind of usefully practical course for them. But they struggled with the theory for the exam and they struggled with some of the written work.* Level 2 Award (Construction)

Some assessors also reported their learners found it harder to engage with education, particularly in the construction qualifications. Although learner engagement was not an explicit aim of any VTQ in our sample, some may be provided by centres with this purpose in mind.

> *You've got to understand the kids that are doing this, they're not like the well-behaved kids. They're doing a vocational course for a reason, because they don't like sitting in classrooms doing theory, they want to do practical stuff.* Level 2 Award (Construction)

Whatever the initial levels of motivation of learners, assessors in almost all centres highlighted the value of these VTQs in promoting learner engagement. The model of internal assessment played a crucial role in this. Most VTQs afford centres flexibility

when setting internally assessed summative tasks. This allows assessors to develop meaningful contexts for tasks, often drawing on links with local business or industry to inspire or engage learners. The continuous nature of internal assessment can also appeal to those who find terminal examinations particularly difficult or intimidating. Learners were also engaged by the practical focus of the course, the accessibility of the qualification and the variety of relevant optional routes through the course.

3.1.3 Alternative to academic route

Assessors talked about the value of these VTQs from many different perspectives. In 7 centres in our sample, interviewees discussed the value of softly vocational qualifications in providing an engaging and accessible alternative to academic qualifications in schools. Some were emphatic about the value these VTQs have for learners who find it hard to access or engage with academic qualifications.

> *A massive proportion of students aren't academically inclined and we're forcing them down a road that is unnatural to them. And some of the students that I'm dealing with now, they're just out of their depth; they need to be doing these sorts of courses where they can learn those sorts of skills.* Level 2 Award (Construction)

Some assessors explained certain learners are instinctively drawn to practical skills and learning. Having a vocational option which allowed these learners to develop such skills was vital in engaging them with education. One Level 2 construction assessor described the course as a "lifeline subject" for learners. Three assessors delivering softly vocational qualifications were emphatic about the importance of retaining these practically-focused qualifications in schools. They all believed these kinds of vocational qualifications had been "devalued" or "side-lined" in recent years.

> *There's a big thing in education where kids are pushed to go further into education when for some of them they need to go and get an apprenticeship or need to get a job and need to learn some skills... I think the practical subjects, especially the last few years, have been massively devalued by government.* Level 2 Award (Construction)

3.1.4 Assessment demand

Although we did not ask assessors specifically about qualification or assessment demand, most touched on this. Where discussed, qualification achievement rates were high – between 90% and 100% of learners taking the course achieved the qualification. Assessors believed everyone should be able to pass the VTQs if they applied themselves and had been appropriately selected for the course. Inevitably, some learners do fail these courses. However, failure tended to be less about learners not having the competence to pass and more about non-completion. Assessors talked about "silly fails" caused by capable learners dropping out. Family, health or financial issues may also contribute.

> *If we retain them, the vast majority will complete. I'm talking 90%, but there is always somebody that then doesn't come back, you know, or they won't complete their online tests or something like that.* Level 3 Diploma (Construction).

In all of the VTQs in our sample, there was a sense that demand was appropriate to the level of the qualification, offering both challenge and accessibility. In the strongly

vocational courses some practical skills may be inherently more challenging (cutting and colouring in hairdressing and roofing in construction for example), however this was felt to be reflective of industry.

In the softly vocational qualifications, assessors expressed more varied views on the level of demand. In the Level 2 construction and IT qualifications, assessors reflected less on qualification demand. There was a sense in both that courses were "pitched at the right level" and allowed assessors to challenge more able learners whilst also providing an accessible qualification for lower attaining learners.

> *I'm full of praise for this qualification. I'm a bit of an evangelist for it. I think it's really, really good. I think it's pitched at the correct level. It's challenging enough, but it's focused on what it needs to be focused on.* Level 2 Award (Construction)

> *We really enjoyed doing them. We really do see as a teacher giving the students the opportunity to work independently and seeing them flourish when they do… It's nice to have a different option whereby if you're a really keen independent learner, no matter your interest, you can succeed.* Level 2 Certificate (IT)

In both of these Level 2 qualifications, there is an option of awarding a learner a Level 1 pass if they were not working at a Level 2 pass. Assessors believed it would be difficult for a learner to fail to achieve this 'Level 1 pass' grade.

> *You can't really fail much. If the person applies themselves through training, not to devalue a level 1, but it is quite simple to achieve because it's an F grade. It's near enough a fail.* Level 2 Award (Construction)

In comparison, assessors delivering the Level 3 IT qualification stressed the high level of technical demand of the course. This was the only VTQ in our sample in which assessors felt learners may genuinely struggle with course content. Assessors in all 3 centres felt the course was equivalent to A level in demand.

> *They have tried to get this course up to the same standard as the computing course…It's a course that you would get something out of; as compared to say the previous course which is a bit of an easy ride. This isn't such an easy ride. So, in a way we've had to… up the standard of the student to do their course. Which isn't necessarily a bad thing.* Level 3 suite of qualifications (IT).

Although demand was high, achievement rates were reported to be good and the assessors we interviewed were mostly positive about the demand involved. That said, assessors in one centre questioned whether the course was pitched at the right level for the cohort taking it.

> *This is supposedly a vocational course that actually is highly technical, but the students aren't of that calibre. We've just finished unit one for the exam in January. Some of the kids are like rabbits in the headlights because of the technical content and the speed we had to go through things. It's just been really tough for them… It's supposed to be vocational but it's not vocational. Some of the areas, the coursework's are vocational; the examined units are not vocational.* Level 3 suite of qualifications (IT).

17

3.2 Area 1 VTQs and real-world practice

A recurring theme of our interviews with assessors was the relationship between the qualification and real-world practice. Essentially this is about validity – the extent to which a qualification's content and assessment reflects the skills and knowledge required for the relevant vocation.

Most assessors believed these qualifications to be highly valid in that they either prepare learners effectively for a specific career or more generally for the world of work. There was, perhaps unsurprisingly, a particular emphasis on this authenticity with regard to the 'strongly' vocational VTQs. For this reason, we focus first on the validity of strongly vocational qualifications before touching briefly on the relationship between softly vocational qualifications and real-world practice.

3.2.1 Strongly vocational qualifications

Assessors delivering strongly vocational qualifications stressed the importance of the validity of the assessment methods used in their qualifications, in particular the internal assessment of practical skills. In hair and beauty, assessors emphasised the validity of assessing learner in realistic working environments. In these courses, learners are assessed on paying clients in a commercial salon environment.

> *The learner would perform a treatment in a realistic working environment. So, we'd be in a salon as they would be if they were obviously out in industry. They could work on people they know at this stage but we often do get clients in for them that are unknown to them so it's a bit more realistic.* Level 2 Diploma (Hair and Beauty).

Hair and beauty assessors felt that an overly standardised assessment environment would fail to reflect the unpredictability and variety that someone would experience working in a salon. The notion that learners should develop a capacity to deal with "any client who walks into the salon", no matter the complexity of their needs, was considered an essential part of the qualification.

> *It just makes it more realistic. We're sending them out into an industry where they need to have worked on different people.* Level 2 Diploma (Hair and Beauty).

This may lead to a tension between assessment validity and reliability. As learners work on different clients during assessments, some may be faced with a client who is particularly challenging. The assessors we interviewed noted a certain "luck of the draw" in this but were nonetheless convinced this was the correct way to build authenticity in assessment.

> *There is that element of luck there I suppose but I think that's what the industry, that's life isn't it? […] I think we'd be wrong if we were teaching that the world is perfect and that every client is going to think that they're wonderful. That's the thing. So having those experiences with clients I think is important.* Level 2 Diploma (Hair and Beauty).

In hair and beauty, learners take dozens of internal summative assessments. Assessors therefore combine as many assessments as possible into one client visit, with learners providing multiple services in each appointment. Although they did this for reasons of manageability, assessors also felt this reflected real-world practice.

> *What should happen within a real working environment is that a number of criteria or a number of assessments can be completed at the same time. So there maybe a number of assessments that are involved in offering that service… They would have to start off with a consultation with the client. And then the service that they're providing, be that a cut or a colour, there's all the health and safety aspects that are involved in that.* Level 2 Diploma (Hair and Beauty).

In the Level 3 Diploma in construction there was less discussion of the validity of the assessment environment. Here, assessments generally took place in a workshop setting. In one centre assessors took learners out of the workshop for certain summative tasks, getting them to build real-life products in the college.

Although our interviews focused specifically on internal assessment, assessors were most likely to discuss validity in relation to qualification content, rather than the model of assessment. Generally, we found assessors to be positive about course content and how this related to real-world practice and industry. Most assessors delivering strongly vocational courses felt they prepared learners well for roles in industry. That said, over half of assessors delivering strongly vocational courses felt course content had not always kept pace with practice in "fast-moving" industries. In most cases disparities between course content and industry practice were felt to be quite minor, a small inconsistency in what is otherwise a valid qualification.

In 7 interviews, assessors highlighted skills which they felt to be "outdated". Five hair and beauty assessors discussed techniques such as cap highlights and mixing face masks which now had little place in industry. Others believed that newer techniques such as gel manicures should be incorporated into specifications.

> *I do feel that they're not moving with the times. They're not evolving. I mean this manicure unit I've been teaching for 11 years, this is still exactly the same as it was 11 years ago. It hasn't changed. And, to me, it's got to change. Again, for instance, I think it should state in the range here, it should have gel polish…because I'd say 70% of the manicures that are done in the industry now, they're having gel polish.* Level 2 Diploma (Hair and Beauty).

Two assessors delivering the Level 3 Diploma in construction also highlighted disparities between the qualification and real-world practice. One group of assessors also noted that industry may be using certain tools and products that assessors don't have access to in centres.

> *I think they produce the standards and the criteria in consultation with employers. Some of it does seem terribly old fashioned I think… again we get that complaint from students, we don't do this at work, we wouldn't do this at work, it's not how we do it. But we're covering the criteria we get from them and they don't necessarily do it like that onsite.* Level 3 Diploma (Construction).

Some assessors highlighted other areas of disparity between real world practice and the qualifications. Seven suggested the level of performance required by employers may not always align with the level required for the qualification. In some cases, qualification requirements were perceived to be higher than standards required by industry. For example, construction assessors observed that employers often want carpenters to work more quickly but to a lower level of quality than required on the

Level 3 course. The changing nature of parts used on building sites also meant that carpenters didn't need to be so skilled to do the job required of them.

> *If you're asking a foreman onsite then he's looking at this thinking this is irrelevant to what I'm doing where I've got to build 4,000 houses. This joint is great if you're making door levels, but they're not anymore; they're coming pre-fabricated. Door sets, doors are coming hung with the locks already in it. Generally, for new builds… to use the word carpenter is a bit, I don't know, it doesn't sit well. On the new builds especially you've got fitters, and that's what they're looking for. They're looking for woodworkers, construction workers that can work safely, use the tools and assemble things, read drawings and assemble.* Level 3 Diploma (Construction).

In other cases industry standards were apparently higher than those defined in the qualification. Often this difference was only ever hinted at, with assessors suggesting learners might need to do further practice "to get that industry standard up a bit". Two hair and beauty assessors felt that timings for beauty treatments were generous compared to industry.

> *You need to be hitting 20 minutes for half leg wax; whereas the awarding body will give them longer. They'll give them 30 minutes. But out in industry you'll get 20.* Level 2 Diploma (Hair and Beauty).

3.2.1.1 Preparation for work

The assessors we interviewed delivering strongly vocational courses emphasised their role in supplying workers to local industry, taking pride of the achievements of their former learners. Many were keen to train learners to a level beyond the standard of the qualification so that those learners would "stand out", find work, and enhance the reputation of the school or college. This concept of preparing learners effectively for industry was absolutely paramount in strongly vocational courses.

> *They've got to compete for jobs… so that's what we try to do really is try to do something different so that our students, it's obvious that they're from this college because they've got all this extra attached to them which would make them the employee of choice.* Level 2 Diploma (Hair and Beauty).

> *The last thing we'd ever want to turn out is a student that won't be employable.* Level 2 Diploma (Hair and Beauty).

Many assessors delivered further technical training to give learners the best chance of gaining employment on completion of the course. Hair and beauty assessors put on extra sessions to ensure learners could carry out treatments to industry timings and had encountered a wider range of clients than specified in the qualification.

> *The other thing we do as well, which wasn't built into the NVQ, is get them to work on male clients. Because again, especially now, male clients regularly visit beauty salons. And I think it wouldn't be fair for us to throw the students out from here and they've never touched a male client, because that's just a very scary experience for them again then. You're not preparing them for the big wide world really.* Level 2 Diploma (Hair and Beauty).

Two groups of Level 3 construction assessors also told us they provided learners with additional technical skills which were not covered in the specification.

> *With that project, it teaches them a little bit of shuttering, which OK isn't in there [the specification]. But every bit of extra knowledge isn't going to hurt… So if somebody says I want to train a shuttering carpenter up, we can then say OK, these guys have done this; let's send one of these over and see if they can get a job and become a shuttering carpenter. It does help them to have that extra knowledge that they need to go out there and have a better chance of getting that job which is what it's all about at the end of the day. We want them to go into full-time employment and be prepared for working life.* Level 3 Diploma (Construction).

Not all of this additional preparation for employment involved technical training. Sometimes it was about improving the general employability skills of learners. This was mentioned in 7 of the 11 centres delivering strongly vocational qualification. Preparation might involve training learners in general standards of professionalism such as appropriate dress and timeliness. Some centres provided learners with training on interview techniques or how to approach "trade tests".

> *We want to be clear that they're going into the industry well equipped and they're going to make, not just a therapist, but a really good therapist... We've said to them when you go for your trade test, you'll have like 10 people, 20 people, there for the same job, but you've got to stand out from the crowd. How do you stand out from the crowd? How do you do that? This is how you stand out.* Level 2 Diploma (Hair and Beauty).

> *You can have anybody you like. They can be a brilliant guy, brilliant lady, whatever. But if they don't turn up, or they're always late, or seem to want to go home every day early, they know that that doesn't work, that it isn't like that out there anymore. They are preparing for working life, which is what it's all about.* Level 3 Diploma (Construction).

Most hair and beauty assessors spoke about the (sometimes extensive) additional training they provided to learners in communication and interpersonal skills in order to turn them into a good hairdresser or beauty therapist. These skills are only touched upon in the learning outcomes for the qualifications, and a number of assessors in our sample felt that, in order to flourish in industry, learners needed additional coaching in this area. For young learners especially, this focused on building confidence in conversing with clients, use of appropriate language and topics of conversation, and dealing with difficult clients.

> *We need them to get a job in this profession. To be meaningful, to be able to get a job at the end of it, is what our purpose is. They're not going to get a job in the salon if they can't speak to somebody.* Level 2 Diploma (Hair and Beauty).

Preparing learners to work in local industry was a central part of these strongly vocational qualifications and how centres operationalise the delivery of them. Links with local businesses were considered to be symbiotic. Local businesses were keen to have access to a source of skilled new employees while the schools and colleges

could use local businesses to enhance their teaching. Some centres had strong formal or informal links with local businesses and drew heavily on these links to secure work experience, gain access to equipment or to organise training from employers to give learners a sense of the realities of industry.

The assessors in our sample delivering strongly vocational qualifications were well placed to understand the needs of local industry given that they all had significant industry expertise. Most still had close links with industry, particularly in hair and beauty VTQs where assessors are required to carry out a certain amount of work in salons each year as part of their CPD.

> *We all still work doing clients. And you've got to, because you've got to pass on what's going on in industry to your students.* Level 2 Diploma (Hair and Beauty).

There appeared to be no similar CPD requirement for construction assessors. Although these assessors had previous industry experience they did not always find it easy to keep their knowledge up to date.

> *Everybody comes from industry, but it's difficult. For somebody like me who's been here in FE now for a long time, it's hard to keep up … Most of us do private work and things like that, but it's not really keeping up to the real cutting edge.* Level 3 Diploma (Construction).

3.2.1.2 Softly vocational qualifications

Perhaps unsurprisingly, given that softly vocational qualifications do not aim to prepare learners to directly enter a specific job role, there was far less discussion of how these qualifications compared to the realities of industry. Where assessors did comment on the validity of these qualifications, they were generally positive that these VTQs give learners the sort of skills employers need.

In the IT qualifications, developing tasks which were authentic to real-world practice was seen as something that could engage and motivate learners. One interviewee also suggested that such tasks offered learners the opportunity to develop the kind of soft skills which help to make learners "work-ready". In IT qualifications in particular, course content was felt to be particularly up to date and relevant. In the Level 2 Award in construction, assessors also believed the course gave a good relevant foundation for learners interested in a range of construction fields.

> *And the nice thing about the course, it was designed with input from a lot of big construction companies and aimed at what they want from somebody entering the construction industry at a young age.* Level 2 Award (Construction).

These construction assessors were especially positive about the validity of the internal assessment of practical skills in the qualification. In 2 cases, however, assessors felt the focus on practical skills could be strengthened even further, suggesting an emphasis on reading and writing skills was reducing the validity of the qualification.

> *So you could have a kid that's an amazing builder, done all the practical absolutely perfectly, but a bit sloppy on their written work and has missed certain things out and they won't do as well as*

> *someone who's been very sloppy with the practical but has been quite thorough with the written work.* Level 2 Award (Construction).

Two assessors noted that the standard of performance required in this Level 2 Award in construction was lower than would be required in industry. Both felt this was appropriate given the age of the cohort.

> *Because 15-year-old kids can't really do that; they're not professional bricklayers. So [AO name] don't particularly specify that. They say the kids have to set their own success criteria.* Level 2 Award (Construction).

Unlike in strongly vocational qualifications, assessors did not discuss any additional enrichment activities to prepare learners for work. Assessors also seemed to have fewer links with local employers, although 2 centres delivering IT qualifications arranged visits and talks from local employers.

3.3 Area 2: Setting Tasks and Supporting Learners

We found great variation in approaches to task setting and task taking in the VTQs in our sample. Often this reflected different AO requirements for each qualification. Even within the same qualifications, however, centres described different approaches to setting tasks and supporting learners. To some degree this was due to the considerable flexibility given to centres in the design and delivery of internally assessed tasks. Although assessment criteria are usually tightly prescribed, the way in which learners can meet them is not. There was also some evidence that centres have differing understanding of AO requirements in the VTQs they deliver.

3.3.1 Task setting

Given the diversity of VTQs in our sample, we found great variation in the nature of internally assessed summative tasks. In hair and beauty and construction VTQs, internally assessed tasks usually comprised practical observations of skills and written assessments of knowledge and understanding. In contrast, in the 2 school-based IT qualifications, assessments generally took the form of written tasks.

In all 6 VTQs, centres have some degree of autonomy in task setting, particularly in written components. The level of flexibility centres are given varied considerably between and even within qualifications. In some units, AOs strictly prescribe the nature of summative tasks. They tightly define the task taking environment and the time allowed for completion. In other units, centres were free to devise their own tasks so long as these met the requirements of the specification. Assessors valued this flexibility highly. Many felt it was central to the delivery of a valid and engaging qualification within the constraints of limited centre resources.

AOs retained by far the highest level of control over summative tasks in the strongly vocational qualifications in our sample. Even here, there was disparity between the practical and theory tasks. Practical tasks are tightly specified. In hairdressing, for example, learners must perform a series of specific cutting and styling services. These must be demonstrated on a pre-defined 'range' of clients to ensure learners are competent with different hair types. Learners are given a strict time limit to complete the task. The only flexibility centres have here is in the scheduling of the assessments and the selection of clients. Centres must use paying clients for summative tasks, but these may include friends or family of the learner.

> *The whole process is laid out in black and white in front of you, so we don't really veer off from that.* Level 2 Diploma (Hair and Beauty).

Although assessments of practical skills in hair and beauty are tightly controlled, centres have more flexibility when assessing knowledge and understanding. Tasks can be adapted to meet the needs of different learners. As long as learners have the knowledge required by the specification, the way in which they can evidence this is not rigid. In one of the hair and beauty VTQs, assessors design workbooks to assess knowledge. These may be supplemented with assignments or oral questions. In the second VTQ learners can choose whether their knowledge is assessed through an external online test or through AO-set and centre-marked assignments.

> *The awarding body give us what the students need to know… That is what we write our work booklets or our assignments around. But*

> *we've found a combination of workbooks, assignments, oral questions covers it best for the students, and they tend to cope with that the best as well.* Level 2 Diploma (Hair and Beauty).

In the 2 construction qualifications all learners undertake exactly the same tasks. Practical tasks are most tightly defined in the strongly vocational Level 3 Diploma where the AO sets a series of prescribed tasks. In 2 of the 3 centres we visited, assessors followed these tasks exactly using a designated assessment 'rig'.

> *They're standard tasks that they have to be assessed on. And the standards that they have to meet, the tolerances and time limits that they have to meet… are set by the awarding body.* Level 3 Diploma (Construction).

In a third centre, assessors believed that whilst tasks are clearly defined, it is possible (and indeed desirable) to be flexible in how they are executed. Where they could, they devised summative tasks where learners were assessed on real-life projects, rather than making a product solely for the purpose of the task. For example, in one unit learners are assessed on a piece of curved carpentry work. At the time of our visit to the centre, learners were building a bridge on the campus for their summative assessment.

The softly vocational Level 2 Award in construction allows more flexibility in task setting. The AO provides model written and practical tasks but assessors can modify these. For the practical element, centres are given the scenario of a guest house in need of a range of repairs. They choose 3 skills from a list to suit their local facilities and expertise. Although the AO provides an assessment brief for each skill, all the centres we visited had modified this.

> *There's quite a lot of flexibility there. Apart from the exam, the other 2 projects were fairly flexible.* Level 2 Award (Construction).

There was some confusion amongst these assessors about the need to gain formal sign-off for modified tasks. In 2 of the 3 centres, assessors believed the AO must formally endorse changes to model tasks before they are used. In the third, the assessor believed no sign off was required, and tasks were checked during the external quality assurance (EQA) process.

In contrast to the other VTQs in our sample, assessors delivering IT qualifications enjoyed great flexibility in task setting across all units. They are given free rein to devise summative assessments as long as these met specification requirements. These centre-designed tasks do not require sign-off before use.

> *We're entirely flexible. The coursework has to meet the grade descriptors set out in the unit descriptors. And it doesn't really matter what that digital product is about as long as it meets the descriptors for the unit.* Level 2 Certificate (IT).

In both IT qualifications, some of this flexibility in task setting is passed onto learners. In 5 of the 6 centres visited, learners are given a certain autonomy over their summative tasks in some units.

> *There's a bit of freedom given to the students as well in terms of, as long as they can justify why they're using the IT, that's the main point. And if they want to go about it in their own direction, that's fine.* Level 2 Certificate (IT).

Overwhelmingly centres found the flexibility within these VTQs to be hugely valuable. In particular, assessors discussed the role flexibility has in promoting engagement of learners with summative tasks. This was mentioned in 8 centres, 5 delivering IT qualifications and 3 delivering construction qualifications. In IT, flexibility allows assessors (and learners) to choose projects and contexts which most engage. In construction, flexibility in task setting can also help make tasks more meaningful.

> *So if we've got to do a relational database for example, we might use Santa's naughty and nice list… We'll do whatever engages kids. And so that flexibility has been really quite useful.* Level 2 Certificate (IT).

> *I mean, last year we produced a staircase, which is used in [college department]… Not only was it an assessed piece of work that they did, the fitting of it was assessed as well. And so it takes it away from being a flat qualification and then you come in, you've done this, you've worked on a bench into something that is real.* Level 3 Diploma (Construction).

The second most cited benefit of flexibility related to the timing of assessments. In 7 centres, assessors praised the flexibility to combine tasks. This was crucial in hair and beauty, where learners are assessed on dozens of tasks. By combining these into larger assessments, qualifications become more manageable to deliver.

> *They've got to start to think it is more seeing that person as a whole to try and see what else they can offer. Because it's clumping the assessments together that makes it a lot easier; otherwise they end up having 40 plus individual practical assessments they have to do.* FE College. Level 2 Diploma (Hair and Beauty).

Six assessors (predominantly in the field of hair and beauty) praised the flexibility they have to tailor written tasks to meet different learner needs and stretch and challenge the most capable. Five interviewees also discussed the value of tailoring tasks to allow learners to get a more 'real-world' experience. IT assessors explained how flexibility in task setting allowed them to use local businesses as clients, creating highly valid task contexts and exposure to the world of industry.

> *Having this proper client, you are having to find out what they want of the application. You're finding out what their problem is, what they want the application to do. It could be a mobile phone app, it could be a website. But it's actually making it as real as we possibly can in our artificial environment.* Level 3 suite of qualifications (IT).

Another important benefit of flexibility was that it allows centres to adapt assessments to their own facilities. All Level 2 construction assessors interviewed experienced some resource or space issues in their schools.

> *So what they say is the brickwork should be a quoin, which is a like a corner of a wall that's damaged and you repair that. But it takes up too much space so I spoke to the chap at [AO name] and sent him plans and we've modified it a bit to do a straight wall… They let you adapt it to suit your centre and facilities.* Level 2 Award (Construction).

Whilst flexibility in task setting promotes the validity, engagement and manageability of assessments, variation in tasks may have some implications for assessment reliability. Only 2 assessors brought up this risk unprompted. Both were positive about flexibility, but questioned how it affects comparability between centres. One felt there was a difference in demand between the practical skills centres can choose to teach, the other questioned whether different interpretations of model tasks might affect assessment demand.

> *It plays to the strength of the teacher. I guess with that comes the concern that is this a level playing field? Is somebody who's got to build a wall doing the same as somebody who's got to tile a wall?* Level 2 Award (Construction).

In other interviews, there was little concern about comparability of task demand. Three hair and beauty assessors acknowledged that differences between clients may have a small effect on the demand of practical tasks. They agreed some clients could be more "difficult" than others, either because of physical characteristics or behaviours. They felt that this was reflective of real-world practice and that learners should be equipped to deal with any client that walks through the door of their salon.

In assessments of knowledge in hair and beauty, assessors believed different methods of eliciting the same technical knowledge from learners were equally valid. Although some learners might find certain assessment methods easier than others, they believed this was unrelated to the construct being assessed, and more about learners' skill in expressing themselves in writing.

> *The majority of our students do the online testing. Mainly because it's easier in the sense of they find it more straightforward to sit and learn and revise than they do to actually try and get their thoughts onto paper in a certain context.* Level 2 Diploma (Hair and Beauty).

In IT qualifications, assessors were similarly unconcerned about comparability. In 4 interviews assessors explained whilst task contexts may vary, the knowledge and skills that learners are assessed on are ultimately the same. For example, in one unit of the Level 3 IT suite of qualifications, learners have to design an application (app) as part of an assessment. Assessors explained that the process learners go through to develop this app would be the same, regardless of its content.

> *Unless they want to make it really hard for themselves for some bizarre reason, it should not be that much different. Because at the end of the day, we will be looking at a student who's done a plumber's interview and a student who has done an interview for a second-hand car shop. We're looking at 2 pieces of interview. So the task is very similar.* Level 3 suite of qualifications (IT).

Overall flexibility is a highly valued aspect of task setting in internal assessment in these VTQs. Although it is quite possible that standardising tasks more tightly may improve assessment reliability, this would have implications for the validity, manageability and engagement of internally assessed summative tasks.

3.3.1.1 Designing summative tasks

Whilst centres talked at length about the nature of tasks, there was little discussion of the process they go through to design these. What limited information we did elicit

suggests assessors use a range of approaches to develop summative tasks, with varying degrees of input from AOs into this. Assessors appeared confident with this part of their role, particularly in the more long-established qualifications such as hair and beauty. No one expressed reservations about their ability to design tasks.

In designing summative tasks, a handful of assessors approach either their External Quality Assurer (EQA) or another AO contact during task setting to confirm the task meets the specification. This might be an informal interaction or as part of a formal sign-off process. Other centres told us they have no direct input from AO personnel during the design process, but view the EQA process as a sign-off for tasks after the assessment.

> *Your formal process will be when you first submit your evidence off. Like, so if I submitted that off to [AO name] and they come back in their moderator's report say the assignment brief doesn't reflect the criteria, or they came back and said it's not fit for purpose, then obviously I haven't satisfied the [AO name] requirements. But I've heard nothing back… so I know what I do is OK.* Level 2 Award (Construction).

Three centres drew on a community of practice when devising tasks, seeking support from neighbouring centres or online communities. This was more prevalent in the early days of a specification. It was also more common in the softly vocational qualifications, where the greatest flexibility in task setting is experienced.

Two centres developed summative assessments by adapting tasks used in other VTQs they have delivered in the past. These pre-existing tasks are tweaked to meet the requirements of the new specifications.

> *I've got lots of resources because I've been doing it a long time. And then the specs change, I just adapt what I've got already, just to ensure that it matches the new spec exactly.* Level 2 Diploma (Hair and Beauty).

Levels of guidance provided by AOs to support task design appeared to vary. It is not clear if this reflects a genuine variation in the amount of guidance or that centres delivering different VTQs simply engaged with this support and guidance differently. In some VTQs assessors were very positive about the amount of written or verbal guidance received, in others opinion was more mixed. Assessors were particularly likely to highlight issues with a lack of written guidance in the early years of a course.

> *I was able to check out with the chief examiner for the course. Right at the beginning, I was talking through the course with him on the phone. He did give me a lot of helpful information, which I'm not sure I would have got otherwise.* Level 2 Award (Construction).

> *I think at first when the course was introduced, it would have been helpful to see maybe theory examples in terms of that level of detail. Because if they're doing a distinction task and it says as an extension of pass level 1, what is an extension? How much is expected for that extension? But we seem to do it all right now.* Level 3 suite of qualifications (IT).

3.3.2 Task taking

Although not a main focus of the interview schedule, most interviews touched on the task taking environment. The level of control associated with this varied significantly between (and within) VTQs. Practical tasks in strongly vocational courses take place in the most controlled environments. In hair and beauty, practical assessments must occur in a realistic working environment; namely a "fee-pay salon*"*. All of the colleges we visited had a fully functioning, commercial salon onsite for this purpose.

> *We cannot assess unless it is in a realistic working environment. If it's done in a workshop session it's not a summative assessment; it has to be in a commercial session, because it has to be done on fee pay.* Level 2 Diploma (Hair and Beauty).

Hair and beauty learners work independently during summative tasks. Assessors in one college described how the salon was sectioned into cubicles to prevent learners looking at one another's work. During tasks, assessors circulate the salon checking on any concurrent assessments. If they miss a stage in an individual learner's work, they use oral questioning to establish a process had been carried out correctly. One private training provider believed oversight of tasks was more rigorous in centres with smaller cohorts. They suggested assessors might struggle to keep track of multiple assessments where larger cohorts were involved.

> *We've only got a small number, so they're all more or less on a one-to-one with you. Whereas, when you teach and you've got 28 students in a class you're teaching, you're assessing, you can't be watching every single student, can you, all the time?* Level 2 Diploma (Hair and Beauty).

In construction, summative tasks take place in a workshop environment, with assessors circulating at a distance and asking learners questions periodically.

> *The teacher is not assessing them constantly - they will be coming backwards and forwards. Which I think actually is better because it is difficult for you if someone's standing over you for the whole 10 hours would be just, it's not practical and not very conducive either.* Level 3 Diploma (Construction).

Unlike hair and beauty, where assessment is more 'on demand' and driven by client availability, learners almost always started tasks at the same time in the construction VTQs in our sample. Learners also have some freedom to talk to each other during tasks and can look at one another's work. Whilst they should not help one another with their work, 2 assessors found it difficult to prevent this.

> *When it's the assessment, it can't be exam conditions. Sometimes they might point things out to each other. I mean, you can't keep a workshop silent... When they're laying bricks, their heads are maybe a foot apart from each other, so you can't really stop them talking. But they get on with their work. They do. They have a little chat.* Level 2 Award (Construction).

In all the practical tasks above, AOs set a time limit for completion. There appeared to be a difference in approach to enforcing this between different sector subject areas. In hair and beauty VTQs the time limit was strictly enforced by assessors; if

learners even slightly exceeded it they fail the assessment. In construction, assessors felt they had more flexibility with interpreting any time limit.

> *It's got to be bang on industry time as they would perform in the workplace.* Level 2 Diploma (Hair and Beauty).

> *If the task is supposed to take 10 hours, I think you are able, if they've taken 10½ hours, that's not necessarily that they would fail; but obviously if they took 20 hours then that's not.* Level 3 Diploma (Construction).

There was much greater variation in task taking conditions in written tasks. At one end of the spectrum was the Level 2 Award in construction. Centres ran the written assessment as a controlled assessment, although one assessor was unsure whether this was mandated by the AO or a centre-devised strategy. Centres enforced strict time limits for tasks and learners worked on their assignments in silence.

> *This is controlled assessment. And they would sit apart from each other, it would be in silence and I'd make it clear that they weren't going to get help.* Level 2 Award (Construction).

At the other end of the spectrum were the IT qualifications. In the Level 3 course, learners discuss their work with one another and may even peer review each other's assignments. Work can be taken home to complete.

> *They do talk to each other. The year 13, particularly, if we would explain something, and you'll hear some of them saying what does such-and-such mean and you will hear them talking about it.* Level 3 suite of qualifications (IT).

As well as variation along qualification lines, there was also some variation in delivery across different centres providing the same VTQ. For example, in the Level 2 Certificate in IT, different assessors described task taking conditions variously as "low control" or "very controlled".

A minority of assessors admitted that they had a slight lack of clarity about AO requirements for task taking. As they were outlining their arrangements to us, 3 assessors questioned their understanding of AO rules. In some cases this appeared to be caused by the sheer number of courses delivered by assessors causing a degree of confusion about individual qualification requirements.

3.3.3 Support for learners from assessors

During our interviews we asked assessors about the level of support and feedback they give to learners during summative tasks. Most assessors told us they were clear about the amount of support they could provide to students during assessments. That said, it was often hard for them to articulate this and these interviews have not always given us a clear sense of what the limits of support are for each qualification.

Once again, we found variation in the timing and nature of support provided to learners. Even within the same VTQs, assessors sometimes had different approaches to supporting learners. In some cases there was perceived to be a lack of clarity in AO guidance, in others there seemed to be a lack of centre familiarity with AO rules.

In assessments of practical skills, the critical time for support is during the practical task itself. For written assessments, discussion focused on feedback provided after work had been submitted by the learner. For that reason we consider practical and written tasks separately.

3.3.3.1 Practical tasks

Learners are assessed on their practical skills in hair and beauty and construction VTQs. In these VTQs, assessors are limited in the amount of support they can give learners during summative tasks. Despite this, we found evidence that different centres may have different interpretations of what support was appropriate.

The lowest levels of support seemed to be provided in hair and beauty VTQs. In one course, assessors told us that they were unable to give *any* feedback to learners during practical tasks.

> *You can't help with the summative. Yeah it's completely unaided.*
> Level 2 Diploma (Hair and Beauty).

In practice, there appeared to be different interpretations of this concept of 'no support'. Whilst one assessor explained they were unable to prompt learners at all if they were failing a task, another told us they give hints to learners that improvements were required before they could be passed.

> *So if they ask us at all during that assessment it's a fail, or if they're doing it wrong it's a fail, but we can't prompt them.* Level 2 Diploma (Hair and Beauty).

> *I will then say to the student right, OK, Kaylie, how would you improve that nail there? So I'm asking for them, I'm prompting them, to say you need to improve that… Then I give them the chance to put it right before I've given them the final check over. I don't tell them what they've got to do.* Level 2 Diploma (Hair and Beauty).

We also found variation in the second hair and beauty VTQ. In 4 of the 7 interviews, assessors told us initially they could give no support to learners during summative tasks. They then went on to describe examples of support they might give. In fact, at some grades in this VTQ, the assessment criteria permit "minimal support" from assessors. Some assessors noted a certain ambiguity around this concept of "minimal support".

> *One person's minimal support is different from another person's minimal support. But that's why we discuss it and we clarify between ourselves this is minimal support, that is not minimal support.* Level 2 Diploma (Hair and Beauty).

In the construction qualifications, there was less clarity amongst assessors about the acceptable level of support during practical tasks. For example, in the Level 2 qualification the assessment criteria reflect the fact assessors may give learners some support during practical tasks. Assessors told us there is a degree of "professional judgement" required here. They admitted they were not always confident they were complying with AO rules when supporting learners.

> *It's something I continually ask myself… I've never been entirely clear where the line is. We were given some examples of what would be acceptable and what wouldn't be acceptable, but there*

> *does always appear to be a bit of a grey area.* Level 2 Award (Construction).

This is illustrated in the fact that 2 assessors described different models of what an acceptable level of support or intervention might look like.

> *Common sense says that means on the odd occasion I might need to intervene and just say no, actually that tile, you need to do a little bit more on that block.* Level 2 Award (Construction).

> *So I make a point of not intervening with them. I'm there. If they want to ask a question, I'll answer questions and things like that. But if I spot a mistake, I won't go and correct them because you wouldn't do that.* Level 2 Award (Construction).

3.3.3.2 Written assessments

Assessors seemed to be clearer about the level of support or feedback they could give to learners undertaking written summative assessments. Perhaps because of this, there appeared to be greater consistency in practice across centres. Despite this, there were still "grey areas" identified by assessors.

Assessment in the IT qualifications in our sample was based on written tasks. Assessors provide limited feedback to learners as they work on tasks; for example they may ask leading questions to learners or respond to their questions.

> *We can't obviously tell them what to do but...you can ask leading questions. So you can say what do you think about this or what might the audience need or?* Level 2 Certificate (IT).

If it became clear that several learners were struggling with a particular concept or skill during task taking, 2 assessors told us they re-teach the relevant area of the specification to the whole class.

We found IT assessors had varying levels of confidence in their approach for supporting learners during tasks. Some reported being very clear on AO requirements for learner support, others were less certain about where the line of acceptability lay in terms of support, feedback and use of model answers.

> *Where do you draw the line? And actually there is no clear 'how much is too much help'.* Level 3 suite of qualifications (IT).

Even where assessors reported greater clarity of AO requirements, we still found that centres took different approaches to supporting learners during tasks. For example, in the Level 2 Certificate in IT, 2 assessors told us they provide very little specific help to learners. In a third centre, the assessor explained their school had a "low level intake" which required sometimes extensive support.

> *No, no. It's not strict… Schools like this, our nature is to help kids as much as they need. Some kids need loads of help; some kids will just go and do it themselves. So it just depends on the kid really.* Level 2 Certificate (IT).

Responses were more consistent when describing the level of feedback provided to learners after the submission of their work. In both IT qualifications, assessors can feedback to individual learners after submission. All assessors agree feedback

should not be too specific. They can identify general areas for improvement or ask questions to learners to prompt them to identify improvements themselves.

> *We can't tell them how to improve, can we? We can say what they need to improve, but we can't say how specifically they've got to go about it. But we can say things like revisit the spec points; check you've covered all areas, things like that. But we can't say ooh you've missed out this particular bit. We can't be too prescribed with it.* Level 3 suite of qualifications (IT).

In the Level 2 IT Certificate, it appears that assessors can give any number of rounds of feedback to individual learners to improve their work. This was not the case in the Level 3 qualification, although only one centre specified any limit (in this case assessors could give 2 rounds of feedback before a final submission was made). Overall, we found understanding of the level of support that assessors can give to learners is variable. Even where assessors claim to be clear on the requirements of the AO, they may then go on to identify areas of ambiguity or describe differing views on the level of acceptable support and feedback. This leads to variation in practice across centres which could have implications for assessment standards. Amongst those who were confident in their approach to supporting learners, there was sometimes a sense this was due to their experience as assessors, rather than their understanding of AO requirements for the specific VTQ.

> *I think it's more experience, isn't it, because we've been doing it for so long, so we kind of know.* Level 3 suite of qualifications (IT).

One potential challenge for assessors when providing support and feedback to learners during summative tasks is the nature of their dual teacher-assessor role. In our sample, all assessors taught the course that they assessed. On one hand, as teachers and lecturers they are responsible for developing and supporting their learners to do their best on the course. On the other, as assessors they are responsible for upholding assessment standards through making sound, objective assessment decisions. None of our participants discussed any such potential for conflict between these roles when discussing learner support. Nonetheless, their dual role may well give rise to tensions when determining an appropriate level of support and feedback, as well as making any resulting assessment decisions.

3.4 Area 3: The assessment process

Almost universally, the assessors we interviewed were confident in their ability to make assessment decisions using the assessment criteria. That said, we found considerable ambiguity and inconsistency in assessor responses. Many initially emphasised the ease of applying the assessment criteria, only to identify vagaries later in the interview. It seemed confidence in applying assessment criteria was not solely determined by the clarity of criteria. AO support and guidance, experience of delivering other qualifications, communities of practice and experience of industry also play a part in determining an assessor's confidence in making decisions.

In most VTQs in our sample, assessment criteria take the form of a series of tightly specified tasks which a learner must complete. Criteria are quite atomised, with each task broken down, element by element. Assessment decisions are usually binary; assessors determine whether competence has been demonstrated (or not) for each discrete element. In construction VTQs, learners may have to produce work to precise measurements, working within tight tolerances.

Generally assessors find these types of decisions straightforward to make, although they usually find some "grey areas". Although we might anticipate assessment decisions being less straightforward in qualifications with a number of grades above a pass, there was no strong evidence that assessors find this. Indeed, there was no pattern evident between different VTQs at all. Assessors delivering the same qualifications expressed different views on the clarity of assessment criteria. In some cases, we even found divergence of opinion on the ease of applying criteria within centres.

Despite this slightly ambiguous picture, most assessors told us they were confident applying the assessment criteria. In 19 of the 25 interviews conducted (covering all 6 VTQs), assessors agreed that judgements were "easy" or "straightforward". Although this is an encouraging finding, we should acknowledge that 'self-selection' may play a part here. It is debatable how many assessors who were not highly confident in their role would consent to take part in an interview with Ofqual.

> *It was very easy to understand what we needed to deliver, first and foremost, and easy to understand what the students needed to do to pass a particular criteria.* Level 2 Certificate (IT).

Most assessors attributed their confidence in making assessment decisions to the design of the qualification or the nature of the skills or knowledge being assessed. Over half explained it stemmed from the clarity of assessment criteria. In hair and beauty, in particular, many assessors described criteria as "black and white".

> *It's a tick box exercise, and if she's been watching what that student's been doing there's not much manoeuvrability.* Level 2 Diploma (Hair and Beauty).

In 4 interviews, assessors suggested the grading model of the VTQ promoted the ease of assessment decisions. In these VTQs, judgements are pass or fail, meaning there is no need to make finer judgements between a number of grade bands.

> *There wasn't really any grey area. Not like with other qualifications which have 6 different mark bands, and it's really difficult at times to know exactly where your work sits within those bands; this was much simpler for a teacher to understand.* Level 2 Certificate (IT).

Although confident, few assessors gave completely consistent responses. In 9 interviews, assessors initially agreed assessment criteria were unambiguous before identifying "grey areas" in them. There was little specific discussion of what makes criteria difficult to apply, but this might include vague wording or a lack of clarity about what characterises performance at different grades. In another 4 interviews assessors explained that whilst they found criteria simple to apply now, this hadn't always been the case. It had taken experience of delivering the VTQ, feedback from EQAs or regular standardisation to build their confidence in making judgements.

> *Until I did my first one I had no idea what the expectations were…*
> *And it's only now after a year and a half that I can look and I have*
> *greater understanding of how it actually works. You come into it cold*
> *and you just don't have a clue.* Level 3 suite of qualifications (IT).

Despite some inconsistency in their responses, the assessors above were still confident overall in making assessment decisions. This was not the case with all interviewees, however. One group of construction assessors described criteria as "vague". Although the AO specifies tight tolerances learners must work within when producing carpentry work, assessors were unclear how these tolerances should be applied. This group felt that different assessors may interpret criteria differently.

> *If you look at the standards in there I think it says within a millimetre.*
> *But it doesn't tell you what part of it within a millimetre. It's a bit*
> *vague… And it doesn't say how many small errors you can have,*
> *just says within. The interpretation basically is up to you.* Level 3
> Diploma (Construction).

A group of hair and beauty assessors also believed assessment criteria were open to interpretation. They felt different assessors might have different interpretations of the wording of the assessment criteria, as well as where the boundaries lie between grades. They undertook a significant amount of standardisation activity as a result.

> *I think it can be quite ambiguous at times because it's the assessor's*
> *interpretation of what the descriptor actually is… One person's*
> *'explain' might be different to another person's 'explain'. So they still*
> *might have included the 'how' and 'why' to explain. But for one*
> *person one 'how' and one 'why' might be enough to explain; for*
> *another person it might be 3 'hows' and 3 'whys'. So it's still open to*
> *interpretation.* Level 2 Diploma (Hair and Beauty).

It was interesting to note that even within centres there could be different views on the clarity of assessment criteria. In 2 centres, interviews revealed very different views on clarity between assessors. For example, in a centre offering the Level 3 IT qualification, one assessor described criteria as "airy fairy" and "open to interpretation", whilst a second assessor explained that "there's no grey area".

3.4.1 Grading

In most VTQs in our sample, assessment decisions were binary. Assessors decide if a learner has met a minimum acceptable standard and, if so, award a pass. Nonetheless, these VTQs still have a range of different approaches to grading at qualification level. Some were entirely ungraded; some were graded on the external exam only; others were graded on internal assessment. Even within the latter group, grading models were very different. In one qualification all tasks are graded, in a

second only practical units are graded and, in a third, assessors do not grade work as learners take different tasks depending on the grades they are aiming for.

Grading above pass potentially adds complexity to assessment decisions. Assessors who graded internal assessment talked at greater length about the complexities of making judgements. However, there was no strong sense they find assessment decisions more difficult. More often than not assessors delivering graded VTQs seemed to be as confident in their decision making as any other. Many articulated the difference between pass and merit, or merit and distinction with assurance. However, assessors delivering graded VTQs were more likely to question whether different assessors would interpret the assessment criteria in the same way.

> *What is an acceptable pass? What the difference is between that pass and a merit and the merit and distinction, and whilst [AO] do give descriptors about what that should be, you still have to have a certain amount as an assessor of discretion as to whether or not you feel that they've actually done that or not.* Level 2 Diploma (Hair and Beauty).

Although grading decisions might be more complex for assessors, interviewees delivering all 6 VTQs saw the benefits of grading internal assessments. They believed the use of grades above a pass stretches and motivates learners and rewards them for their hard work. It also gives employers more detail on the competency of the learner. Because of this, at least 3 centres in our sample had introduced *their own* internal grades above pass in internal assessment.

> *Some of them you can go the extra mile and you still get a pass and some just scrape through and they've still got a pass and it doesn't seem fair. So we thought if we issued students with a college certificate saying actually yes you passed at [AO] NVQ level but we feel you are a merit grade student or a distinction grade student then it gives them a little bit of extra value to take out to employers.* Level 2 Diploma (Hair and Beauty).

3.4.2 AO guidance on applying assessment criteria

Assessors had varying experiences of the nature or quality of AO guidance to support the application of the assessment criteria. Some assessors had accessed useful written guidance or verbal support. Many others identified a lack of guidance and exemplars of the standard, particularly in the early days of a qualification.

When discussing the nature of AO guidance, assessors were most likely to refer to informal, one-to-one verbal support provided by the EQA. In 9 centres, assessors contact their EQA with ad hoc questions about applying the assessment criteria. In another 8 centres assessors discussed the use of written exemplars. Some assessors found their understanding of the assessment standard had been strongly influenced by AO exemplars. Others found exemplars useful but did not completely agree with them. In these instances the assessor clearly had their own idea of the standard which did not completely align with that communicated by the AO.

> *The thing that really gave me an understanding of the standard that I was looking at were the 2 exemplars. That for me is the ultimate resource when it comes to marking.* Level 2 Award (Construction).

> *If you were starting the course today, you'd …see exemplar work of unit, a pass 2, a merit, a distinction, and you can link that across. I don't think I agree with all of it, but that's my perception and that's theirs.* Level 2 Award (Construction).

Some assessors discussed a lack of written exemplars for certain VTQs. Of the 3 centres delivering the Level 3 IT qualification, 2 told us the AO provides no exemplars for the units they deliver. A third was aware of exemplars in some units but found them unhelpful. Overall, these particular groups of assessors were the least positive about the availability of guidance. Some felt that the lack of guidance had been very problematic when they first started delivering the course.

> *The teachers do not have a clue at what level we are aiming this at. And the documentation is nonsense. It doesn't help us at all.* Level 3 suite of qualifications (IT).

Other sources of guidance provided by AOs to support assessors in assessment decisions included training courses and one to one training visits. Two Level 2 construction assessors praised a training course they attended in the early days of the qualification. Level 2 IT assessors talked about one-to-one training visits provided by the AO in how to apply the assessment criteria. Both of these forms of training were valuable to assessors.

> *We all group marked questions on that training event…that was useful because it gave us an understanding of what was expected for the marks. So I did feel like I had a good idea of what was expected for each level. But it all came from that training session. It was absolutely essential for me.* Level 2 Award (Construction).

Assessors most rely on guidance and exemplars during the early days of delivering a course. Perversely, this seemed to be the time when guidance was most lacking. Six assessors (all delivering the softly vocational qualifications) felt the guidance and exemplars had improved over the lifetime of the course but were lacking at the start.

> *They should have had more exemplar work on there. That should have all been accessible from the word go. And I never really felt like I'd got my hands on exactly what I needed sometimes.* Level 2 Award (Construction).

3.4.3 Conceptions of the assessment standard

During our interviews we noted a slight contradiction between the level of confidence assessors report in making assessment judgements and their perceptions of the clarity of the assessment criteria. Confidence in making assessment decisions does not always coincide with a belief in the clarity of the criteria. Instead, it became clear there are external factors which also shape assessor conceptions of the standard. These include assessors' experience of delivering similar qualifications in the past; experience of what is required by industry; a shared understanding developed by communities of practice; as well as informal or formal AO guidance. By drawing on these sources of information, assessors help to resolve any ambiguities in the assessment criteria and crystallise their understanding of the assessment standard.

Most assessors discussed a range of factors which influence their interpretation of the standard. The relative influences of these were rarely discussed, but apparently varied from assessor to assessor. The most common factor which influenced

assessors' interpretation of the assessment standard was the concept of an "industry standard". This was mentioned by 11 centres, all of which delivered construction or hair and beauty VTQs. Here assessors generally have recent or current experience of working in industry. They (consciously or unconsciously) bring their interpretation of industry standards to the assessment criteria. Where consciously done, this was less about resolving ambiguities in the assessment criteria, and more about ensuring learners were suitably prepared for local employment.

In the strongly vocational VTQs, assessors' application of the assessment criteria may be influenced greatly by this concept of the industry standard. One assessor suggested their concept of the assessment standard was 50% influenced by the AO requirements and 50% influenced by industry standard.

> *I feel like I know the qualification, I've worked in the industry a very long time and I know what I expect to see, what I want to see, I know what [AO name] want to see as well. So I try and work to that standard all the way through.* Level 2 Diploma (Hair and Beauty).

> *Commercial standard, but meeting the requirements of the qualification.* Level 2 Diploma (Hair and Beauty).

There was some evidence that assessors used their idea of industry standards to directly inform assessment decisions, even if this meant consciously departing from the assessment criteria. For example, 2 assessors explained how some of their learners had work placements or external jobs. As long as learners were working safely, assessors would not penalise a learner for following their employer's procedures even if it was different from the assessment criteria.

> *If you want an example of what's competency simply look at that door that's behind you there, there is an industrial standard that that door needs to be hung to. OK? Now, that's down in black and white. What will happen in the real world is that a site manager or client will be happy with how the door's been hung and the fact that it's operating but it might not meet the industrial standards…If we were going bang to the letter of the law, that wouldn't pass. But then because that's what they've been used to and some of them will have part-time jobs that gets accepted outside, that will pass.* Level 3 Diploma (Construction).

> *Because I still work in industry there'll be things that I'm thinking oh I can't possibly pass that. The water that they used to remove the face mask was dirty, and even though that's not a performance criteria that they use clean water, obviously that's what's expected. So then it would be I can't pass you because the water was dirty that you were using to remove the mask with.* Level 2 Diploma (Hair and Beauty).

Few assessors gave a sense of the degree to which the assessment criteria varied from industry standard. Where they did, differences between industry and qualification standard appeared minor. In some cases the AO requirements were perceived to be slightly more stringent, in others industry requirements were stricter.

Although most assessors who discussed the concept of industry standard delivered strongly vocational qualifications, they also included 3 Level 2 construction assessors. Here assessors used some idea of acceptability to industry in their

decisions, however, they agreed any concept of industry standard must be tempered by realism about the age and experience of the learners.

> *This is a 15-year-old boy that's made a bit of furniture for the first time in his life, you know, we can't take shop quality as the standard for marking. We might tell them that's what we're aiming for, but if there's a little mark or it's slightly off-centre or it's not quite square, you know, we have to remember who's building it here. And that can be an area for a bit of diversity amongst markers.* Level 2 Award (Construction).

In addition to industry standards, there are other factors which shape conceptions of the assessment standard. These usually came into play where assessors found criteria to be ambiguous or guidance to be insufficient. In 8 interviews assessors discussed how their understanding of the assessment standard was influenced by their experience of delivering qualifications in the past. In most cases, assessors talked generally about how delivering similar courses had shaped their interpretation of assessment criteria. As centres change the VTQs they offer over time, the lines between different qualifications can start to blur.

> *When they first introduced this new course, it was quite difficult, wasn't it, making that judgement and thinking in our first moderation is this going to get through?… And we went in the end with looking at the previous course, the standard that we would have expected.* Level 3 suite of qualifications (IT).

In other interviews with construction assessors, interviewees referenced a specific VTQ they were influenced by. In more than one centre assessors 'borrowed' criteria or tolerances from these qualifications.

> *I think that it was better what we got from [previous AO], you know, and basically that's what we are using still what were the [previous AO] criteria and drawings and everything like that.* Level 3 Diploma (Construction).

Finally, 8 centres discussed the influence of a community of practice in building a shared understanding of assessment criteria. All except one delivered softly vocational qualifications. Communities of practice took the form of collaborations of centres linked by geography or assessor relationships, or online communities of assessors. In the latter, centres may informally share learner work to seek advice on specific assessment decisions or to contribute to a shared understanding of the assessment standard. Again, this seemed particularly important for assessors in the early days of a specification where AO guidance was perceived to be more limited.

> *You're getting a lot of sharing of work outside of, how legitimate that is, I don't know. Obviously the exam board doesn't like that. But then you're learning it is about learning best practice. And what is a distinction level and what is not. So if the exam board aren't giving that guidance, we're having to get it from somewhere.* Level 3 suite of qualifications (IT).

3.5 Area 4: the quality assurance process

All the VTQs in our sample incorporated various layers of quality assurance and quality control. These aim to assure the accuracy and reliability of centre-based assessment judgements in internally assessed units. This section explores the nature and focus of quality assurance systems put in place by centres and AOs.

3.5.1 Internal quality assurance

Within a centre there are 2 main mechanisms for ensuring quality in assessment decisions. The first is standardisation. This is an 'upstream' control which builds quality through training assessors in the correct application of assessment criteria. The second is a downstream check of assessment decisions after they have been made. In this report we refer to the latter as Internal Quality Assurance (IQA). IQA and standardisation are hard to disentangle and were discussed interchangeably. In some cases, there appeared to be a conflation between these processes.

3.5.1.1 Standardisation

Where multiple assessors deliver a qualification, centres carry out standardisation to ensure they apply assessment criteria consistently. We asked interviewees to describe standardisation arrangements. Responses were limited, with many focusing instead on the IQA of assessment judgements. Certainly, there was evidence of an overlap between standardisation and IQA in some centres. At one extreme, centres had distinct standardisation and IQA processes. At the other, centres used IQA as the sole mechanism for standardising assessment decisions. Standardisation as it is operationalised here is closer to a quality control; a post hoc check of judgements.

Ten centres provided some detail of their approach to standardisation. Seven have regular, formal standardisation meetings. In one of the hair and beauty VTQs assessors placed a particularly strong emphasis on standardisation. Four of the 5 centres delivering this qualification had regular standardisation meetings. These cover the assessment criteria and the level of support assessors can give to leaners during tasks. In these centres, assessors recognised the importance of this standardisation in promoting accurate and consistent judgements.

> *We do a lot of standardisation activity. We have a Wednesday afternoon where we don't have any lessons timetabled across college, and we do have sort of whole staff development days as well where they do get together and they will standardise. So they'll do the usual kind of demonstrate a cut or show a video or whatever and actually say right what do you think this would be?* Level 2 Diploma (Hair and Beauty).

Not all centres have standardisation meetings. At 8 centres (delivering 5 of the VTQs), IQA of learner work *was* the standardisation process. There was no formal standardisation before assessment decisions were made, instead it was subsumed into the centre's quality control mechanisms. All of the Level 2 IT assessors described "cross moderation meetings" when asked about standardisation. Here, assessors look at recently graded work and reviewed one another's decisions. This took place *after* assessment decisions had been made. The discussion at these meetings resulted in assessors gathering further evidence of learner competence or

submitting work to the AO. There was a similar picture in the second hair and beauty course. Here, there is a clear conflation of standardisation and IQA mechanisms.

In addition to (or in place of) internal standardisation, assessors in 7 centres discussed how they carry out a kind of informal standardisation with other centres. By sharing examples of learner work they "benchmark" with other centres to ensure they are assessing to a common standard. These were rather informal and impromptu interactions and tended to take place more in softly vocational courses.

3.5.1.2 IQA of assessment judgements

Although not all centres had a distinct standardisation process, all except one had some check of assessment decisions. We refer to this as IQA. Assessors also described this as internal verification (IV), moderation or cross-moderation. Those delivering strongly vocational courses talked in terms of 'verification' whereas in school settings, they talked about 'moderation'. This different terminology highlights again the variation in the sector. It may also reflect different focuses of quality assurance. IQA processes varied across our sample often along qualification lines. In strongly vocational qualifications IQA seemed highly formalised. In softly vocational qualifications, IQA was more informal.

The resourcing of the IQA role also varied. IQA was performed by subject specialists within a centre; by sector (but not subject) specialists within a centre; by subject specialists from other centres or by independent consultants. IQAs often teach and assess the same courses they quality assure. Most assessors we interviewed also acted as IQAs, and so we draw no artificial distinction between assessors and IQAs in this section. In order to act as an IQA in the strongly vocational qualifications in our sample, assessors must hold a Level 4 qualification. There appears no such requirement in softly vocational qualifications in our sample.

Hair and beauty assessors described their IQA in the greatest detail. Processes here were clearly structured and well established. IQA happened on at least 3 occasions during the year. IQAs review a sample of written and practical work, (including observing live practical tasks). Usually samples were structured to cover all assessors, all units and all learners, although this was not always the case.

Where assessors discussed sample sizes, centres were fairly consistent in their approach with between 20 and 30% of learners reportedly having their work sampled. That said, sampling was often risk based. In 6 interviews, assessors explained that new assessors have up to 100% of their assessments sampled.

> *We've got horizontal and vertical sampling. We look at different sites and different assessors and risk assess the level of experience of assessors. Whether it's a new course or a new assessor… Last year we did Level 3 barbering for the first time. So that was sampled as at 100%.* Level 2 Diploma (Hair and Beauty).

In all but one centre delivering hair and beauty VTQs, IQA was carried out within the teams delivering the course. Within these, it was common for all assessors to be qualified IQAs. Depending on the cohort size, IQAs may or may not have taught the learner whose work they check. In another case, merged college status meant IQA was carried out between campuses allowing greater independence of judgements.

> *It's become slightly easier now we have got these 2 campuses both doing the qualifications, you can actually get people to come across*

and verify and act slightly more externally than they would be in a staffroom where you've got four or five people who work together all of the time. They're friends, they're mates. That's when you've got to really ensure that standards are being met. Level 2 Diploma (Hair and Beauty).

One small private training provider delivering a hair and beauty qualification took a different approach to IQA. At the time of the interview there was no qualified IQA in the centre. Instead, this centre used an experienced external IQA who worked as an assessor in a local FE college. Again, this IQA reviewed a sample of learner work across all units on an ongoing basis throughout the year.

We have less detail about IQA in construction VTQs, although it appeared less tightly defined than in hair and beauty. There was also greater variation in practice between centres. In 2 centres delivering the Level 3 course, IQA focused on portfolios, with reportedly no live observations of practical tasks. In the third centre regular observations of practical assessments took place. Assessors delivering the same course did not IQA each other's work and only cross-department IQA was allowed. Assessors felt this approach allowed for more independent judgements.

Other colleges like you to do it in-house, don't they, in your department, which is not always ideal. Because like last year we were all teaching more or less the same groups on different days weren't we? Level 3 Diploma (Construction).

Assessors delivering the Level 2 construction course used a different model of IQA. Again, it seemed to focus on checking learners' written work and portfolios, with no live practical observations. In 2 centres, IQA took place across centres, with assessors comparing examples of assessed work with neighbouring colleges. In the third centre, the assessor said that there was no formal IQA process.

They do like it to be moderated internally. I was the only one teaching it. So what I did was I moderated with the teacher from the other school. They were quite keen to emphasise the fact that this needed to be cross-moderated. Level 2 Award (Construction).

In the IT courses, IQA arrangements were less formal. Again, different models were used in different centres. In the Level 2 course assessors IQA each other's work via the "cross-moderation meetings" described previously. The IQA process in the Level 3 course depended on the number of assessors on the course. Where one assessor delivered the course, IQA was carried out by others in the department. In the third centre, the 2 assessors delivering the course moderated each other's work.

In their description of the IQA process, centres revealed different attitudes towards its purpose and focus. In most cases, IQA focused (at least in part) on ensuring that assessors had reached the correct assessment decision. In 7 centres (6 of which delivered softly vocational qualifications), moderation of assessment decisions seemed to be the sole focus of IQA. In another 7 centres, assessors described IQA as having a dual purpose. It involved checking assessment decisions, whilst also checking processes had been correctly followed. This might include checking all forms had been signed, all assessments had been completed and learners had been given appropriate feedback. Six of these centres delivered hair and beauty VTQs.

They go in and they verify the standards of practical assessment and making sure that the assessment process has been adhered to.

> *They also look at a range of theoretical assignments and make sure that those have been marked and any feedback is there and credible.* Level 2 Diploma (Hair and Beauty).

Assessors in another 3 centres believed the sole aim of IQA was to check correct process had been followed. Checking assessment decisions was not a key part of this process. These assessors delivered 3 different VTQs. This suggests that whilst the focus of IQA did vary slightly on qualification lines, there was some divergence of opinion within the same VTQs as to what IQA is intended to achieve.

> *When you're an internal verifier, you don't necessarily have to be an expert in that field. It helps, but you're looking at the process. You're making sure the things are ticked.* Level 2 Award (Construction).

Many assessors reflected on the effectiveness of their centre's IQA arrangements. In 11 centres, assessors believed their IQA processes to be "robust" or "rigorous". This was particularly true in hair and beauty where processes were most formalised. Assessors gave a variety of explanations as to why they believed their IQA to be robust. Five referred to the sheer amount of IQA they carry out.

> *And we do a lot of IQA. We do more than lots of other colleges for example, so I think we would see every student we would see 3 times… So that's for every student and then for one in 10 we look at their whole, everything they've done. Which is more than I think the awarding body asks for, they usually ask for a percentage, but it's a bit belt and braces really.* Level 3 Diploma (Construction).

Four centres believed the independence of their IQA arrangements added rigour to the process. In these centres, assessors do not IQA the work of learners they teach. One of these centres had assessors spread over 2 campuses. Assessors IQA the judgements of assessors based at the other site. They believe this allowed more objective checks to be made. They suggested it can be difficult for assessors who work closely together to critique their peers' work effectively.

> *That particular process is one of the most difficult to do within a college. Because what you're actually saying is to a colleague that you need to improve this, this and this, you know, your grading needs to improve… Asking people to do that, especially when you've got colleagues of the same sort of demographic if you like, the same amount of experience and one of them isn't maintaining the standard, that's quite a difficult thing to do.* Level 2 Diploma (Hair and Beauty).

In contrast, 2 groups of hair and beauty assessors told us unprompted they were very happy to challenge one another's assessment judgements.

> *When you've got your IV hat on, it doesn't matter if you're friends or whatever, you would then have to say I feel that you're not doing this or that.* Level 2 Diploma (Hair and Beauty).

There was some tension evident between the independence of the IQA and their level of relevant knowledge. In some centres, IQA was carried out by assessors in the department who are not industry experts. Two assessors saw this as a strength of their arrangements, suggesting that this can provide a different perspective. Not

everyone felt this way, however. One construction assessor believed an industry expert would be better placed to decide whether assessment criteria had been met.

> *We're not looking at rocket science here at the moment; we're only looking at level 2 qualifications. And I'm sure if I gave you the stuff, you would probably be 60% confident that what that kid has put down is sufficed for the criteria. However, it would be much better if I had somebody from a college or school that was delivering the course who has background in construction as an internal verifier.*
> Level 2 Award (Construction).

3.5.2 External quality assurance

In addition to IQA, all centres are subject to external checks by their AO. We refer to this as External Quality Assurance (EQA)[4]. Again, the use of this term does not reflect any common language amongst assessors. In our interviews assessors delivering softly vocational qualifications talked in terms of "external moderation" whilst those delivering strongly vocational courses mentioned "external verification". Again, this may well reflect different purposes of EQA activity. Assessor perceptions of the aim of EQA generally mirrored that of their IQA. So if the focus of IQA was the moderation of assessment decisions, assessors believed EQA has the same focus.

We examine the different approaches to EQA in the following section. It is important to note that this information is based on interviewees' perceptions and experiences of the EQA process. We should not assume it is a fully accurate representation of each awarding organisation's EQA process. That said, assessors delivering the same qualifications did give fairly consistent descriptions of EQA.

We found variation in the EQA of centre assessment decisions between strongly and softly vocational qualifications. In many strongly vocational courses, centres claim certificates for learners without prior moderation of assessment decisions having taken place. To do this they must be granted Direct Claims Status (DCS). Of the 11 centres delivering strongly vocational courses, 10 had DCS. In the softly vocational qualifications all cohorts are subject to moderation prior to certificates being issued. This inevitably leads to differences in the design and focus of EQA activity.

EQA is just one control an AO has over centre delivery of internal assessment in VTQs. As discussed in previous sections, they also retain varying levels of control over task setting, task taking or IQA. We should consider EQA in the context of these wider controls. For example, the greatest flexibility in task setting and taking was experienced in IT qualifications. These also had the most informal IQA processes. However, these VTQs appear to have the most intensive EQA in terms of frequency and coverage of sampling. These VTQs also require assessment decisions to be signed off by the AO before certification for a cohort can be claimed.

In contrast, hair and beauty VTQs allow little flexibility in task setting and taking. IQA arrangements are highly formalised. Whilst EQA is intensive, it involves smaller samples of work. In these VTQs, there is more emphasis on assuring centres go through the correct processes to make and quality assure their decisions.

---

[4] We should note that the external quality assurance (EQA) by AOs of centre-based assessment delivery referred to here has no connection with the EQA of apprenticeship end-point assessments.

3.5.2.1 EQA of hair and beauty VTQs

In hair and beauty qualifications, EQA involved prearranged centre visits once or twice a year. The focus of EQA appeared to be broad. EQAs review the process of delivering assessments and the quality of assessment judgements. This is consistent with the focus of IQA. Hair and beauty assessors felt EQAs have a particularly strong focus on checking the effectiveness of IQA. They check sampling plans, meeting minutes and standardisation record and observe the IQA of a live practical task.

> *She will come in normally on her first visit and check all the processes are in place, all the students are registered, CPD has been recorded, minutes of meetings, standardisation. All the processes are normally done on her first visit. And then her second visit, we have all of the students in performing a range of services.* Level 2 Diploma (Hair and Beauty).

The second strand of EQA was checking assessment decisions. AOs appeared to have slightly different approaches to this. According to assessors, the EQAs from one AO spend a full day at a centre watching a wide range of live practical summative tasks and checking written work. EQAs pre-select a sample of written work to review but practice varies in terms of sampling practical work. One centre suggested that it was difficult to completely pre-select the treatments each learner might demonstrate as "it depends what walks through the door" on the day. Another described how the AO used to specify learners they wished to see demonstrate practical skills, but this did not always allow them to see a wide range of treatments. As a result they were given more flexibility to select learners in order to demonstrate a range of treatments.

> *So she'll see a facial, a wax, a manicure, a pedicure, an eye treatment, a makeup… then we just allocate the students, the ones that haven't completed that unit yet will be doing the client. We don't give her students that have already passed that unit, because that's a bit pointless.* Level 2 Diploma (Hair and Beauty).

At the other AO, EQAs pre-select the portfolios of learners they wish to sample. They appear to sample fewer practical assessments than the first AO, however, they reportedly retain higher control over sample selection. In addition to selecting the learners they wish to sample, EQAs also select assessors and IQAs to observe.

> *She can look at whoever she wants on the day. We all assess and IV on the day, but she will usually home in on one person because with so many of us it would take her a week not a day to see everyone.* Level 2 Diploma (Hair and Beauty).

In both hair and beauty VTQs, there is a dual focus on reviewing the assessment process and the quality of assessment decisions. The exact emphasis of this varied by AO, however. Whilst one AO's EQAs sample summative assessments only, at the second AO EQAs might sample formative or summative assessments. This latter AO appears to place more weight on reviewing the *process* of assessment during EQA. Various assessors delivering this particular course explained the purpose of EQA is not to quality assure summative assessments as such, but instead to ensure that a centre is capable of making and assuring accurate and reliable judgements.

> *Usually if we can she'll look at least one or two of those that are actually on summative assessment, or formative assessment. It doesn't really matter. It's just the assessment process she's looking at. And then she feeds back to us. She listens to our feedback to the students and she listens to the IQA's feedback to the assessor.*
> Level 2 Diploma (Hair and Beauty).

3.5.2.1 EQA of IT VTQs

EQA of the IT qualifications had a different focus. Here the primary aim of EQA was to assure the quality of assessment decisions for a cohort prior to certification. Less attention was seemingly paid to the assessment process. Whilst the organisation of EQA varied between the courses, control over selecting samples is high. Sampling ratios were also high, with up to 100% of learners' work sampled.

In the Level 3 IT qualification, centres received 2 visits a year. As well as checking a sample of assessment decisions, EQAs may also quality check assignment briefs and the level of feedback given to learners. Samples of learners are selected by the EQA prior to the visits, but centres are still required to make all learners' work available to EQAs. Where specified, sample sizes reviewed by EQAs ranged from 50% to 100% of learners on each visit.

> *With [AO] they look at every student. So it's not just your sample. So every piece of work for every student is looked at.* Level 3 suite of qualifications (IT).

EQA in the Level 2 course also focused on moderation of assessment decisions, however, practice appeared to be more varied. One centre told us they received regular centre visits (announced and unannounced), a second had only experienced remote moderation and a third had experienced both. In one centre, all learner work was moderated, in another the rate of sampling ranged between 50% and 100%.

> *On occasions we would get a sample. So if we submitted 30 students, we would be asked for 15. On other occasions we got asked for all 30.* Level 2 Certificate (IT).

In both courses, the EQA process was "fluid". Centres submit work or arrange an EQA visit on completion of their units. EQAs review assessments and may ask for more evidence of learner competence. This is resubmitted to the EQAs to review.

3.5.2.2 EQA of construction VTQs

We have a less clear picture of EQA in construction qualifications. To a degree this is due to differences between the 2 VTQs. One consistent theme across both VTQs was the fact that EQAs rarely check live practical tasks in the centres in our sample. Construction assessors were also the least positive about the rigour of EQA.

All of the colleges delivering the Level 3 Diploma in construction had been granted DCS. EQA was similar to hair and beauty VTQs, with a dual focus on assessment decisions and centre processes. EQA took the form of a centre visit conducted once or twice each year. EQAs pre-select samples although sampling ratios were low.

> *He'll pick at least one person from each level and qualification… So he will look at their training book and assessment book. He also will go and speak with students. He may observe an assessment. He*

may want to observe an assessor assessing and an IQA assessing the assessor. Level 3 Diploma (Construction).

In 2 centres, assessors told us that EQAs occasionally sample live practical tasks. In the third, practical assessments had never been sampled for logistical reasons.

The external verifier does request it, but the trouble is that when he comes in for some reason he tends to come on a day when that particular one he wants to speak to isn't actually in. Level 3 Diploma (Construction).

In the Level 2 course EQA appeared quite different in both process and focus. As with the IT qualifications, assessment decisions for each cohort are checked prior to certification. As such the focus of EQA appears to be the quality of assessment decisions. Assessors send written work and photographs of practical work to moderators. In 2 centres, assessors had to send all learner work to a moderator. In a third, the AO requested a specified sample. Of the 6 VTQs in our sample, these assessors appear to receive the least feedback from the EQA.

It all gets sent to a moderator. I mark it. The folders get bundled up and sent off to a moderator. They check that it's legit and the last 2 years it has been. They've not sent anything back. Level 2 Award (Construction).

3.5.2.3 EQA support role

When we asked assessors to describe EQA, we anticipated responses would focus on its monitoring function. In fact, assessors perceived EQA to have another vital role. In 14 of the 20 centres, assessors discussed the support they receive as part of EQA. Seven assessors believed centre support was a central aim of the EQA process. Some felt the EQA was less of a check of centre performance and more of a form of centre reassurance and assistance.

They have become better at giving advice. It used to be almost like a punishment and they would find something, but now they do [give advice]. They're supposed to be like a critical friend and I think they are generally more like that. Level 3 Diploma (Construction).

I don't panic now when she comes. I know her now, so I know she's there to support and she's been a really good support, to be honest. Level 2 Diploma (Hair and Beauty).

Assessors found the support provided by EQAs during centre visits or remote moderation to be invaluable. They used EQA as an opportunity to ask questions about assessment criteria and qualification delivery. EQAs also made suggestions about how to improve delivery of courses and shared good practice in other centres.

Outside of the formal EQA process, assessors in 11 centres talked about the support they receive from EQAs between visits. These assessors seek advice from their EQA via phone or email, usually on task setting or interpretation of assessment criteria. In some cases, a centre's understanding of the standard required had been influenced by an informal dialogue with their EQA spanning a number of years.

Some assessors described a limit on the length of time an EQA was allocated to a centre (between 3 and 5 years). Often EQAs had been with centres for a period of time at the upper end of this limit. This allowed assessors to build up a "good

working relationship" with their EQA. Two had been given more freedom to select samples of learner work during the EQA process due to this relationship of trust.

> *Because she's been with us a long time she'll say I want to see this IV, this assessor…or you choose what you want me to see.* Level 2 Diploma (Hair and Beauty).

Although some interviewees suggested centre support is a key aim of the EQA's role, we have no evidence from AOs to confirm this. Support and monitoring are clearly very different functions, and ones which it may be difficult to reconcile in a single EQA activity. This raises a question about whether there is any tension between the 'support' and 'custodian of standards' roles reportedly held by the EQA.

> *There can be some conflict there between, you're hoping for advice and they're considering that what you're doing is sanctionable.* Level 3 Diploma (Construction).

3.5.2.4 Effectiveness of EQA

Overwhelmingly the assessors we interviewed were positive about the effectiveness of the EQA process. Fifteen of 19 who expressed a view believed it to be a robust and effective process. Most told us they trusted the process to detect inconsistencies in standards between centres. Assessors believed EQA involved high levels of scrutiny; often using variations on the phrases "with a fine-toothed comb" and "nowhere to hide". Some believed EQAs would readily sanction centres if required. Interestingly, these perceptions stand in contrast to the limited personal experience our interviewees had of changes to assessment decisions or sanctions.

> *They scrutinise everything. So if there's something to be found, especially with the EQAs that we've had, if there's something to be found they will find it.* Level 2 Diploma (Hair and Beauty).

Another 4 assessors gave more nuanced responses. On balance these concluded they trusted EQA, but less emphatically so. Two assessors told us no quality assurance process was perfect, but EQA was probably as good as it could be.

> *Without being here every minute of every day it's very difficult to get away from that. So I think with the resource they have they do as well as they can do.* Level 2 Diploma (Hair and Beauty).

Two construction assessors talked about potential points of vulnerability in EQA. Both questioned the ability of the EQA to authenticate assessments, believing centres could fake evidence of practical work. This would be difficult for the EQA to identify unless they observed live practical assessments. This was not thought to be a problem specific to these particular VTQs, however. One also perceived that EQAs would tend to give learners the benefit of the doubt in any borderline decisions.

> *If people wanted to cheat then it would be easy enough to do that if you get away with it. But it was the same with the other course I taught and it's the same with the engineering course I'm teaching now: it's just part of the system. I don't see how this one would be any easier to get away with it. It feels robust enough.* Level 2 Award (Construction).

> *If it's a grey area, you always err on the side of positivity… And I believe that the moderation is probably the same as that. They're*

more inclined to go with awarding the grade rather than not. Level 2 Award (Construction).

Some other areas for improvement were identified in even the most positive interviews on EQA. In 6 centres delivering hair and beauty VTQs, assessors had experienced inconsistencies in the EQA process, either between different EQAs or over time. They suggested different EQAs can have different approaches to the process and even slightly different interpretations of the specification. Although these differences were minor, some assessors called for better standardisation of EQAs.

> *I do think the standardisation of the EV process needs better scrutiny to be honest with you; only because we've experienced a few different verifiers, EV'ers and sometimes have been told conflicting things.* Level 2 Diploma (Hair and Beauty).

3.5.3 Adjustments to assessment decisions

The assessors we interviewed trusted the EQA process to maintain standards across centres. It was therefore interesting to note how rarely adjustments to decisions had been experienced in their centres. Where discussed, 'adjustments' usually took the form of AOs asking centres to provide further evidence of learner competence. No one we interviewed discussed any cohort-level changes to grades awarded. Assessors in 15 centres discussed whether they had ever had their assessment decisions adjusted or overturned. Nine told us they had no experience of changes to assessment decisions in these specific VTQs.

> *She could actually overturn something if one of us didn't make the right decision. But I don't think that's ever happened. I mean all the times I've been here and at my last college we had [AO name] and they never really overturned anything we've decided on.* Level 2 Diploma (Hair and Beauty).

Assessors in 5 centres had some experience of EQAs challenging assessment decisions. Four of these delivered IT qualifications. Here, the EQA did not adjust decisions but asked assessors to provide further evidence of learner competence before certificates could be claimed.

> *They can mark it down to say you've been too lenient in your marking so we don't agree with that. But most of the time they'll just say, if it does need a bit of tweaking, go away and do this and we'll check it next time.* Level 3 suite of qualifications (IT).

One group of construction assessors described an incident where a learner had their certificate revoked by the AO. This was because the assessors had not recorded prior learning completed by the learner at a different centre. The learner had to retake a 12 hour assessment and subsequently had their certificate reissued.

Finally, 2 centres suggested that whilst EQAs may not overturn assessment decisions, they might recommend that future decisions are made slightly differently.

> *I've never had experience of them coming and saying I completely disagree that that is a pass or something like that. But they may say what I'd like to see you do in the future is make sure that you include this in the assessment.* Level 3 Diploma (Construction).

As well as changes to assessment decisions, AOs may also give formal actions to centres to improve delivery or sanction them in some way. Again, this was rare at the centres we visited. Two assessors discussed receiving a sanction as a result of issues identified during EQA. In both cases, the AO removed the centres' DCS as a result – a sanction described as "*debilitating*".

3.5.3.1 Variation in standards between centres

Just over half of centres perceived there to be some variation in standards between centres delivering the same VTQs. This might relate to differences in assessment standards, quality of teaching or task demand. This perceived difference did not necessarily reflect a belief that 'other' centres set low standards. Often there was a general feeling that minimum standards were secure, but some centres aspired to deliver higher standards in order to best equip their learners to find employment.

Most perceptions of standards in other centres were based on second hand information. Some assessors had been told by an EQA their standards were superior to other centres, others had been told by learners about poor teaching elsewhere. Others suspected disparities between colleges due to the standard of learners who had previously been at another college.

> *When we noticed the quality is when we've had students that have done a previous level at another college, and come onto us. And you think how did you pass?* Level 2 Diploma (Hair and Beauty).

In contrast, assessors in 7 centres perceived standards to be broadly the same across different centres delivering the same VTQs.

> *Well we just assume that it's done in the same way that that process is being followed. It's quite clear. We are being told by our awarding body how we do things and we assume that it is, that other centres are doing it in the same way.* Level 2 Diploma (Hair and Beauty).

It was interesting to note that most centres seemed unconcerned about any differences in standards, even where they believed these existed. Five groups of assessors told us they were indifferent to standards in other centres and simply focused on elevating their own standards.

> *I hope that people all teach well and obviously that the standards are met. But I just worry about my students and think as long as I know that mine are good and our standards are good I'm not too worried about other people.* Level 2 Diploma (Hair and Beauty).

When discussing variation in standards between centres, 11 assessors referred to cheating or gaming. Most referred to cheating in a fairly abstract way, musing on the potential for malpractice in these VTQs. Three assessors were aware of individual incidents of malpractice in centres, experienced either first or second hand. One group of hair and beauty assessors felt a real sense of injustice about this.

> *You feel a bit cheated that they're doing that, that that sort of thing is happening. It makes me feel well hang on, this is the way it should be done… It's making a mockery then of the qualification, isn't it?* Level 2 Diploma (Hair and Beauty).

Not all responses focused on explicit cheating. Two assessors referred to a subtler form of gaming the system. One construction assessor told us they always erred on

the side of positivity when making assessment decisions, leaving it to the EQA to adjust grades downwards if required. In a second case, a group of IT assessors told us they used their experience to "play the system" by selecting the least demanding route through for learners, rather than the route which would give the learner the most relevant knowledge and skills.

> *You want to do everything you can to get the kids through, to let them achieve. And if it's a grey area, you always err on the side of positivity. Because there's a lot riding on it. It's the kid's grade. It's your performance management. It's the school's resources.* Level 2 Award (Construction).

When discussing variation in standards in centres, many assessors reflected naturally on some of the causes of this. In 9 interviews, assessors discussed the enormous pressure they are under to achieve good results. This can act as a perverse incentive for malpractice or gaming. Most told us this pressure comes from centre management.

Three assessors talked specifically about the role of centre funding in this. One assessor noted that where learners are paying full-costs for a course, assessors may feel a responsibility to make sure they get their certificate. Another felt some centres place learners onto higher levels of a qualification than they should as they attract more funding for these. Finally, one group of assessors discussed how if the college lost funding (through learners failing courses), their jobs may be at risk.

> *And we know that if we don't get figures for next year then possibly [assessor name] is out of a job… There's a lot of pressures. The college could save £25,000 and one of us will be going.* Level 3 Diploma (Construction).

Another 4 interviewees felt many assessors had an inherent desire for learners to achieve. They may struggle to manage their dual teacher-assessor role and find maintaining independence of judgements more difficult where they have good relationships with learners.

> *There are always borderlines. We never want anyone to fail; we always want our students to get a grade if possible.* Level 2 Certificate (IT).

Although there are pressures on assessors, most interviewees highlighted a range of protective factors which encourage them to uphold assessment standards. They include drivers which act as both 'carrot' and 'stick'. The top protective factor cited by assessors acts as a 'virtuous incentive' whereby good consequences follow from standards being upheld. In this case, the desire to uphold centre reputation incentivised assessors to promote high standards. This was mentioned by assessors in 11 centres, most of which delivered strongly vocational courses. Assessors spoke at length about the need to maintain their reputation locally by producing highly competent learners. This went beyond simply meeting assessment standards, and resulted in assessors delivering additional learning which they felt would give learners extra desirability to employers. They explained that any fall in centre reputation could have implications for learner numbers and centre resources.

> *We're not going to turn them out of here without having the skills that they need to do the job. They're going out as our representatives and so we need companies, our local companies to*

> *know that if they have an apprentice or a trainee from us then they can rely on that standard.* Level 3 Diploma (Construction).

> *We are very tight here, because we have a good relationship with our employers around here for hair and beauty… I can't go to any beauty salon within about a 12-mile radius because it's got my students working there, which is nice. But we need to be sending them out to a good standard; otherwise it's no good for us*. Level 2 Diploma (Hair and Beauty).

A similarly positive influence on behaviour was the professional integrity of assessors. In 9 centres, assessors told us they were passionate about their teaching role and had a strong professional pride in ensuring that their learners have the highest levels of competence possible.

> *It's all down to our integrity at the end of the day. You could be as crooked as you like, we could sign all the students off now if we really wanted to. All the way up to level 3 probably. And then do a runner. We could but we don't. We try to be as honest. It's down to personal integrity.* Level 3 Diploma (Construction).

The other protective factors mentioned by assessors related to the likelihood (and consequences) of being caught cheating. Assessors in 10 centres believed the scrutiny of the EQA process deters centres from deliberate cheating. Six also mentioned the threat of sanctions as a deterrent, particularly the removal of DCS.

> *They will sanction centres readily. It's a real risk for centres because it's their reputation and a nightmare if you've got a load of students that have completed the year and you can't claim for them.* FE College. Level 3 Diploma (Construction).

> *The threat of malpractice or having something disallowed sharpens the mind in this area…We've seen one of our colleagues actually charged with malpractice and had that upheld in the past, OK, and I'd never want to go there again. Never want to see that. It's horrible.* Level 2 Certificate (IT).

3.6 Area 5: External assessment

The focus of this study was the internal assessment of VTQs. Despite this, all of the assessors focused some of their responses on the externally assessed elements of the VTQs they deliver. In this section we provide a brief summary of this discussion.

All the VTQs in our sample include internal and external assessment. The weighting of external assessment in these ranged from approximately 33% to 70%.

The way in which external assessment contributes to the overall qualification grade depends on the VTQ in question. In some qualifications, the grades available for external exams mirror those available in the internally assessed units. In others, a learner's grade is derived entirely from their performance in the online tests.

The VTQs in our sample are characterised by a high level of flexibility in the delivery of internal assessment. External assessments also allowed for flexibility, although to a lesser degree. In the strongly vocational qualifications, external exams are on demand, with learners entered for these at any point where assessors deem them ready. There is also some flexibility in the format of exams in hair and beauty VTQs. In one such course assessors can decide whether knowledge is assessed through external exams at all. Although the centres we spoke to preferred to use online tests, they could set written assignments instead for learners who found exams daunting.

When discussing external exams, assessors were most likely to focus on their demand or difficulty. Generally they believed exams to be demanding, particularly in the softly vocational qualifications. There was no sense this was excessive; most seemed to believe exams were "pitched at the right level". IT assessors were most likely to discuss the high level of demand in the exams, particularly in the Level 3 qualification where one group of assessors questioned whether the demand in these exams may be excessive.

In contrast with the VTQs above, 3 assessors delivering strongly vocational courses felt the level of demand of the external exams was low compared to the internally assessed elements. They discussed highly predictable or "easy" questions, the multiple choice format and the number of retakes available.

> *Sometimes we've despaired with the exam, the tests that [AO name] set. I think some of the wording of the tests, and it tends to be the, we're probably shooting ourselves in the foot saying this, but it's been the same one year on year on year on…You can teach to an exam can't you if you had to.* Level 2 Diploma (Hair and Beauty).

Overall, assessors were more likely to believe external assessment was more demanding than internal assessment, however. Five assessors delivering softly vocational qualifications explicitly stated this was the case.

> *Dare I say it's easier to get a distinction in the coursework than it is in the exams?* Level 3 suite of qualifications (IT).

In 6 interviews, assessors discussed the validity of online tests. Three assessors delivering softly vocational qualifications praised the validity of exams. This included 2 IT assessors who were impressed with the up to date content of the tests. Three groups of assessors delivering each of the strongly vocational qualifications were less positive about test validity. Two groups criticised the wording of exams. They felt this was a barrier to learners demonstrating technical knowledge.

The questions on the tests are pitched at a level that's very difficult for some students…I'm not talking about technical words now because technical words, we teach them. I'm talking about the actual language… I think would disadvantage somebody who is actually quite a good hairdresser. Level 2 Diploma (hair and beauty).

One group delivering the Level 3 Diploma in construction were particularly critical of test validity. As well as issues with test wording, they criticised the testing of content they perceived to be out of date or irrelevant to the unit as well as ambiguous questions for which more than one answer could actually be correct.

3.7 Cross cutting themes

During the course of the interviews we identified a number of common threads underpinned or helped to shape delivery of internal assessment in the 6 VTQs in our sample. We briefly explore these below.

3.7.1 Diversity in the VTQ sector

One of the most notable features of this study was the great complexity within the VTQ sector. It incorporates a vast number and range of qualifications with various purposes and end users. Although we only included 6 VTQs in our study we found huge diversity within these along almost every dimension. They were delivered in different settings by assessors with different backgrounds and experience. They had different purposes and were aimed at different cohorts. They also had different assessment requirements and grading structures, in part to cater for the specific needs of the sector. As well as the variation between VTQs, there was also variation in how different centres delivering the *same* VTQs administered, assessed and quality assured internally assessed tasks. Because of this diversity, the experiences of assessors we interviewed were often very different.

The complexity of the sector undoubtedly raises challenges for Ofqual, awarding organisations and other organisations working in the field. One may also imagine it raises challenges for assessors. At any one time, assessors might be delivering a suite of qualifications provided by different AOs, at varying Levels and in different subjects. Internalising the requirements associated with each of these VTQs could be challenging. Assessors talked very little about any such difficulties, however. They seemed either unconcerned about, or accepting of, this. That said, we observed some confusion or unfamiliarity with qualification requirements amongst a minority of assessors, and there was a sense of the lines between different VTQs starting to blur. At times this appeared to be an unintended consequence of the sheer number of qualifications delivered by assessors. In other cases, it appeared to be more of a deliberate policy, with assessors taking aspects of other qualifications they found to be useful and applying them to these specific VTQs.

> *We've started up a few new courses recently, as you'll understand. It's difficult to remember exactly, but I'm pretty sure it was clear from the board that this is how the assessment is to be undertaken. I might have glossed over it, to be honest, in the documentation.* Level 2 Award (Construction).

Another cause of diversity is the flexibility associated with the VTQs in our sample. This includes the variety of routes centres can take through a qualification and the freedom afforded to assessors in task design. Although one assessor felt that the number of routes through a VTQ could be confusing, both of these aspects of flexibility were considered strengths of the VTQs. Different optional routes allow centres to deliver units which suit their resources or assessor background, best engage their cohort or offer the most useful route through the qualification.

> *It is very confusing. There are so many routes through and for some of the routes that is actually mandatory. I think if you're doing a two or three equivalent A-levels then that becomes mandatory…There's so many, oh gosh..*. Level 3 suite of qualifications (IT).

Another potential source of complexity is the links between similar VTQs. Again, assessors only tended to refer to the useful aspects of this. In 3 different VTQs assessors explained that learners struggling to achieve a Diploma may be dropped down to a Certificate. Assessors appreciated this "safety net" which ensures that learners are still able to gain something for their work on the course.

> *We've had students who have struggled because of ability, because of other issues and we've said right, it's not fair on the student to go out of this without anything. So we've downgraded their entry from a diploma to a certificate.* Level 3 suite of qualifications (IT).

Whilst the diversity of the VTQ sector might create challenges for some, it appears to be less of an issue for individual assessors than we might suppose. With diversity comes choice and flexibility. Assessors in our sample valued this and there was little discussion about any negative aspects of the range of, and routes through, VTQs. That is not to say that assessors do not experience complexity in the system. In fact, many assessors talked at great length about the complexities of delivering VTQs. These did not appear to be a product of the intrinsic diversity of the system, however, but due to a sense of constant change and reform in the VTQ sector.

3.7.2 Frequent change and reform

Many assessors talked at length about the level of change and reform in the sector. There was a clear divide here in the views of assessors delivering softly and strongly vocational qualifications. The strongly vocational qualifications in our sample had experienced great stability in recent years. In contrast, there was a strong sense of frequent change amongst those delivering softly vocational qualifications. This might be the result of government reform, changes in the VTQs which contribute to performance tables or AO-led change to qualification content. Assessors delivering softly vocational qualifications had experienced frequent turnover of courses in their schools and regular change in the content and structure of the VTQs they deliver.

> *The courses usually only last for, in my experience, four or five years before they have to revise them.* Level 2 Award (Construction).

In 5 centres, assessors discussed the impact of changes to the VTQs which contribute to performance tables. At the time of the interviews, 2 of the softly vocational qualifications in our sample no longer contributed to performance tables. All 5 centres had decided to stop providing these VTQs as a result. There was some frustration about this, not least because these VTQs had been so well regarded by assessors.

> *I think it's you lot actually keep messing around with the qualifications we're allowed to offer. And so basically we did constructing the built environment for 2 years and then DfE decided it wouldn't count towards Progress 8 anymore. So we went to another [AO name] thing that they assured me was quite similar… Not as good. Not as much practical. The kids don't enjoy it as much.* Level 2 Award (Construction).

Changes to the qualifications which contribute to performance tables can cause disruption to centres. Running new courses has considerable resource implications,

both in terms of the cost of training and resources and in terms of the assessor time required to develop teaching and assessment materials.

> *Setting up a course is expensive and it's a lot of work. And then if you have to restart again the next year… It's really tricky. So we've got lots of equipment sitting there not being used now. Because they re-ratified it, we might switch back, but it causes us a heck of a lot of work.* Level 2 Award (Construction).

Changes to qualifications can also make it difficult for assessors to internalise qualification requirements and assessment standards. Certainly, many interviewees told us they are far less confident in applying the standard in the early days of delivering a course. They talked about learning from their mistakes and refining their understanding of criteria as they gained experience of delivering a qualification.

> *With these kind of qualifications, especially when you're doing it the first time, you never quite know if what you're doing is what the exam board expect to see.* Level 2 Certificate (IT).

As discussed previously, where there is a lack of clarity about qualification requirements or assessment criteria, assessors may draw on 'unofficial' sources of information to shape their interpretation of this. We found evidence of this throughout our interviews, particularly when it came to applying assessment criteria. In 8 interviews assessors discussed how their understanding of the assessment standard was influenced by their experience of delivering qualifications in the past. Usually assessors talked quite generally about how delivering similar qualifications had shaped their interpretation of assessment criteria. Others talked about how they directly borrowed guidance or tolerances from a previous qualification when making assessment decisions for the VTQs in our sample. In either case, there is an obvious risk that lines between different qualifications start to blur.

> *What I always keep in my mind is that shampooing is shampooing whatever you call it. Blow drying is blow drying. Setting is setting. So it doesn't really matter too much about individual [qualification] numbers.* Level 2 Diploma (Hair and Beauty).

In an apparent attempt to minimise the work associated with qualification change, a handful of centres re-used resources or processes associated with previous qualifications. This might involve revising and reusing internally assessed tasks or keeping IQA arrangements developed for other qualifications. Again, this may lead to a blurring of the lines between different VTQs.

With one or two exceptions, most of the discussion above was observed in schools delivering softly vocational qualifications. The strongly vocational qualifications in our sample had enjoyed greater stability over the years. Some assessors felt this had reinforced their familiarity with qualification requirements and assessment criteria.

> *You could get caught out if you didn't know the specification. But it's the fact that we've been with* [AO name] *for absolutely years.* Level 2 Diploma (Hair and Beauty).

Although qualification stability was useful from the perspective of assessor familiarity with AO requirements, as previously discussed some assessors felt there had been a trade-off in terms of the currency of qualification content. Some assessors

delivering strongly vocational courses believed these had not always kept pace with constantly evolving industries and included "out of date" techniques and services.

3.7.3 Devolved decision making and control

Although awarding organisations remain responsible for the standards of the qualifications delivered in centres, interviews with assessors highlight that, in practice, many decisions that underpin these standards are devolved to centres. Depending on the VTQ, assessors are trusted to make important decisions about setting summative assessment tasks, the task-taking environment, interpreting the assessment criteria and ensuring quality across assessors and cohorts. The level of oversight and control that AOs have (or that assessors believe that AOs have) over each aspect of this process appears to vary.

It is hard to draw any general conclusions about the overall level of control AOs have across the VTQs in our sample. Points of control seem to be prioritised differently by different AOs. There is no indication what model offers the most effective oversight. Broadly speaking, the nature of devolved decision making varied depending on whether VTQs were strongly or softly vocational. In the strongly vocational courses, AOs place an emphasis on maintaining control over the early part of the internal assessment process. They tightly specify the nature of practical tasks and the task taking environment, and require centres to have robust IQA arrangements. Some centres are then able to directly claim certificates for their learners. In contrast, in the softly vocational qualifications, AO controls are more focused at the end of the process. Centres have far greater freedom to develop and set summative tasks, task taking controls are often lower and IQA is less tightly specified. However, AOs check extensive samples of learner work before certificates can be claimed by the centre.

In either model, AOs place considerable trust in the knowledge and integrity of assessors. This might be viewed by some as a strength of the system. The assessors we interviewed were generally highly experienced, passionate about their job and confident in their ability to perform their role effectively. A minority of assessors acknowledged and praised this level of trust from AOs.

> *It's good that they're trusting us as the experts to actually make that judgement.* Level 2 Award (Construction).

This level of devolved decision-making inevitably introduces risks to the validity and reliability of internal assessment. A small number of centres recognised this potential for standards between centres to vary as a result of different interpretations of AO requirements or understanding of the assessment criteria. Similarly, where pressure on assessors outweighs 'virtuous' incentives, even the most experienced assessors might skew their understanding of the assessment standard.

> *What we're doing here, another centre say …could be awarding kids distinctions and we're giving them passes.* Level 2 Award (Construction).

Linked to this devolution, there is also a question about the degree to which AOs 'own' the assessment standard and the degree to which it is owned by assessors. It is clear there are other factors which shape assessors' conceptions of the standard which are external to assessment criteria. This included experience of delivering similar qualifications, experience of industry requirements and a shared conception of the assessment standard developed by communities of practice. Using these

sources of information, assessors form their own understanding of the assessment standard. This may well be subtly different from the standard communicated by the AO. In a small number of cases, assessors consciously applied a slightly different interpretation of the standard than set out in AO guidance. In this sense some assessors feel they have a certain ownership of the assessment standard.

3.7.4 Communities of practice

A theme which permeated many of our interviews was the importance of communities of practice when delivering internal assessment. In almost all centres (18 of the 20 we visited), assessors' networks were discussed in some way. Many of these centres relied strongly on communities of practice to shape their approaches to teaching and assessment. Communities of practice took a range of forms. They might be formal, established arrangements or more informal, sporadic interactions. Some are based around online forums, whilst others rely on personal links between assessors. Whatever their form, they facilitate the sharing of best practice and resources. The influence of communities of practice was particularly evident in the softly vocational qualifications. These VTQs offer more flexibility in task setting and were also less established than the strongly vocational courses in our sample.

Although some networking opportunities were facilitated by AOs, most were centre-led. Usually they take the form of informal networks between individual assessors or centres based on geographical location or on pre-existing relationships. Many assessors told us they stay in touch with former colleagues with whom they share ideas and resources.

> *FE is a very small community. People don't realise, if they've not been involved in it, they haven't realised how small it is. I've worked for a lot of colleges… You don't lose contact with the people that you've worked with before. You'll swap things around. You'll throw them something and go have a look at this and see what you think.* Level 3 Diploma (Construction).

Other assessors described communities of practice based around teacher-led online networks. This was prevalent in IT courses, where assessors shared teaching resources, assignment briefs and even anonymised pieces of learner work. Some AOs actively monitor these channels, others appeared 'unregulated'.

> *They're not set up by the exam board, they're set up by teachers, but there's different support groups and you can post pictures on there. They're really good for resource sharing and just general discussion, am I doing this right, that kind of thing.* Level 2 Award (Construction).

Although many networks were quite informal, some had become more formalised. For example, one AO requires centres to carry out moderation of their assessment decisions. As the course in question was only delivered by one assessor in each centre visited, assessors had arranged cross-moderation with other local centres.

Assessors in 7 centres explained that some networking took place through AO facilitated channels. This usually took the form of AO-led online forums. Again, this was most common in IT qualifications. Other networking opportunities occur as a result of attending AO training events and workshops.

Whatever form they take, communities of practice serve a number of useful purposes. Most assessors talked very generally about how they use networks to exchange ideas and advice and to share best practice. Others were more specific explaining how communities of practice support the sharing of teaching resources, assignment briefs or examples of student work. The networks used by these centres therefore helped to inform most aspects of their delivery of internal assessment.

Assessors in 9 centres (7 of which deliver softly vocational qualifications) used communities of practice to act as a kind of informal and impromptu standardisation process. By sharing examples of learner work they "benchmark" with other centres to ensure they are assessing to a common standard. Assessors may also use networks to gain a second opinion on specific pieces of learner work when they are unsure in their assessment decision. In this way, communities of practice help to shape assessors' conception of the assessment criteria.

> *[We are] part of ICT Teachers Facebook Group. Which sounds quite sad, but it's actually really useful…You moderate through other people as well. So I would say I can give you an example of an anonymous piece of work to someone and say would you agree that this is a higher or a secure, and get some feedback. Just not en masse but just for the odd task here and there, just to secure my knowledge from people that are more experienced of delivering it.* Level 2 Certificate (IT).

For many assessors delivering softly vocational qualifications, communities of practice seemed to have a particularly crucial role in qualification delivery. Many of these centres described quite extensive use of networks and emphasised the importance these have in developing their teaching and assessment practice.

> *In my position now, you rely on it…The networking is a big thing. I'm working with a lot of different schools.* Level 2 Certificate (IT).

Some assessors alluded to the reasons why they drew on communities of practice so extensively. Five explained that in the early days of delivering a qualification (or a specific unit), there can be a lack of clarity about how to deliver or assess internal tasks. Sharing experience and resources with colleagues was an important way to build capacity and confidence. Another 3 assessors found communities of practice to be crucial where AO guidance is lacking. In these cases, assignment briefs and examples of learner work informally replace AO guidance and exemplars.

> *The most useful thing I did was - and I'm very grateful that we had the contact - one of the schools in the trust taught the same course and I went and spoke to the guy and he'd been teaching it for years. If I hadn't had him, I would have really floundered with this course. I would have had a lot of trouble. Because…I found it very hard to get information or find example work and my impression was it was because it was a new course.* Level 2 Award (Construction).

In the strongly vocational courses in our sample, communities of practice were less central to qualification delivery. Assessors were less likely to spend time proactively developing and contributing to networks; instead interaction with other centres was often discussed as a by-product of attending training. These VTQs were much more established in the centres we visited, however, and it is possible that communities of practice were more influential during the early days of qualification delivery.

Only one centre in our sample stated they did not use communities of practice in delivering their course. These assessors felt they simply did not have the time for networking.

> *We're very insular here, like every college I think. We don't really have any contact with outside colleges much do we?* Level 3 Diploma (Construction).

Although rarely explicit, there was a sense from our interviews of a collaborative mind-set between centres delivering these qualifications. Many appeared to be open to sharing experience and resources with other schools, both giving and receiving advice. Only 2 assessors alluded to any level of competition between centres. Both drew on communities of practice, but avoided working with a named local centre due to the more competitive nature of their relationship with this provider.

4 Discussion

One of the most striking features of this study was just how diverse, complex and dynamic the VTQ sector is. The variation between the 6 VTQs in our sample was considerable. Even within the same qualifications, we found variation in how assessors delivered courses. This was due in part to the level of devolved decision making afforded to centres. The variation in the VTQs in our sample has shaped the volume and complexity of data generated in the study. On one hand this has made analysis and reporting of results challenging. Attempting to distil a series of common themes which apply across this disparate landscape is difficult. On the other hand, this is a useful finding in itself. What we might take from this is that it is difficult to specify any one single approach when regulating this sector.

One of the aims of this study was to identify points of vulnerability in the delivery of internal assessment in VTQs. Whilst we did identify some potential vulnerabilities, our findings do not point to a straightforward set of improvements or course of action. Instead they highlight some of the tensions and trade-offs in the system. Often the very points of vulnerability in these VTQs double as their great strengths. Devolution of decision making to centres inevitably introduces risks to the reliability of internal assessments, with centres potentially having quite different approaches to setting tasks, supporting learners, interpreting assessment criteria and quality assuring judgements. However, this freedom also allows assessors to deliver qualifications which engage learners and best prepare them for the needs of local employers. It places assessment decisions in the hands of the experts. It is possible that introducing stricter controls over some aspects of internal assessment may enhance assessment reliability. At the same time, it could adversely affect the value of certain qualifications whether in terms of leaner engagement or assessment validity.

There are other challenges when considering how we might strengthen internal assessment in VTQs. This study has shown the internal assessment system to be a complex network of interconnecting parts. In many ways it is a finely balanced ecosystem. Any changes to such a system risks upsetting its delicate state of ecological balance. In this context we must think carefully about how any intended enhancements to internal assessment might undermine other parts of the system.

Nonetheless, there are some aspects of the system which we may choose to look at more closely. One of these is the variation in support and feedback given to learners by assessors as they undertake summative assessments. Centres delivering the same course often had different interpretations of what was permissible in terms of learner support. These differences have implications for assessment standards. At times, variation appeared to be caused by ambiguity in AO guidance. At others, it seemed to be a product of assessor unfamiliarity with such guidance.

This was not the only area where assessors highlighted some issues with the clarity or sufficiency of AO guidance. Assessors also identified issues with the guidance on task setting and making assessment judgements. Improving the clarity and availability of guidance and exemplars, particularly in the early years of qualification delivery, might seem a reasonably straightforward way of enhancing the delivery of internal assessment. However, given that we found some evidence of assessor unfamiliarity with existing AO guidance, it is unclear to what degree assessors would engage with additional guidance.

A third potential area of vulnerability was the quality assurance processes designed by centres and AOs alike. In both internal and external quality assurance there appeared to be some conflation of capacity building, on one hand, and quality control on the other. This may be particularly problematic in the EQA process. Where AOs devolve (sometimes significant) control of internal assessment to centres, it is vital they can assure that standards in these centres are being upheld. Although EQA appeared to be comprehensive and well-regarded by assessors, interviewees had little personal experience of EQAs ever overturning or adjusting assessment judgements. In the experience of some assessors in our sample, the EQA process was as much about supporting centres in their delivery of internal assessment as it was about scrutinising them.

EQAs are not the only agents in the system with dual roles. Whilst we refer to 'assessors' throughout the report, in reality our interviewees wear many hats. Indeed it is questionable if many would even see assessment as their primary role. Almost all assessors in our sample also taught the learners they assess. Some IQA the same courses. These multiple roles may be difficult to reconcile. As teachers, they have relationships with their learners and preconceptions of their competence, diligence or suitability for a particular job. With teachers and centres judged on the level of achievement of their learners, they also have a stake in their success. This may colour the judgements of even the most conscientious assessor or IQA. Although a handful of assessors noted the pressure of wanting their learners to perform highly, it was interesting that there was little real discussion of this inherent role tension. Certainly this issue is not unique to VTQs and is characteristic of internal assessment in any context. Nonetheless, any potential vulnerabilities which arise are best managed where all of those involved are conscious of their existence.

One interesting feature of this study was the suggestion that qualifications do not always work in the manner one might assume from reading specifications or assessment criteria. There was a sense that the delivery of these VTQs relied on a certain amount of tacit knowledge and understanding. This was particularly true in relation to the application of the assessment criteria where assessors may consciously or unconsciously apply an assessment standard which is subtly different from that intended by the AO. Although assessors told us they found assessment criteria simple to apply, they often drew on other sources of information when making assessment decisions. These may include their own interpretation of industry standards or their experience of delivering qualifications in the past. This was not always easy for assessors to verbalise and certainly we could have dedicated the full interview to unpicking the process of applying and conceptualising assessment criteria. What was evident from the interviews is that some assessors feel they have a certain 'ownership' of assessment standards.

Although we identified certain points of vulnerability and inconsistency in the delivery of internal assessment in these VTQs, we also found the qualifications were highly valued by assessors. They were seen as valid, engaging and pitched at an appropriate level of demand, whilst providing learners with the specific or more general skills employers need. In the strongly vocational qualifications, the importance of preparing learners to meet the needs of local employers was paramount. Almost every assessor delivering these VTQs stressed the importance of maintaining centre reputation and relationships with local employers by ensuring any learner gaining a particular qualification is highly competent. This often went beyond upholding the assessment standards and extended to delivering extra technical and

employability skills to learners to help them "stand out" to employers. In this sense, the centres in our sample delivering these qualifications seemed highly 'self-regulating'.

This study represents one of the few contemporary pieces of published research into the implementation of internal assessment of VTQs. It is interesting to note that our observations do tend to support the findings of historical research in the field. Some of the challenges of delivering internal assessment some 30 years ago still hold true today. For example, our study suggests that ambiguities exist in even the most apparently explicit assessment criteria, echoing findings reported by Wolf and Silver (1986), Black et al (1989) and Carter and Bathmaker (2017). We also found strong evidence of the central role of assessor judgement in the assessment process. Although the assessors in our sample endeavour to meet AO requirements in the form of assessment criteria, at the same time they view these criteria through the lens of their own experience, beliefs and interactions with other assessors. This gives each assessor a unique perspective on the assessment standard (see research by Wolf, 1995; Shay, 2005). Others have observed previously how this can cause variation in assessment delivery between centres (e.g. Torrance et al, 2005).

The study also serves to demonstrate some of the stated benefits of internal assessment. The assessors in our study were able to develop 'real-world' assessment tasks which help to prepare learners effectively for employment or help to engage and motivate them (see Ecclestone, 2000; Vitello and Williamson, 2017). Many assessors we interviewed also believed the focus on internal assessment can also improve the motivation and performance of learners who felt particularly anxious about exams, mirroring findings reported by Dufaux (2012).

One area where our research does not replicate previous studies in quite the same way is in the concept of 'instrumentalism' (whereby clarity in assessment criteria and processes can lead to a culture of criteria compliance in centres). In our study there was little evidence of the "assessment as learning" approach described by Torrance et al (2005) in their review of the sector. Indeed, in some of the centres in our study delivering strongly vocational qualifications, there was an emphasis on delivering learning above and beyond the course requirements to best prepare learners for employment. That said, we did find evidence of a variation in the level of support and feedback provided to learners when completing summative assessments.

Overall, this study has enhanced our understanding of how the delivery of internal assessment is influenced by both the nature of the specific qualification and the exact context in which that qualification operates. The findings will be crucial for directing further research into internal assessment in VTQs and informing future decisions around how best to regulate such qualifications.

4.1 Limitations of the study

The nature of our research questions lent themselves to a qualitative methodology which allowed us to explore the experiences of assessors in depth. There are some limitations associated with this methodology however. We present these below.

This was a small-scale, exploratory study. For each of our VTQs we interviewed assessors in a minimum of 3 centres. Given the small size of this sample, we cannot claim it is representative of the wider population. Because of this we cannot draw any strong comparisons between the different VTQs in our study. This is

compounded by the slightly different nature of evidence gathering in the different qualification types. For instance, almost all interaction with assessors delivering strongly vocational qualifications took the form of focus groups. In contrast, the softly vocational qualifications were more likely to involve one-to-one interviews. This may have affected the nature of the information collected from participants.

Given the considerations above, this research only allows tentative generalisations beyond the sample. Nonetheless, it does allow us to better understand the mechanisms through which assessment arrangements may work in practice. This can help us to identify whether there are systemic vulnerabilities which might threaten the validity or reliability of internal assessment in VTQs.

There were also limitations associated with the recruitment of research participants. Although we invited centres to participate in our study based on a randomly generated sample, we were reliant on centres agreeing to take part in the work. Self-selection can play a part here. It is unlikely that assessors who are not confident in their role (or are knowingly non-compliant with AO rules) would volunteer to take part in an interview with the qualification regulator. Our sample is therefore likely to comprise some of the more confident, conscientious and experienced assessors.

In this kind of qualitative research the presence of a researcher in the data gathering process is unavoidable and can influence the responses of participants. Prior to the interview we emphasised the confidentiality of the interviews and the fact that any information gathered would *only* be used for the purposes of research. Nonetheless, it is possible that participants were keen to present themselves and their centre in the best possible light, minimising any issues they have in the delivery of VTQs.

The exploratory nature of the study, and the breadth of subjects covered, meant that we needed to take a slightly fluid approach to conducting the interviews. Although strongly guided by an interview schedule, interviews were not limited to particular questions and were redirected by researchers in real time to follow up emerging issues. Some interviewees chose to focus much of their discussion on a particular aspect of the internal assessment process, others were more interested in exploring different areas. This flexibility was useful in generating additional findings but meant that each interview had slightly different areas of focus.

The breadth of the study also prevented us drilling down into great depth into any one topic. Whilst this is a limitation of the study it is also one of its strengths. The broad focus allowed us to capture a high-level picture of this internal assessment 'ecosystem' and to better understand how various factors in its context, design and delivery inter-relate with and influence one another.

Finally, we should acknowledge the 'situatedness' of participants may have impacted the nature of the information gathered. Our research participants were so immersed in the reality of delivering internal assessment in VTQs that it was sometimes difficult for them to reflect objectively on the process of this. To take one example, many assessors initially told us how straightforward they found the application of the assessment criteria, before going on to identify ambiguities on further reflection. It also was interesting to note a slight disconnect between the perceptions of assessors and any points of vulnerability in the system; the "rigorous" EQA process which apparently very rarely resulted in any changes to assessment decisions, for instance.

5 References

Aronson, J. (1995). A Pragmatic View of Thematic Analysis. *The Qualitative Report*, *2*(1), 1–3. Retrieved from http://nsuworks.nova.edu/tqr/vol2/iss1/3

Black, J.H; Hall, J.; Martin, S. and Yates, J. (1989).The quality of assessments: case studies in the National Certificate. Scottish Council for Research in Education.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Carter, B., & Bathmaker, A. (2017) Prioritising progression over proficiency: limitations of teacher-based assessment within technician-level vocational education, *Journal of Further and Higher Education*, 41:4, 460-474, DOI: 10.1080/0309877X.2015.1135881

Cuff, B., Zanini, N., and Black, B. (2018). An exploration of grade inflation in 'older style' level 3 BTEC Nationals. Ofqual/18/6459/2

Dufaux, S. (2012) Assessment for Qualification and Certification in Upper Secondary Education: A review of country practices and research evidence (OECD Education Working Paper 83). Paris: OECD Publishing.

Ecclestone, K. (2000) 'Bewitched, bothered and bewildered: A policy analysis of the GNVQ assessment regime 1992–2000'. Journal of Education Policy, 15 (5), 539–58.

Ecclestone, K. (2002). *Leaner Autonomy in Post-16 Education: The Politics and Practice of Formative Assessment*. London: Routledge Falmer.

Garland, P. (1998) Assessment in GNVQs: learning the hard way. *Research in Post-Compulsory Education*, 3:3, 329-344, DOI: 10.1080/13596749800200040

Harth, H & Hemker, B (2013). On the reliability of vocational workplace-based certifications, *Research Papers in Education,* 28:1, 75-90. DOI: 10.1080/02671522.2012.754228

Jessup, G. (1991). *Outcomes: NVQs and the Emerging Model of Education and Training*. London: The Falmer Press.

Johnson, M. (2008) Exploring assessor consistency in a Health and Social Care qualification using a sociocultural perspective, *Journal of Vocational Education and Training*, 60:2, 173-187, DOI: 10.1080/13636820802042446

Kelly, S. (2017) Reforming BTECs: Applied General qualifications as a route to higher education. Higher Education Policy Institute. HEPI Report 94.

Shay, S. (2005). The assessment of complex tasks: A double reading. *Studies in Higher Education* 30, no. 6: 663–679.

Sutherland, R., and Pozzi, S. (1995). The Changing Mathematical Background of Undergraduate Engineers: A Review of the Issues. London: Engineering Council

Torrance, H. (2007). Assessment *as* Learning? How the Use of Explicit Learning Objectives, Assessment Criteria and Feedback in Post-Secondary Education and Training Can Come to Dominate Learning. *Assessment in Education* 14 (3): 281–294

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

Vitello, S., and Williamson, J. (2017) Internal versus external assessment in vocational qualifications: A commentary on the government's reforms in England. *London Review of Education* 15 (3): 536 - 548

Wahlberg, M., & Gleeson, D. (2003) 'Doing the business': paradox and irony in vocational education – GNVQ business studies as a case in point. *Journal of Vocational Education and Training*, 55:4, 423-446, DOI: 10.1080/13636820300200238

Wolf, A., and Silver, R. (1986) Work Based Learning: Trainee Assessment by Supervisors. Sheffield: MSC R & D Series No. 33

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

Wolf, A. (1995). Competence-Based Assessment. Buckingham: Open University Press.

Annex A: Interview schedule

1. **Please describe a typical task (or set of tasks) that a learner might undertake for their summative internal assessment. Please talk us through a detailed example of such a task.**
   a. At what stage in the course of study do learners typically undertake this/these task(s)?
   b. Please describe what the learner physically produces for the assessment (e.g. a report, a portfolio, a product).
   c. When during the course of the learner taking the qualification does the assessment take place?
   d. To what extent do teaching and assessment overlap?
   e. (*If they have only described one assessment task*) How does this compare to other assessment tasks a learner undertakes?

2. **Please describe how summative assessment tasks are set for individual learners.**
   a. Are learners allocated a pre-set task or do they devise their own bespoke task?

       i.   If tasks are set by you, how do you develop them?
      ii.   If tasks are pre-set, have you used these tasks before?
    iii.   Are examples of previous learners' work for that task available?

b. What support or guidance do you have from your awarding organisation when setting tasks?

c. How much variety is there in the tasks which are set/undertaken?
       i.   Do you think these tasks are generally comparable?

d. How do you ensure that these different tasks allow you to make consistent assessment judgements?

e. Do all learners have to attempt the same assessment tasks, and provide the same kind of assessment evidence?
       i.   Is there any differentiation for learners working towards different grades?

f. Do awarding organisations check the tasks that you have set?
       i.   If so, how and when do they check your tasks?

3. **Please describe the type and frequency of support that learners receive throughout the assessment process.**

a. How much feedback are assessors able to provide learners with during their summative assessments (portfolios, practical tasks, etc.)?

b. What are the limits of the support that learners can receive from teachers/assessors?

c. To what extent do learners support each other?

d. Are learners able to look at work that was completed by other learners (i.e. from previous years)?
       i.   What level of support would you consider to be desirable?
      ii.   Is there a level of support which you would consider to be unacceptable?

4. **Please can you describe the process that you go through to apply the assessment criteria to the work produced by the learner in their summative assessments?**

a. Are you clear what 'standard' of performance you are looking at from learners?

b. Where have you got this understanding of the 'standard' from? (Experience? Industry? Assessment criteria?)

c. Are there any challenges or issues when applying the assessment criteria/marking the learners?
       i.   Are the assessment criteria usually clear? Have you got any examples where they are not?
      ii.   Are there a manageable number of criteria to assess?
    iii.   How effectively do criteria for different grades differentiate between learners of differing levels of competence?

d. How do you aggregate assessment evidence to come up with an overall grade/decision? Are there clear rules to support this?

5. **How do you standardise your assessment judgements across different teachers in the centre (and over time)?**
   a. What guidance or training do you receive from the awarding organisation in applying the assessment criteria/standard?
   b. What processes are in place to ensure that assessors are consistent in their judgement across learners?
   c. What does your centre do to ensure that assessment decisions are comparable across different assessors?
   d. What opportunity is there to discuss the assessment process with other assessors, in your centre or others?
   e. Do you see work from centres other than your own?

6. **Do you have any sense of how this qualification is assessed in other colleges or training providers?**
   a. If so do you believe the assessment standards are the same in other colleges/training providers?
   b. Do you have any concerns about how this qualification is assessed in other colleges/training providers?

7. **What processes are in place within your centre/training provider to internally quality assure or verify assessment judgements?**
   a. Who performs the IQA/IV role?
   b. How confident do you feel with regard to the accuracy and consistency of your own assessment judgements?
   c. Where does the IQA process work well?
   d. Where might it be improved?

8. **If you can, please describe the external quality assurance and external verification of assessment carried out by your awarding organisation?**
   a. How much confidence do you have in the effectiveness of this process?
   b. Where does this process work well?
   c. Where might it be improved?

9. Is there anything else you would like to tell us?

Note: It is only necessary to ask the main (numbered) questions. The sub questions (a/b., i/ii.) are intended as *possible* follow-ups. They should only be asked where appropriate. You may think of alternative questions that are more relevant.

Annex B: Consent Form

Ofqual

Reference number (for Ofqual use): _____

Consent form

Project Title: Evaluation of the functioning of internal assessment in technical and vocational qualifications

Research Team: Charlotte Lockyer, Stuart Cadwallader, Ben Cuff, Wendy Cotton, Paul Newton, Beth Black

By signing below I confirm that I agree with the following statements:

1. I understand the purpose of this research and by continuing with the study I will be consenting to participate.
2. I have had the opportunity to consider the information I have been given, ask questions and have had these questions answered satisfactorily.
3. I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason. I also understand that I have the right to change my mind about participating in the study and can withdraw from the study at any time.
4. I understand that this interview will be audio recorded, and this recording will be transcribed by an external transcribing company (under contract with Ofqual).
5. I understand that the data will be treated in a confidential manner and my anonymity will be maintained in any report of the research. I also understand that this data will be securely stored (by Ofqual) for up to 7 years before being destroyed.
6. I understand how Ofqual may use my personal data.
7. I understand that the statements I give in this study will be analysed with the hope to publish any findings.
8. I agree to take part in the above study.

Print Name: _____

Signed: _____ Date: _____

Researcher's Signature: _____

Published by:

ofqual

Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

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

February 2020

Ofqual/20/6594