

## **Internal versus external assessment in vocational qualifications: A commentary on the government's reforms in England**

Sylvia Vitello\* and Joanna Williamson  
*Cambridge Assessment*

### **Abstract**

The distinction between external assessment and internal assessment underpins a major reform to vocational qualifications underway in England. To be approved by the Department for Education, vocational qualifications must now include a minimum proportion of external assessment, regardless of subject. This paper discusses the nature and implications of this constraint on qualification design. First, it clarifies the meaning of external assessment and the key arguments underpinning the reform. Second, it evaluates the use and implementation of this blanket rule. The final section discusses the nature of internal assessment in more detail, highlighting its heterogeneity and potential advantages over external assessment.

**Keywords:** vocational education; vocational qualifications; technical education; external assessment; qualifications; governmental reforms

### **Background**

The education system in England is currently undergoing major reform by the Department for Education (DfE), affecting both vocational and academic qualifications. This paper provides a commentary on the DfE's decision regarding external assessment in vocational qualifications taken by 14- to 19-year-olds, with a focus on school-based Key Stage 4 provision (typical age: 14 to 16). Before vocational qualifications can be approved for inclusion in school performance tables, they must now include a minimum amount of external assessment, regardless of subject area (meaning that this is a blanket rule). We discuss this reform from two angles. First, we discuss the use and implications of the blanket rule of external assessment for vocational qualifications and the use of blanket rules more generally in qualification design. Second, we evaluate the affordances of internal assessment as an assessment method, since the implication of the reform is that it can never be sufficient unless coupled with external assessment.

### **External versus internal assessment**

The distinction between external and internal assessment underpins a key reform to vocational qualifications. The DfE (2015a) has defined external assessment as:

---

\* Corresponding author – email: [vitello.s@cambridgeassessment.org.uk](mailto:vitello.s@cambridgeassessment.org.uk)

©Copyright 2017 Vitello and Williamson. This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

a form of assessment in which question papers, assignments and tasks are specified by the awarding organisation, then taken under specified conditions ... and marking or assessment judgements are made by the awarding organisation.

(DfE, 2015a: 18)

It is clear that this definition concerns the processes of setting, taking and marking assessments but does not specify the type of task that a student should take. This distinction is important to remember because external assessments are often associated with written, time-bound examinations but do not have to be.

Internal assessment is, hence, any form of assessment in which any of the three above assessment processes are controlled by the institution where the student is studying (for example, school or workplace). One well-known example is teacher-based assessment.

As part of the reforms, the DfE has categorized vocational qualifications taken by 14- to 19-year-olds into four types (DfE, 2015a; DfE, 2015b), and specified that they must now contain a minimum amount of external assessment in order to be approved for use in school performance tables. For 14- to 16-year-olds, there is only one type: Technical Awards. These are broad, applied qualifications that do not focus on a specific occupation. For 16- to 19-year-olds, there are three types: Applied General, Tech Levels and Technical Certificates. Applied General qualifications enable students to continue their general education through applied learning. Tech Levels and Technical Certificates are technical qualifications that equip students with specialist knowledge of a specific industry or occupation. As Table 1 shows, for all Technical Awards, regardless of subject, external assessment must contribute to at least 25 per cent of the overall grade for 2017 performance tables, and this amount needs to rise to 40 per cent for 2018. In contrast, Tech Levels only need 30 per cent external assessment. Maintained schools in England may still offer other vocational qualifications (that is, not meeting the criteria for performance tables), especially for students with particular education needs. However, the critical importance of performance tables for school accountability in England means that schools are likely to choose qualifications approved for performance table recognition in all but exceptional cases, as the reform intends.

**Table 1:** Percentage of external assessment required by the DfE (2015a; 2015b)

Qualification	Technical Award	Technical Certificate	Tech Level	Applied General
<b>External assessment (%)</b>	25 (2017 tables) 40 (2018 tables)	25	30	40

One of the prominent reasons for the government’s decision to use more external assessment in vocational qualifications is concern with assessment quality. The reform followed the Wolf (2011) review, which argued that external assessment is able to ‘safeguard against downward pressure on standards’ (112). The DfE (2015a) subsequently emphasized a connection between this type of assessment, rigour and the esteem of academic qualifications. More recently, external assessment was again highlighted as an important feature of the post-16 skills plan that followed the Sainsbury report on technical education; this time it was considered essential for ‘comparability and reliability’ (BIS and DfE, 2016: 52). These reports do not give an explicit definition of assessment ‘quality’ but draw attention to multiple quality-related constructs. We similarly take a multifaceted view of assessment quality, but try to be precise in terms of the facet of quality to which we refer in each instance. For example, ‘quality’ can mean that qualification results are valid, fit for purpose, reliable, lead to progression within the labour market or further education, and/or increase earnings.

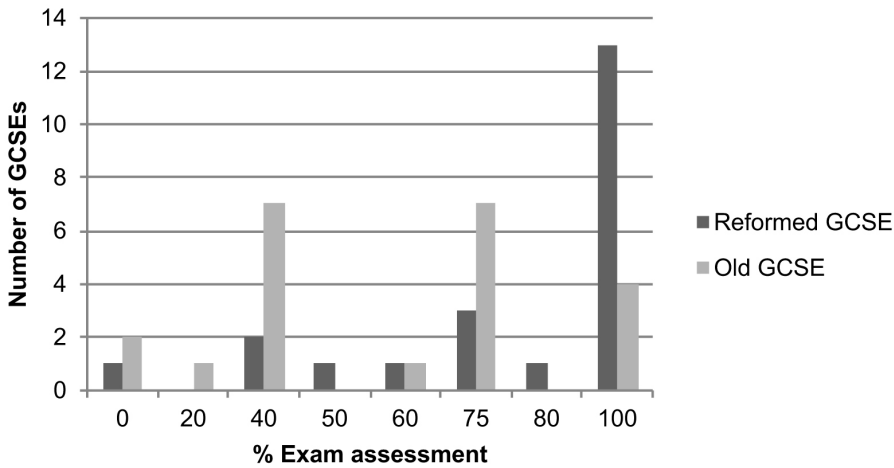
There are various reasons why external assessment may increase facets of quality. For example, if the awarding body has control over the assessment processes, it could reduce threats to validity such as malpractice, and may reduce variability in standards between institutions (Wolf, 2011). While external assessment could reduce some threats to quality, in certain situations the overall effect of using external assessment may not necessarily be positive. For example, the level of control that is needed may limit the types of knowledge and skills that can be assessed (this is discussed further below). More generally, because many factors affect the appropriateness of an assessment, it does not seem straightforward to justify having blanket rules that specify a minimum amount of external assessment that does not take into account the subject area, unless, of course, the advantages override the disadvantages overall. Unfortunately, it is not clear whether this is the case for vocational qualifications, since much of the debate about external/internal assessment has focused on academic qualifications (for example, S. Johnson, 2013; QCA, 2006) or on the distinction between summative and formative assessments (for example, SQA, 2007). It is complicated, and may even be inappropriate, to generalize those research findings because vocational qualifications differ from other qualifications in many ways, including the purpose of the qualification and the type of candidates.

### **A blanket rule of external assessment**

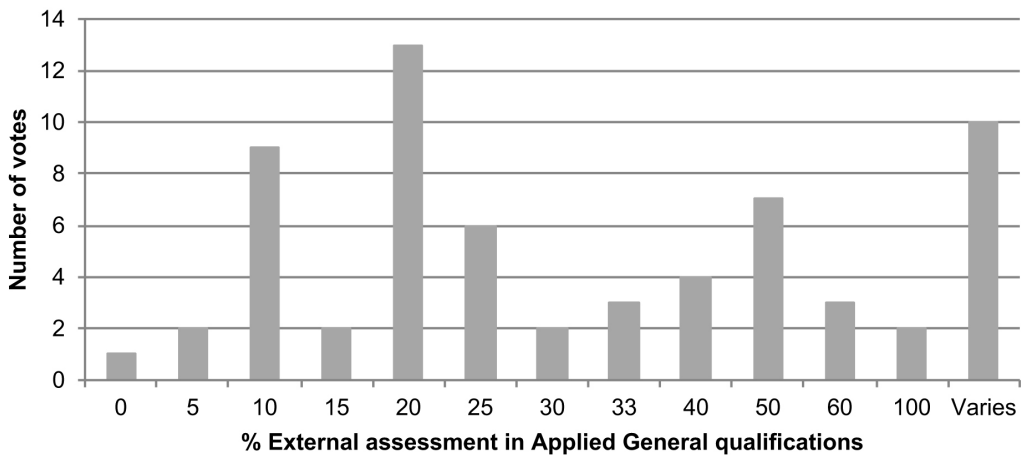
From a policy perspective, this blanket rule is surprising because it contrasts with the government's and regulator's positions on other assessment practices. For example, in 2006, the QCA (Ofqual's predecessor) advised against the use of a blanket rule of 66 per cent internal assessment for GCSEs in vocational subjects (QCA, 2006). In the recent reforms to GCSEs, Ofqual (2017) has allowed some degree of flexibility between the assessment practices of different subjects. Ofqual required exams to be the default method of assessment for all GCSEs, but still considered the use of non-exam assessment on a subject-by-subject basis. The importance of that flexibility is highlighted by the variability in exam assessment among the reformed GCSEs, as shown in Figure 1. The proportion of exam assessment in reformed GCSEs ranges from 0 per cent to 100 per cent. Although the majority of GCSEs now have 100 per cent exam assessment (13 out of 22 subjects, including all English, maths and science GCSEs), there are multiple GCSEs that include non-exam assessment. Even more noteworthy is that one GCSE (art and design) is not required to have any exam assessment at all, and, in fact, is not even required to include *external* assessment of any kind. This seems to conflict with the requirement of a minimum proportion of external assessment for vocational qualifications.

Contrasting GCSEs with vocational qualifications in terms of assessment practices draws attention to differences in the corresponding governmental reforms, especially the use of a blanket (subject-general) rule for the vocational qualifications compared to a subject-driven consideration process for GCSEs. However, it is debatable whether we can, or should, generalize the assessment practices used for academic qualifications to vocational qualifications (Acquah and Malpass, 2017). Another way to evaluate the proportion of external assessment appropriate for vocational qualifications is to ask stakeholders. In 2013, the DfE organized a consultation to assess stakeholders' views of the reforms it was then proposing (DfE, 2013). Respondents (organizations, teachers and employers) were asked their views on the minimum proportion of external assessment in Applied General qualifications. Figure 2 shows a lack of consensus among the respondents. The proportions ranged from 0 per cent to 100 per cent, with 10 out of 64 respondents stating that the level should vary. Almost two-thirds of respondents stated that the proportion of external assessment should be 33 per cent or less. The DfE decided on 40 per cent for Applied General qualifications, but it is unclear why this percentage was chosen from

the distribution of responses. Furthermore, the consultation did not ask stakeholders about other types of vocational qualifications, and it is unknown how decisions were made for those qualifications.



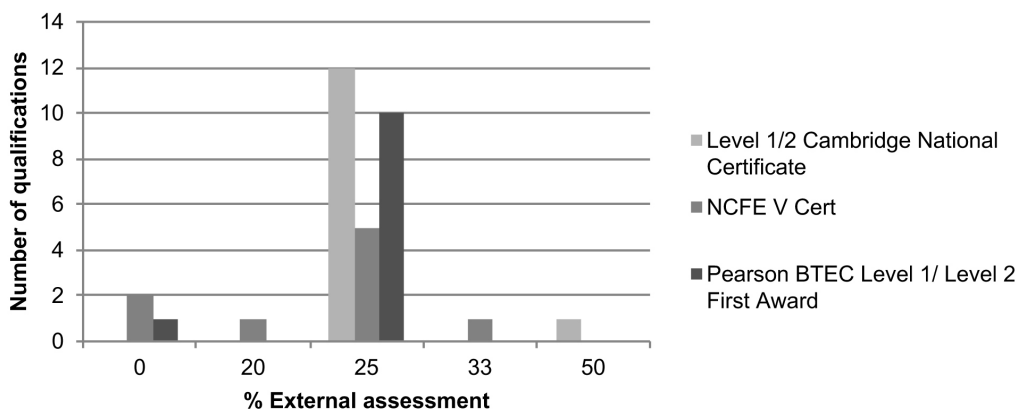
**Figure 1:** Percentage of exam assessment in GCSEs



**Figure 2:** Consultation responses to 'What is the minimum proportion of the content of an Applied General qualification that should be subject to external assessment?' (figure created by authors from data in DfE, 2013)

A third way to evaluate the use of external assessment is to consider practice in existing qualifications, which may provide some insight into the level of demand for external assessment among stakeholders. We investigated Level 1 and 2 vocational qualifications because they target the same age group as GCSEs and would fall into the DfE's Technical Award category. We primarily focused on Cambridge Nationals, offered by OCR (Oxford, Cambridge and RSA). We also investigated comparable qualifications offered by two other awarding bodies in similar subjects: Pearson BTEC Firsts and NCFE V Certs.

Figure 3 shows that there is limited use of external assessment in these vocational qualifications. Although the proportion of external assessment ranges from 0 per cent to 50 per cent, the majority of qualifications have 25 per cent external assessment, which is much less than the 40 per cent requirement for 2018 performance tables. Many V Certs (in similar subjects to the Cambridge Nationals) have less than the 25 per cent requirement. This overview of current qualifications suggests the possibility that certain subjects may not be suited to external assessment. Of course, it is important to regularly re-evaluate assessment in light of current context and research evidence, as this may change the consensus on the most appropriate assessment methods for a given domain of skills/knowledge.



**Figure 3:** Frequency of vocational qualifications by percentage of external assessment, grouped by awarding organization

Although many researchers argue against a ‘one size fits all’ approach to qualification design, the suitability of any blanket rule is affected by both the homogeneity of the objects of the rule (that is, what is affected by the rule) and the scope of the rule itself (that is, the variation in the outcome of the rule). Vocational qualifications, the object of the external assessment rule, remain by nature a heterogeneous set of qualifications (for example, in terms of subject area and specific occupational content), despite recent reforms reducing their number.

The lack of homogeneity among vocational qualifications may not necessarily be problematic for implementing the blanket rule, if the rule is broad enough in scope. The rule needs to ensure that the core skills/knowledge can be assessed appropriately for all subjects and that their assessment is not limited by having to include a certain proportion of external assessment. In principle, external assessment has wider scope than, for example, the requirement of exam assessment that has affected GCSE reform, because it does not specify the type of task that a student needs to do. Besides an exam, an external assessment could be a speaking test that is recorded by the teacher and marked by an external examiner (as in GCSE modern foreign languages), or a performance assessment that is set by the awarding body and marked by a visiting examiner (as in GCSE drama). Both of those examples fall outside Ofqual’s definition of exam assessment (‘taken by all students at once, under formal supervision, and are set and marked by exam boards’ (Ofqual, 2014: 10)) because students are not all assessed at the same time. Computer-based assessments could also be implemented as external assessments. For example, shipping licence examinations in some countries involve PC-based simulation tests that are externally assessed (Gekara *et al.*, 2011).

Although examples exist, the current use of externally assessed non-exams is limited. For example, our research into OCR Cambridge Nationals and Pearson's BTEC Firsts (in similar subjects) showed that all their external assessments were exams, either written or on-screen. One reason is that non-exams may not be feasible operationally. Non-exams may increase demands on centres' resources in terms of the physical environment and equipment needed for the assessment, which may affect the manageability and costs of the process. They may also increase demands on the awarding body, including examiner recruitment, setting and marking processes. Therefore, even if the blanket rule is theoretically broad in scope, which allows assessments to be, to some extent, tailored to the specifics of the qualifications, only a limited range of variants may be viable. This is problematic if those variants are not appropriate for the qualification. For example, written exams may not validly assess certain constructs or may induce negative 'washback' effects on teaching and learning (Alderson and Wall, 1993), both of which can threaten the validity of qualifications. The ultimate unfortunate consequence may be to reduce the provision and diversity of qualifications on offer to students.

These issues surrounding how to implement external assessment raise the more general concern with a blanket rule (irrespective of scope), which is that it affects, and to some extent conflicts with, approaches to qualification development. The Cambridge Assessment Group, for example, employs an integrated model of assessment (the Cambridge Approach) in which numerous factors are considered during assessment design, including a clear statement of purpose, identification of candidate populations and cataloguing of the constructs that are the focus of the assessment. It acknowledges the importance of complying with national and international criteria but emphasizes that the process should be evidence-based (Cambridge Assessment, 2009). Other approaches to assessment development similarly stress that it should be grounded on a model of cognition and learning (Pellegrino *et al.*, 2001). This clearly contrasts with the use of a blanket rule, which puts the type of assessment at the forefront of qualification design processes.

### **The useful heterogeneity of internal assessments**

The introduction of a blanket rule of external assessment for vocational qualifications assumes that internal assessment is not fully adequate to assess the constructs of the qualifications. To some extent this seems tenuous, considering the fact that internal assessment is not a uniform construct. Based on the DfE's definition of external assessment, there can be seven variants of *internal* assessment, which differ in terms of the level of control that the awarding body has over three stages of the assessment process: task setting, task taking and task marking (see Table 2). It is important to note that our discussion of 'control' is concerned with core assessment processes (those that determine an assessment to be internal or external) and not moderation or verification procedures. Some of these alternatives may offer advantages over the awarding body having full control, for certain subjects and qualifications. The variety of internal assessment and its potential usefulness has been acknowledged by national regulators for general academic qualifications (Ofqual, 2013; QCA, 2006) but has not been given attention in discussions of vocational qualifications. Instead, even in the most recent governmental document on vocational (technical) education (the *Post-16 Skills Plan*), external assessment is, again, stressed as necessary for ensuring comparability and reliability of qualifications (BIS and DfE, 2016).

The following sections consider each stage of the assessment process in turn and evaluate the impact of internal control (that is, not by the awarding body). The aim is to examine whether internal control at any of the three stages necessarily results in poor-quality vocational

qualifications, or whether in some cases the positive benefits offered by forms of internal assessment might achieve an overall better model.

**Table 2:** Locus of control in types of external and internal assessment ('external' means the process is under full control of the awarding body)

Type of assessment	Assessment process		
	Task setting	Task taking	Task marking
External assessment	external	external	external
	external	external	internal
	external	internal	external
	external	internal	internal
Internal assessment	internal	external	internal
	internal	external	external
	internal	internal	external
	internal	internal	internal

### **Task setting**

The aim of task setting is to ensure that students take tasks that are valid and are marked reliably. A student's performance should reflect his/her level of understanding of the topic being assessed (validity) and he/she should perform comparably on assessments that test comparable content (reliability).

#### *Arguments against internal control*

One major concern of tasks that are internally set is that they may have lower validity and reliability than externally set tasks (Wolf, 2011), which would occur if teachers or training providers set tasks that are different in standards. Assessments could also differ in terms of the content or learning outcomes that are chosen to be assessed. Various studies on performance assessments and portfolio use have found that a large proportion of the variability in student scores can be attributed to the sample of tasks with which a student is tested (task-sampling variability), combined with the fact that tasks are typically only performed on one occasion (occasion-sampling variability) (for example, Shavelson *et al.*, 1999).

However, negative effects of task variability can be minimized. Shavelson (2013) argues that score variability caused by sampling of tasks can be minimized either by making items more homogeneous or by having a larger number of tasks. He argues that complex domains, such as performance assessment, require the latter option. In line with that argument, because vocational qualifications may be conceived as complex domains, internal assessments may be effective if teachers or trainers are skilled at setting a variety of tasks. The concern is that they might *not* set appropriate tasks; this contributed to the removal of internal assessment for most GCSEs and A levels (S. Johnson, 2013; QCA, 2005; QCA, 2006). There are many reasons why tasks might not be set appropriately. For example, internal setters might lack experience in designing summative assessments. Teachers might set tasks that they believe suit their students' style of learning, which would threaten validity if the student's performance is not reflective of their level of understanding. The high-stakes accountability system in which vocational qualifications are provided may put pressure on teachers to choose assessments/content that they believe are

the easiest for students, such as highly artificial tasks that control the successful production of work that meets grading criteria (Ofqual, 2013). Nevertheless, various procedures can be put in place to help internal setters; for example, awarding bodies could provide advisers, although this has not always been successful (QCA, 2005: 12). Ultimately, the above arguments against internal control rest on the assumption that it is difficult to ensure comparability of standards through external verification or training by the awarding body.

Even if validity and reliability could be assured, there may be other disadvantages of asking teachers or training providers to set assessment tasks for their students. A widely used assessment method in vocational qualifications in the UK is evidence accumulation, also known as portfolios. Evidence accumulation has been criticized for being time-consuming, focusing students' attention on assessment and distracting them from their learning (de Bruler, 2001; Wolf, 2011). Similarly, some GCSE teachers have reported that task setting for certain summative assessments is burdensome (Ofqual, 2013). These kinds of negative perceptions might change to positive if assessments are viewed as tools that can facilitate the learning process, rather than merely for measurement (Earl, 2013).

#### *Arguments for internal control*

One of the main arguments for internal control at task setting is that it could produce assessments that are *more* valid than ones that are centrally (that is, externally) set. In the context of National Vocational Qualifications (NVQs), Jessup and colleagues have argued that specific assessment criteria (that is, an external reference point) and authentic tasks should be at the forefront of assessment design, and that reliability would naturally result from valid assessments (Jessup, 1991). This weighting of validity versus reliability is controversial; ideally, assessments should be both valid and reliable.

The idea that validity may increase with internally set tasks is one worthy of further consideration, especially if reliability can also be ensured. There are several ways by which validity may be enhanced. First, it may enable teachers or employers to devise assessments that reflect workplace contexts (that is, assessments are more authentic); this may be especially valuable for technical qualifications, given their vocational nature. Centres may have different facilities or infrastructure in place to assess students, and awarding bodies may not be able to exhaustively specify each possible arrangement.

Validity may also increase by allowing the assessment to take into account the characteristics of the students who take the qualification. Despite governmental desires for parity of esteem, vocational qualifications are still disproportionately taken by low-attaining or disaffected students compared to academic qualifications (for example, Smith *et al.*, 2015). Student-centred assessments may enhance performance, motivation and learning (Ecclestone, 2000) and hence constitute a better representation of the student's understanding. This argument is not intended to suggest that each individual student should have a different task but, instead, to highlight the effects of task formats on performance, which could be reduced by internal setters. There are many examples of interactions between the type of assessment that students take and their performance in non-vocational contexts. For example, research has found several types of gender differences, including that boys and girls perform differently on examinations and coursework (Elwood, 1999). The question is whether internal setters have enough expertise to provide tasks that increase validity by giving candidates the opportunity to demonstrate their understanding, rather than merely setting tasks that unfairly facilitate student performance. A compromise may be for the awarding body to set the task (that is, an external assessment approach) but on the



condition that they provide a range of tasks. A certain amount of constrained task choice is typical in academic and vocational qualifications.

The regulator, Ofqual, appears to recognize that internally set tasks have value, at least in certain cases. The subject-level requirements for GCSE art and design specify that ‘an awarding organisation must ensure that of the total marks available for a GCSE Qualification in Art and Design, 60 per cent of those marks shall be made available through tasks set by a Centre (“Internally Set Assessments”)’ (Ofqual, 2015c: 12). More specifically, the awarding body ‘must ensure that each Internally Set Assessment is designed to (a) require a portfolio of work to be completed by the learner’. In this case, despite Ofqual’s overall preference for external exam assessment, specific subject needs have led not only to allowing but to requiring 60 per cent of marks to be allocated to internally set, portfolio assessment.

### **Task taking**

#### *Arguments against internal control*

The major concerns of low control at task taking are malpractice, such as cheating and plagiarism by students. Without external control, there may be more opportunity for students to receive inappropriate help and advice from parents and/or teachers during the task. This would reduce the value of the assessment grade as an indicator of students’ understanding of the topic. The QCA’s (2005) study of GCSE coursework found variability in the extent to which teachers encouraged revision and redrafting of work, and supplied writing frames, templates and checklists to students. Some parents even reported having drafted their own children’s coursework.

#### *Arguments for internal control*

The way in which the task is taken may have effects on students. If the conditions of the tasks are restrictive, then it may lower students’ motivation and increase their anxiety, which may ultimately affect their level of performance or increase the risk that they drop out part way through the course (Stasz *et al.*, 2004). For example, in a review of assessment practices in upper secondary education, Dufaux (2012) argued that low-performing students may be affected by the type of task, feeling more discouraged and stressed under the pressure of exams.

The conditions under which an assessment is taken may have different challenges for different subject domains and different types of learning outcomes. This may be particularly problematic for vocational qualifications, where the learning outcomes relate to practical and technical skills rather than only to knowledge. For example, it may be difficult to determine the length of time that students should take to complete the task, where students should complete the task (for example, in a classroom, work environment or at home), what information they should have access to when completing the task (for example, book chapters or internet-based resources) and the amount of feedback that teachers should give them. If these conditions are restrictive, then it could affect the skills that students are able to develop or demonstrate through the task, affecting the validity of the assessment. For example, in a review of GCSE controlled (internal) assessment, Ofqual (2013) highlighted that timing and feedback restrictions placed on English literature coursework tasks limited students’ opportunity to draft and redraft work, and therefore prevented the assessment from testing those skills.

Another advantage of internal control is that it may minimize disruption to teaching, depending on where and when students can do the task. This has also been used as an argument for reducing the quantity of exams in general (Johnson, 2013).

## **Task marking**

### *Arguments against internal control*

The major concerns about internal control at task marking are that it increases the risk of malpractice (deliberate or unintentional) and may lower the reliability of the marking. The risk of malpractice during marking is high for qualifications for which teachers are under pressure to give high marks, from their students or from the education system (grade inflation). Vocational qualifications offered to 14- to 19-year-olds are of that kind. They are high-stakes for both students and teachers; in particular, students' grades affect their progression to further study and form part of school accountability and funding regimes. However, although malpractice threatens the validity of assessments, awarding bodies put in place procedures to minimize its incidence by, for example, moderating teachers' marks or statistically screening for malpractice (Ofqual, 2011).

Several reviews have been conducted into the reliability of teacher marking for summative or high-stake purposes, but few in the context of vocational qualifications (M. Johnson, 2006; S. Johnson *et al.*, 2013). Harlen (2005) discusses research that has shown high reliability of marking by teachers in school-based assessments but concludes that the evidence is not strongly favourable. Similarly, S. Johnson (2013) concludes that the evidence is limited and often ambiguous. Both authors suggest ways in which it is possible to achieve higher reliability, such as by consensus moderation, training to make markers aware of potential biases in their decision making, or more detailed marking criteria and assessment guidance.

Even if reliability could be assured, there may be other disadvantages of teachers marking the assessments. In particular, teachers have complained that marking can be time-consuming (Ofqual, 2013).

### *Arguments for internal control*

S. Johnson (2013) argues that permitting teachers to mark could broaden the scope of the assessment by exploiting 'the rich base of evidence that teachers have available to them ... by virtue of the time spent interacting with ... their students, that could in principle lead to greater validity and reliability' (2013: 92). The same argument could be made for making judgements about students' knowledge and technical skills in a vocational context.

An internal marker may have positive effects on students if the marking needs to occur in their presence (such as assessing live drama performance), for example by reducing a student's level of anxiety compared to an external marker. Little research has examined the effects of external markers on students' performance. There is some evidence that the perceived attitude of the examiner is noticed by students during an exam (Siddiqui, 2013), and therefore could, in theory, affect performance if it is negative. In other research, students have been found to be more anxious in assessments that are assessed by examiners *in situ*, such as oral examinations (Huxham *et al.*, 2012; Pearce and Lee, 2009), although this research did not assess whether this is moderated by the type of examiner. In contrast, another study, in this case with primary school children, has found some evidence that an external examiner may reduce, not increase, levels of anxiety (Bertoni *et al.*, 2013).

There may be operational advantages to using internal markers in certain circumstances. For example, an internal marker may be more manageable logistically and less costly when the marking needs to occur during the assessment. It may be time-consuming, if not impossible, for a sufficient number of external examiners to attend a large number of test centres and/or a large cohort of students. Visiting examiners may not be required if the assessments can be

delivered to external examiners. This would be straightforward for written assessments, which is the practice for written examinations, coursework and portfolios. It can also be achieved by recording oral and performance assessments. However, those types of recordings might be difficult to achieve if test centres do not have access to the required equipment or lack confidence or expertise in using it. Although such potential difficulties may be overcome through adequate training procedures, internal markers, who can mark *in situ*, may be more efficient.

Once again, it is interesting to consider the example of GCSE art and design. The reformed GCSE does not require any external marking (Ofqual, 2015a: 12), after respondents to the consultation on reformed GCSE art and design 'raised concerns about the practicality and validity of external marking in art and design' (Ofqual, 2015b: 2). Although GCSE and vocational qualifications differ, it is difficult to see how concerns about the 'practicality and validity of external marking' deemed valid for the GCSE in art and design would not also apply to a Level 2 vocational qualification in art and design.

## **Conclusion**

The DfE has specified that all vocational qualifications in England must include a minimum amount of external assessment in order to gain government recognition in performance tables. This requirement is a blanket rule with no apparent flexibility for the proportion to be modified on a subject-by-subject basis, although it does vary by type of qualification. This blanket rule contrasts with the government's (slightly) more flexible position on assessment regulations for other qualifications (for example, the exam requirement for GCSEs). The DfE has provided little evidence of the rationale or consultation responses that underpin the proportions of external assessment that have been chosen, which is especially important because they diverge from the current use of external assessment in vocational qualifications. Despite these concerns, the blanket rule may not necessarily be problematic to implement, even though vocational qualifications are heterogeneous in nature, because external assessment is theoretically wide in scope. However, in practice, practical and economic factors may mean that external assessment is operationalized as an examination, which may not be appropriate for all subjects.

The requirement for a minimum proportion of external assessment implies that internal assessment is unsatisfactory. The main argument against internal assessment at all three main stages of the assessment process (task setting, task taking and task marking) is quality assurance. It is argued that quality is more at risk of being compromised if the awarding body does not have control of the process. Although there is some evidence supporting this possibility, there are also mechanisms that can be put in place to minimize this risk. For each stage of the assessment process, it has also been argued that internal assessment may enhance the quality of the qualification, in particular when we consider the characteristics of the cohort that typically take vocational qualifications and the heterogeneous nature of work environments that students might be exposed to.

Assessment decisions must take into account the specific context in which the qualifications are provided. Solutions posited to address concerns about internal assessment are inevitably constrained by practical factors (for example, finances) and their success is likely to be moderated by the high-stakes accountability and funding system in England, which may put pressure on lowering standards (for example, by grade inflation). Since internal assessment has a variety of potential merits, it is critical that it is not dismissed as an assessment method and, instead, that efforts continue to be made to devise feasible ways by which to ensure its validity and reliability (for example, *AlphaPlus*, 2014). The external versus internal debate exists alongside other key debates on vocational assessments more generally. For example, there is

ongoing controversy surrounding the authenticity (or lack of authenticity) of school-based tasks for vocational understanding and whether school teachers have the professional competence to provide vocational courses. It is likely that these debates interact such that advances in resolving one might help enhance the others. It is plausible that more authentic tasks could lead to more valid internal assessment, but also that better internal assessment guidance could free up centres to use more authentic tasks.

This paper calls into question the idea that external assessment will inevitably be of higher quality than internal assessment. It highlights the need for any evaluation of internal assessment to include a more comprehensive list of advantages and disadvantages that takes into account the nature of vocational qualifications (for example, type of cohort) and, equally as importantly, evaluates potential disadvantages against potential solutions. This type of evaluation is likely to lead to different conclusions for different qualifications and subjects.

### Notes on the contributors

Sylvia Vitello and Joanna Williamson are research officers in the Assessment Research and Development division at Cambridge Assessment. Their current research focuses on aspects of educational measurement and qualification reform.

### References

- Acquah, D.K. and Malpass, D. (2017) 'The technical baccalaureate: Providing excellence in vocational education?'. *Assessment in Education: Principles, Policy and Practice*, 24 (1), 96–117.
- Alderson, J.C. and Wall, D. (1993) 'Does washback exist?'. *Applied Linguistics*, 14 (2), 115–29.
- AlphaPlus (2014) *Validation of Vocational Qualifications: Final report*. Coventry: Office of Qualifications and Examinations Regulation.
- Bertoni, M., Brunello, G. and Rocco, L. (2013) 'When the cat is near, the mice won't play: The effect of external examiners in Italian schools'. *Journal of Public Economics*, 104, 65–77.
- BIS (Department for Business, Innovation and Skills) and DfE (Department for Education) (2016) *Post-16 Skills Plan: Presented to Parliament by the Minister of State for Skills by command of Her Majesty: July 2016*. London: Department for Education.
- Cambridge Assessment (2009) *The Cambridge Approach: Principles for designing, administering and evaluating assessment*. Cambridge: Cambridge Assessment.
- de Bruler, C.A. (2001) 'Assessment, Knowledge and the Curriculum: The effects of a competence-based approach to the training of teachers in further and adult education'. EdD thesis, The Open University.
- DfE (Department for Education) (2013) *Consultation on the Reform of Level 3 Vocational Qualifications for 16–19 Year Olds: Government response*. London: Department for Education.
- DfE (Department for Education) (2015a) *Technical Awards for 14 to 16 Year Olds: 2017 and 2018 performance tables: Technical guidance for awarding organisations*. London: Department for Education.
- DfE (Department for Education) (2015b) *Vocational Qualifications for 16 to 19 Year Olds: 2017 and 2018 performance tables: Technical guidance for awarding organisations*. London: Department for Education.
- 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.
- Earl, L.M. (2013) *Assessment as Learning: Using classroom assessment to maximize student learning*. 2nd ed. Thousand Oaks, CA: Corwin Press.
- 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.
- Elwood, J. (1999) 'Equity issues in performance assessment: The contribution of teacher-assessed coursework to gender-related differences in examination performance'. *Educational Research and Evaluation*, 5 (4), 321–44.

- Gekara, V.O., Bloor, M. and Sampson, H. (2011) 'Computer-based assessment in safety-critical industries: The case of shipping'. *Journal of Vocational Education and Training*, 63 (1), 87–100.
- Harlen, W. (2005) 'Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes'. *Research Papers in Education*, 20 (3), 245–70.
- Huxham, M., Campbell, F. and Westwood, J. (2012) 'Oral versus written assessments: A test of student performance and attitudes'. *Assessment and Evaluation in Higher Education*, 37 (1), 125–36.
- Jessup, G. (1991) *Outcomes: NVQs and the emerging model of education and training*. London: Falmer Press.
- Johnson, M. (2006) 'A review of vocational research in the UK 2002–2006: Measurement and accessibility issues'. *International Journal of Training Research*, 4 (2), 48–71.
- Johnson, S. (2013) 'On the reliability of high-stakes teacher assessment'. *Research Papers in Education*, 28 (1), 91–105.
- Johnson, S., Johnson, R., Miller, L. and Boyle, A. (2013) *Reliability of Vocational Assessment: An evaluation of level 3 electro-technical qualifications*. Coventry: Office of Qualifications and Examinations Regulation.
- Ofqual (Office of Qualifications and Examinations Regulation) (2011) *GCSE, GCE, Principal Learning and Project Code of Practice (Ofqual/11/4850)*. Coventry: Office of Qualifications and Examinations Regulation.
- Ofqual (Office of Qualifications and Examinations Regulation) (2013) *Review of Controlled Assessment in GCSEs (Ofqual/13/5291)*. Coventry: Office of Qualifications and Examinations Regulation.
- Ofqual (Office of Qualifications and Examinations Regulation) (2014) *Developing New GCSE, A Level and AS Qualifications for First Teaching in 2016 (Ofqual/14/5487)*. Coventry: Office of Qualifications and Examinations Regulation.
- Ofqual (Office of Qualifications and Examinations Regulation) (2015a) *Confirmed Assessment Arrangements for Reformed GCSE, AS and A Level Qualifications (Ofqual/15/5590)*. Coventry: Office of Qualifications and Examinations Regulation.
- Ofqual (Office of Qualifications and Examinations Regulation) (2015b) *GCSE Art and Design: Conditions and guidance (Ofqual/15/5667)*. Coventry: Office of Qualifications and Examinations Regulation.
- Ofqual (Office of Qualifications and Examinations Regulation) (2015c) *GCSE Subject Level Conditions and Requirements for Art and Design (Ofqual/15/5643)*. Coventry: Office of Qualifications and Examinations Regulation.
- Ofqual (Office of Qualifications and Examinations Regulation) (2017) 'Guidance: Summary of changes to GCSEs from 2015'. Online. [www.gov.uk/government/publications/gcse-changes-a-summary/summary-of-changes-to-gcse-from-2015](http://www.gov.uk/government/publications/gcse-changes-a-summary/summary-of-changes-to-gcse-from-2015) (accessed 28 July 2017).
- Pearce, G. and Lee, G. (2009) 'Viva voce (oral examination) as an assessment method: Insights from marketing students'. *Journal of Marketing Education*, 31 (2), 120–30.
- Pellegrino, J.W., Chudowsky, N. and Glaser, R. (eds) (2001) *Knowing What Students Know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- QCA (Qualifications and Curriculum Authority) (2005) *A Review of GCE and GCSE Coursework Arrangements*. London: Qualifications and Curriculum Authority.
- QCA (Qualifications and Curriculum Authority) (2006) *A Review of GCSE Coursework*. London: Qualifications and Curriculum Authority.
- Shavelson, R.J. (2013) 'On an approach to testing and modeling competence'. *Educational Psychologist*, 48 (2), 73–86.
- Shavelson, R.J., Ruiz-Primo, M.A. and Wiley, E.W. (1999) 'Note on sources of sampling variability in science performance assessments'. *Journal of Educational Measurement*, 36 (1), 61–71.
- Siddiqui, F.G. (2013) 'Final year MBBS students' perception for observed structured clinical examination'. *Journal of the College of Physicians and Surgeons Pakistan*, 23 (1), 20–4.
- Smith, S., Joslin, H. and Jameson, J. (2015) *Progression of College Students in England to Higher Education (BIS Research Paper 239)*. London: Department for Business, Innovation and Skills.
- SQA (Scottish Qualifications Authority) (2007) *Internal Assessment, External Assessment, and Assessment for Learning: A think piece for those involved in developing assessments (Research Report 14)*. Glasgow: Scottish Qualifications Authority.
- Stasz, C., Hayward, G., Oh, S.-A. and Wright, S. (2004) *Outcomes and Processes in Vocational Learning*. London: Learning and Skills Research Centre.
- Wolf, A. (2011) *Review of Vocational Education: The Wolf Report*. London: Department for Education.