The fairness of internal assessment in the GCSE: the value of students'

accounts

The use of internal assessment within GCSE qualifications has always aroused controversy, primarily because of concerns about the authorship of students' work. This has led to tighter controls and a reduction in its use over the last decade. The paper argues that the focus on the authentication of student work has neglected other aspects of fairness, such as ensuring that assessments elicit students' best performance. Students' accounts can contribute to our understanding of test fairness by showing what affects their performance and how these assessments are enacted within their contexts. The paper draws on a study which surveyed 1600 GCSE students and held focus groups with 128 others in Northern Ireland and Wales, in addition to data from the WISERDEducation multi-cohort study. A number of factors, including the assessment environment, reportedly affected students' performance. According to student comments, the ways in which the assessment guidelines were enacted varied considerably between contexts, with significant fairness implications.

Keywords: internal assessment; controlled assessment; fair testing; GCSEs; qualifications

Introduction

Internal assessment has been included in some form as part of the General Certificate of Secondary Education (GCSE) since it was introduced in 1986. An assessment is defined as being 'internal' when at least one of the processes of setting, taking or marking are controlled by the student's learning institution (Vitello and Williamson, 2017). Coursework, the first form of internal assessment to be introduced, was designed to facilitate the assessment of 'positive achievement' (Tattersall, 1994: 14) and to test skills which could not be assessed with examinations (Elwood, 1999). It was also hoped that it would engender learning benefits, by enhancing communication skills and encouraging students to take responsibility for their own learning (Bullock et al., 2002). This mirrored the introduction of 'performance assessments' in some states in the USA

during the 1990s, and the growth in the use of teacher assessment in Victoria, Australia, and some Nordic countries, including Finland and Sweden (Darling-Hammond & Adamson, 2010). The reasons for the introduction of internal assessments in the UK were varied, but included concerns about impacts of external testing on the curriculum (Harlen, 2005). However, in many US states and in the UK the use of internal assessment was controversial. In the US there were concerns that removing the anonymity of standardised testing would disadvantage children from minority ethnic groups (Baker and O'Neill, 1994). In addition, there was scepticism about the likelihood that comparable administration settings could be implemented for all students (Baker and O'Neill, 1994). However, the main reason for the decline of performance assessment in the US was the growing use of testing for accountability purposes, with the more expensive and time-intensive teacher assessments conflicting with the demands of accountability regimes (Chung Wei, Pecheone & Wilczak, 2014).

In the UK there were particular concerns raised regarding the fairness of the assessments, mainly relating to the authenticity of student work (Tattersall, 1994). There was a popular view among students (Bishop et al., 1999) and teachers (Ipsos Mori, 2006) that there was too much cheating in the GCSE coursework and that students could submit work authored by others. Alongside these, there were also concerns among teachers about how fairly and consistently assessments were administered in different contexts (Ipsos MORI, 2006), leading to calls for 'greater consistency in the interests of fairness and natural justice to candidates' (Colwill, 2007).

As a result, over the last decade, the regulation around internal assessment in GCSEs has been continually tightening, with policy-makers appearing to be engaged in a perennial search for a form of assessment which brings the perceived learning benefits of internal assessment alongside the rigour of examinations. Introduced for first

teaching in September 2009, controlled assessment was the latest step in this series of reforms, as a type of internal assessment which shared many features with examinations. Ofqual defined controlled assessment as 'a form of teacher-supervised internal assessment that aims to encourage a more integrated approach to teaching, learning and assessment, and to enable teachers to confirm that students have carried out the work themselves' (Ofqual, 2013b). Controlled assessments introduced a system of controls, which related to the setting of tasks, the taking of tasks, and the marking of tasks. Controls were set at each stage of the assessment process, at either limited, medium or high (Ofqual, 2013b). While the rules varied for different subjects, for the majority of written controlled assessments, students were permitted to see the task in advance and prepare; however, they were required to undertake the assessment in silence under teacher supervision. While there were also changes to the setting and marking of tasks, this research paper has focused on the changes to the taking of tasks, as students were directly involved in this stage and were able to make informed comments on this aspect of the assessments.

Controlled assessment was introduced in order to make the assessment process fairer by ensuring that everyone had the same amount of time and teacher input. However, research findings suggest that inconsistency in the interpretation of the new guidance meant that some students were more advantaged than others under the system (Crisp and Green, 2013). One of the most significant issues raised by teachers was the difficulty of interpreting the guidance regarding controlled assessment (Ipsos MORI, 2011). Some research has attributed this to a lack of clarity in the regulators' guidelines, leading awarding bodies to interpret them differently, thereby engendering a range of different interpretations at school level (AlphaPlus, 2012). These inconsistencies meant that depending on which guidance was followed, students could sit an assessment in an

examination hall or classroom; be permitted to word-process responses or not; be obliged to complete the entire assessment in one sitting or allowed to undertake it over several different dates (Vowles, 2012).

While many of these issues were related to difficulties in understanding controlled assessment guidelines, there were also concerns that some schools might exploit the situation by purposely interpreting guidance in ways that would advantage their own students. In Ipsos MORI's (2011) research, teachers were worried that some schools might not supervise students as strictly as the guidelines required, and might allow more notes than permitted. AlphaPlus's (2012) study also suggests that teachers were using strategies to maximise performance on assessments. These included holding an assessment with a very similar title prior to the controlled assessment task, so that they could provide feedback to students.

Concerns about the authenticity of student work also remain. Crisp and Green's (2013) research found that teachers tended to think that controlled assessment had reduced the risk of plagiarism, although a fifth of French teachers believed that the risk had increased. Responses to a survey of science teachers in England (Wilson, 2013) found that teachers considered controlled assessment to be problematic as it reduced the time available for practical work and 'encourages cheating' (p. 5). Moreover, in Ipsos MORI's (2011) study, teachers reported that controlled assessment was easier to authenticate than coursework. However, some pointed out that they could not prevent students from simply memorising essays and regurgitating them in the assessment. In addition, 27 per cent of subject teacher respondents to Ofqual's call for evidence on controlled assessment thought that it primarily measured students' skills of recall and memorisation of essays, which has serious implications for the fairness of the assessments (Ofqual, 2013b).

As a result of these issues, the use of internal assessment in the UK has declined over the last five years, with England imposing the most substantial restrictions on its use. Until 2013, England, Northern Ireland and Wales jointly regulated all aspects of GCSEs, with GCSE subjects following the same assessment formats over the three countries. However, since 2013, the end of three country regulation has meant that each region now sets its own rules for national qualifications and the three countries have taken different approaches to the regulation of internal assessment. England has limited the use of internal assessment to a small number of subjects such as Art and Design and Design and Technology, for which it considers it to be necessary, stating that examinations are now 'the default method of assessment' (Ofqual, 2013). The first of the reformed qualifications with less internal assessment in England began teaching in September 2015. It was at this point, when the first new specifications were introduced, that the term 'controlled assessment' began to be replaced by 'non-examination assessment' in England (and later in Wales), although it was not clear that there was much functional difference between them.

In Wales, subject specifications were released more gradually, with some retaining internal assessment and others removing them or reducing their contribution to the overall grade. For example, in 2015 when English Literature and English Language specifications were released for first teaching, the former retained the same proportion of writing internal assessment as it had used previously (25 per cent), whereas for the latter a writing internal assessment task that had previously counted towards 20 per cent of grades was removed (WJEC, 2015a; 2015b). A similar change was made for Modern Languages, which also removed the writing internal assessment task in their 2017 specifications (WJEC, 2017a). Other subjects made smaller changes, such as History,

for which the proportion of internal assessment was reduced from 25 per cent to 20 per cent in 2017 (WJEC, 2017b).

In Northern Ireland, reformed GCSEs were not introduced until 2017. This also saw reductions in controlled assessment, although changes were not as drastic as in England. For example, the proportion of internal assessment in English Language was reduced from 60 per cent to 40 per cent (CCEA, 2017a), and in English Literature from 25 per cent to 20 per cent (CCEA, 2017b). A writing internal assessment that counted towards 30 per cent of the grade was also removed for French GCSE in 2017 (CCEA, 2017c).

Following the introduction of the new specifications in all three countries between 2015 and 2017, concerns about internal assessment remain. In one example in 2018, Ofqual announced that the programming task for GCSE Computer Science would no longer contribute any marks towards the qualification, following concerns about malpractice, as students were found to be using code found online after the task was posted on online forums (Ofqual, 2018). This case highlights the complexity of designing assessments that are fair to all: as Ofqual itself acknowledged, professional programmers often draw on existing coding repositories and adapt them for their own purposes, and so students who were committing 'malpractice' according to the regulations were in fact deploying the same skills as experts in their fields.

The potential contribution of students' perspectives to the policy debates

Despite the impact on students of these changes, and the value of their perspectives, so far such debates have not been informed by their views and experiences (Elwood, 2012; Author & Other, 2018b).. As discussed, the use of internal assessment

in national qualifications in the UK is continually shifting, with a great deal of controversy over how much internal assessment should be used, and what form this assessment should take. While the current trend has been for a reduction in its use, the development of a new curriculum in Wales with an emphasis on formative assessment (Donaldson, 2015) is likely to reignite debates regarding the use of internal assessment within national qualifications. It is essential that students' perspectives are understood before any new assessment reforms are implemented.

The exclusion of students' perspectives from this debate is particularly problematic if we consider assessment policy as being enacted within school and classroom contexts by teachers and students so that students are not passive recipients of policy, but are recognised as policy actors (Ball et al., 2012; Braun et al. 2011; Elwood, 2012; Author & Other, 2018b). Thus, the practices and behaviours they engage in around assessment need to be understood, particularly when changes are made to the 'task-taking' element of assessments, where the ways that students prepare and complete assessments have such an impact. This is not to say that students alone can determine whether or not an assessment is fair, but it is remiss not to consider their perspectives as evidence alongside those of teachers and other actors (Elwood and Lundy, 2012; Author & Other, 2018b). Without research data on young people's views and experiences of the assessments, we have a limited picture of how they are enacted in practice (Author & Other, 2018a). Much current research relates only to teachers' views of students' behaviour around assessments, which, although valuable, are insufficient. This is particularly the case when many of these questions rely on an understanding of students' behaviours and practices outside of the classroom. For example, the question of whether the work students submit is their own could certainly benefit from the input of students, and could provide us with a more nuanced

understanding of the support available to students from different backgrounds at home. Previous research has also shown that students are concerned about teacher bias in the marking of internal assessments (Elwood, 2012). The following section will discuss contemporary models of fair assessment before illustrating further contributions that student views can make to evaluating the fairness of internal assessments based on the literature on fair assessment.

Fair assessment

Standardised examinations have historically been seen as the 'gold standard' of fair testing, as they provide all students with access to the same test under strictly controlled conditions (Mislevy et al., 2013). However, while the authentication of student work is still considered to be a core element of test fairness, the focus of fair testing has shifted to ensuring that assessments provide all students with opportunities to demonstrate their knowledge and skills (Gipps and Murphy, 1994). One key question is: '*is it plausible to believe that with an alternative assessment (test or item) or under different test conditions an individual or groups of individuals may have fared better?*' (Camilli, 2013). This has led to the introduction of alternative types of assessment, such as practical assessment and forms of internal assessment which, together with other more traditional assessments such as examinations, are intended to capture the knowledge and skills of all students (Cole and Zieky, 2001; Gipps and Stobart, 2009).

The concepts of validity and fairness have become intertwined in the assessment literature, and so where this paper uses the term 'fairness' it relates to theories of test fairness which are underpinned by validity. These theories see fairness and validity as being interdependent: so that 'a test has to be fair to be valid' (Xi, 2010). Validity and fairness are seen as similar ways of approaching the question 'Are the proposed

interpretations and uses of test scores appropriate for a population over some range of contexts?' (Kane, 2010: p. 177). Similarly, fairness can also be understood as 'comparable validity for all individuals and groups' (Willingham and Cole, 1997). This perspective on test fairness is rooted in theories of validity which incorporate the consequences of test use into the general definition of validity (Messick, 1990). Thus, test developers have a responsibility to ensure that the instruments they design are appropriate for their intended uses, and that any adverse impact for an individual or group cannot be attributed to 'defects in the test or testing procedures' (Kane, 2013: 56). In particular, construct-irrelevant variance, where the test is too broad and requires test-takers to deploy skills which are not targeted by the assessment in order to access the task (e.g. computer skills), can cause adverse impacts for groups or individuals (Kane, 2013). This may cause them to face unnecessary barriers to displaying their knowledge and skills.

The appropriateness of interpretations of test results can be differentially valid for particular groups, based on factors such as gender, ethnicity and class (Cole and Moss, 1989; Stobart, 2005). Thus, consideration should be given to whether test items contain any contextual information which would be more appealing or familiar to certain groups (Gipps and Murphy, 1994). The language used, the item type or the assessment format should be taken into account to ensure that they are not biased towards or against a specific group (Murphy, 1991). Moreover, the use of test results from different types of assessment techniques and formats such as coursework, examinations and tiering may be differentially valid for particular groups (Elwood, 1999, 2005; Elwood and Lundy, 2010; Stobart et al., 1992).

To date, the majority of the fair testing literature has focused on group differences; however, there is now a growing awareness of fairness issues relating to

individual differences (Cole and Zieky, 2001; Camilli, 2013). It has been argued that statistical information provides more evidence of variation between individuals than between groups, and so there should be more of a focus on the former in fairness analyses (Cole and Zieky, 2001). The focus of individual fairness is the use of standardised conditions wherein 'students are treated comparably' (Camilli, 2013; 105) with scope for accommodations for those who might require more time or support to demonstrate their knowledge and skills.

Furthermore, the fairness of assessments can be affected by the conditions of test-taking (Crooks et al., 1996). The testing environment should be comfortable in terms of physical conditions such as light or temperature, with the aim of ensuring that each student can perform to their best (Wollack and Case, 2016). Test administration procedures should also be consistent across different contexts (Baker and O'Neill, 1994): generally it is thought that each candidate should be permitted the same materials, and allowed the same amount of time for preparation and examination (Kunnan, 2004). However, it has been argued that it would be simplistic to assume that test administration is fair simply because it is standardised (Gipps and Murphy, 1994). For example, students might respond differently to the same instructions or feel more comfortable and relaxed within a particular environment. Thus, understanding a students' experience of an assessment situation is key.

Student views and experiences as evidence of fairness
It has already been recognised that the perspectives of test-takers on the skills necessary to succeed on an assessment can help inform validation processes (Kane, 1992; Michaelides, 2014; Murphy, 1995; Ryan, 2002). Previous research has found that students believed that a broader range of skills was required than what was specified on

the examination syllabus (Xie, 2011). Furthermore, some of this research has argued that student perceptions should be considered in evaluations of the fairness of assessments (Elwood and Lundy, 2012; Michaelides, 2014). Students' views and experiences can make a particular contribution to our understanding of the fairness of the task-taking and preparation stages of internal assessments.

This current study has collected students' views and experiences of internal assessment. Three main themes were identified relating to the fairness of the assessments, as will be discussed further in the methodology section. The themes are:

(1) Fairness of internal assessment in testing appropriate skills

(2) Fairness of internal assessment and authenticity of student work

(3) Fairness of internal assessment in facilitating 'best performance'

The first theme relates to issues of validity and fairness. It considers whether students believe that it is necessary to deploy the skills outlined on the specifications in order to succeed, or whether they use 'gaming strategies' (Stobart, 2008) which often rely on the memorisation of answers written in advance. When we consider the skills that are being used by students, we can compare them with those targeted on the qualification specifications. We can also decide whether we believe that it is fair for further educational and employment opportunities to be distributed based on the skills students use to succeed on these assessments. The second theme, which is closely related to the first, addresses the fairness issue that has driven the most recent reforms – it asks whether teachers can confidently authenticate students' work to ensure that no student succeeds on an assessment with work they have not produced. It also requires us to consider whether students have been given comparable opportunities to succeed.

Students' accounts of how they prepare for and complete the assessments are thus key to addressing this issue.

The third theme relates to whether students believe that they have been given sufficient opportunity to demonstrate their knowledge and skills. Thus, students can tell us whether any other factors affect their attainment, such as lack of adequate preparation by teachers, poor understanding of the test requirements, test anxiety, or problems with the administration of the assessments (Michaelides, 2014). Since GCSE results are used to distribute future employment and education opportunities, it is vital that we understand what factors within the test-taking and preparation stages lead to success, and whether these are in line with what we would expect from a fair assessment.

Methodology

This paper uses data from two research projects. The first is a mixed-methods study which investigated the views and experiences of young people on GCSEs in Northern Ireland and Wales. The second is the WISERD Education project: a longitudinal multi-cohort study which annually surveys secondary school students in Wales.

The data from Study 1 was collected in 2014-15, prior to the introduction of any reformed qualifications in Wales. However, the WISERD Education data was collected in Spring 2017, when some of the new specifications with less internal assessment had been released, but many of the old ones remained. As the new specifications replaced 'controlled assessment' with 'non-examination assessment', both terms would have been in use during the time that the data for the WISERD Education study was collected, although controlled assessment was more established. Thus, I will use the term controlled assessment when referring specifically to data collected in the two

studies, but as controlled assessment is a type of internal assessment, I will use this term when discussing the implications of the findings more generally.

Study 1: Survey and focus groups

The survey and focus group questions were developed and analysed in collaboration with young persons' advisory groups which were established in each country, as a key element of the children's rights approach of the research project. This holds that children have a right to have their views taken into account in decisions that affect them under the United Nations Convention on the Rights of the Child (UNCRC) (Lundy and McEvoy, 2012). The results of national assessments such as GCSEs have a major impact on their future life trajectories, and so students' views on this topic should be considered (Elwood and Lundy, 2012). Thus it is important that they have an input into the research project, to help ensure that the questions are addressing issues that are of concern to young people. It is also argued that involving young people in research in this way can improve the validity of the research findings, since they can help the researcher design questions which their peers would understand, and assist in the interpretation of the data (Lundy and McEvoy, 2012).

The survey included a mix of open-ended and closed questions about respondents' views and experiences of GCSE assessment features such as internal assessment. In the focus groups the issue of fairness was delved into more deeply, with questions about what kind of assessments participants thought were the fairest way of assessing their performance.

Sampling: Survey

Random stratified sampling was used to select the schools for the survey. Differences in the school systems in Northern Ireland and Wales led to the adoption of different sampling strategies in both. Northern Ireland operates a selective school system, with the majority of students either attending a grammar school, a non-grammar school, or an integrated school, which were introduced so that students from Catholic and Protestant backgrounds could be educated together. Thus, the Northern Ireland sample was stratified to include grammar, non-grammar and integrated schools. As there are no grammar schools in Wales, schools were sampled by the proportion of students above and below the national free school meal average (FSM) in Wales. Eight non-grammars and six grammars were recruited in Northern Ireland, and in Wales six schools with above average FSM were recruited, as well as five below. The response rate from contacted schools was approximately 10 per cent.

Sampling: Focus Groups

In Northern Ireland, five non-grammar schools and one grammar school were recruited for the focus groups, whereas in Wales four schools with above average FSM were recruited, and two below. As recruitment was challenging in Wales for this age group (due to the pressures of GCSE preparation), a youth forum was also selected to obtain the necessary number of focus groups.

Participants

Participants in Study 1 were all final year GCSE students, aged 15-16 years of age. In Northern Ireland 699 students took part in the survey. Of these, 379 students from grammar schools participated: 79 girls and 300 boys. 320 students in non-grammar schools also completed the survey, including 139 girls and 181 boys.

In Wales, 901 students participated. 494 students from schools with a below average proportion of students eligible for FSM took part: 225 girls and 269 boys. In the schools with above average proportion of FSM eligible students 199 girls and 208 boys participated: 407 in total.

65 students took part in focus groups for the study in Northern Ireland. This included 13 students in grammar schools: five girls and eight boys. In the non-grammar schools 40 students participated: 20 boys and 20 girls. In addition, 12 students in an integrated school took part: six girls and six boys.
 In Wales 68 students participated in the focus groups. Of these, 62 were from comprehensive schools: 30 girls and 32 boys. Six students, three girls and three boys, from a youth forum also took part.

Data collection

The survey was administered online, except in cases where schools had no computer facilities available and requested print versions of the survey. Infographics on GCSE assessments were integrated into surveys and given to students during the focus groups to ensure that students were able to give informed answers. A number of questions

relating to the topic of the fairness of controlled assessment were asked in the focus group discussions.

Data analysis

Once inputted and cleaned, the quantitative data were analysed using SPSS to identify any significant differences between groups of students. The qualitative data for the project were analysed using the thematic analysis approach outlined by Braun and Clarke (2006). The answers to open-ended questions from the survey and the focus group transcripts were coded using MaxQDA data analysis software. General codes relating to topics identified in the research questions, such as controlled assessment, tiering, course structure and usefulness, were inputted into the coding system at the beginning. This meant that a more theoretical thematic approach was used, as opposed to an inductive one, since many of the codes were linked to the research questions and to the questions asked in the focus groups (Braun and Clarke, 2006).

The advisory groups then assisted the researcher in the next stage of analysing the qualitative data. The groups were given training in coding and thematic analysis, and were then asked to code a sample of quotations to help ensure that the researcher's interpretations of young people's views were valid. They then organised their codes into broad themes, which the researcher drew upon to produce the final thematic map. Three main themes were identified – consequences, choice and fairness. This paper will focus on the data relating to the fairness of internal assessment. The sub-themes relating to this were:

1   Fairness of internal assessment in testing appropriate skills
2   Fairness of internal assessment and the authenticity of student work

3   Fairness of internal assessment in facilitating 'best performance'

The results section will be organised into three parts based on these sub-themes.

Study 2: WISERD Education

Following the completion of Study 1, questions were included in the WISERDEducation study in order to elicit further quantitative data on some of the issues that emerged from the focus groups from Study 1. The study, which has been running since 2012, annually surveys three cohorts of secondary school students within 12 schools in Wales. The schools were selected using random stratified sampling: the sampling frames were stratified according to FSM and urban/rural area. Data were gathered using online surveys and completed on tablets. The data collection was administered by researchers on school premises. In Spring 2017, 331 fourth year GCSE students were given two closed questions on controlled assessment within the broader WISERD Education survey. They asked about how they prepared for controlled assessment, and whether they received support from others with their tasks.

Following completion of the project, the data were analysed using SPSS to produce frequency charts. They were also analysed to see whether there were any statistically significant differences between groups of young people.

Presentation of data

When data are presented below, each quotation has been labelled to indicate whether they are from students in Northern Ireland (NI) or Wales, and whether they derive from the focus group (FG) or the survey. In Northern Ireland and Wales there is a difference

in the way that the school years are numbered – in Wales the first year of secondary school is year 7, whereas this is year 8 in Northern Ireland. In order to avoid confusion the years have been labelled first, second and third and so on, beginning from the year of entry into secondary school at age 11. Data from the WISERD Education study have been clearly labelled as such.

Findings

Fairness of internal assessment in testing appropriate skills

Some students reported that they considered controlled assessment to be a good test of their knowledge and skills. In one of the focus groups in Wales, students discussed how in English 'the controlled assessment actually tests your English skills more than the exam does' (Female, FG, Wales). In Northern Ireland also there was a perception that controlled assessment had the potential to assess 'a wider range of skills for people rather than just exams' (Male, FG, NI). This resonates with research on controlled assessment which has suggested that it can test skills which examinations cannot (Tremain, 2011). Students also suggested that they required a deeper understanding of a topic:

> I think controlled assessments…require more understanding than exams - if I go
> into my exams in the summer having looked over all of my work and memorised it
> off by heart I'd be able to recite to you what everything means and I wouldn't need
> an understanding of it and I can go into my exam write it all down get full marks
> and not have a clue what I've just written…I know if I go to a controlled
> assessment you need to have an understanding of what you're writing about
> (Female, FG, Wales)

Another positive aspect of controlled assessment that was identified by students was its capacity to test more real-life skills. Students in Wales discussed the idea that controlled assessments are 'more like real things you actually do when exams are just testing' (Female, FG, Wales), and in Northern Ireland commented '[c]ontrolled assessments are great in GCSE because they are a lot like what it would be in the real world of work...' (Male, Survey, NI).
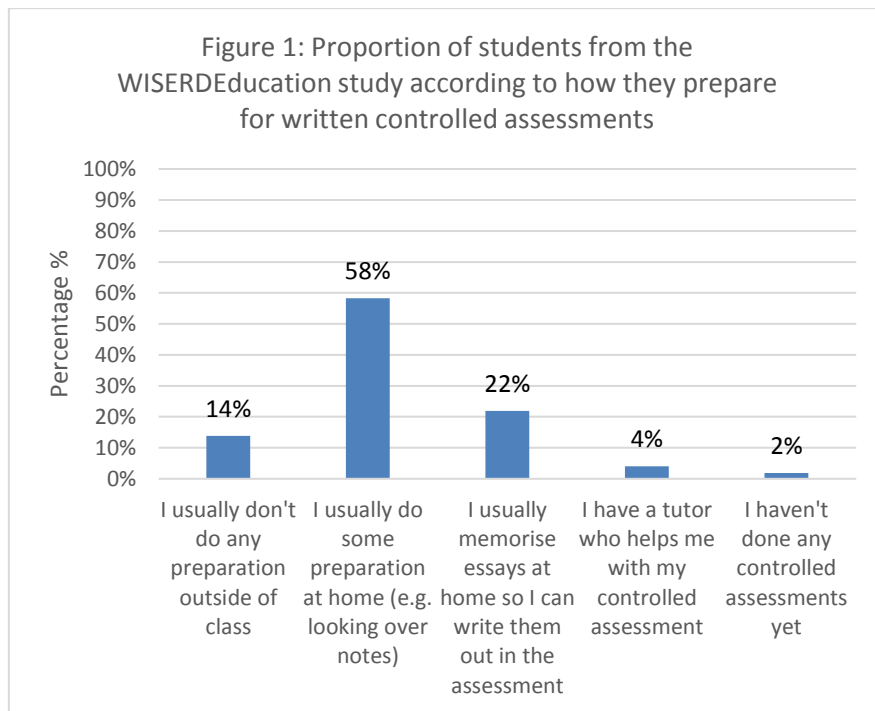
While some students believed that controlled assessments tested more real-life skills than examinations, a number reported that the main skill they deployed on controlled assessment was memorisation. Several of the Welsh focus groups discussed controlled assessment as 'a memory test' (Male, FG, Wales). In Northern Ireland students complained that there was an expectation that they memorise essays:

'it does not test your intelligence or knowledge of the subject - it tests your memory which is unfair for people like me who find it hard to learn controlled assessments off' (Female, Survey, NI).

Controlled assessment in language subjects was considered to be particularly problematic, with students reporting a great deal of rote learning. The views expressed by students in this research corroborate those of teachers in earlier research (Ipsos MORI, 2011). Students often drew on examples from language subjects to illustrate their view that controlled assessment did not facilitate the development of language skills, suggesting instead that they tested their memory:

I don't know what I was writing I went in and I came out with full marks not because I understood what I'm saying because I knew exactly what to write and what order it had to be in' (Female, FG, Wales).

Data from the WISERDEducation study illustrates the variety of approaches taken by students when preparing for and completing written controlled assessment tasks for subjects such as History and English (which are generally essay questions), as can be seen in Figure 1 below. Students could choose two options out of the five outlined in Figure 1 below.

Figure 1: Proportion of students from the WISERDEducation study according to how they prepare for written controlled assessments

Response	Percentage
I usually don't do any preparation outside of class	14%
I usually do some preparation at home (e.g. looking over notes)	58%
I usually memorise essays at home so I can write them out in the assessment	22%
I have a tutor who helps me with my controlled assessment	4%
I haven't done any controlled assessments yet	2%

As we can see, the majority of students did some form of preparation at home, such as scanning notes. However, over a fifth of students reported that they memorised essays and wrote them out in the assessment, which is problematic as it would mean that for these students the assessment is at least partly assessing memorisation, which is not amongst the higher order skills which controlled assessment is designed to assess. These findings around the skills deployed by students for controlled assessment reinforce the findings of Ipsos MORI's (2011) research, which found that over a quarter of teachers they surveyed believed that some of their students memorised essays for the controlled assessment.

Fairness of internal assessment and the authenticity of student work

The focus group discussions indicated that participants were concerned about the variety of support available to students at home. Participants suggested that the knowledge and education level of parents had a considerable impact on students' performance, as while some parents might be able to provide academic support: 'some people's parents might not be able to help them at all' (Male, FG, NI). Students could also develop strategies allowing them to capitalise on the support available to them at home:

> '[p]eople can still get help at home — it's not testing their own ability — people get help at home and just learn it and write it up...' (Female, Survey, NI).

In the two Welsh schools that had the lowest levels of students eligible for free school meals from the sample, participants discussed how they benefitted from the support they received at home:

> S1: my nan speaks fluent French, which helps for French controlled assessment if I'm gonna be honest I would have done nowhere near as well as I did without her
> S2: my mum's a science teacher so...
> S3: it helps if you've got a mum who's a teacher
> S4: that's the thing when you know what you gotta do you can find extra sources outside to help you kinda do well in it so it's not really all about you ...it's about who you know and what you can do with it (FG, Wales)

In order to probe this further, we asked students in the 2017 sweep of the WISERDEducation study about how often their parents helped them with controlled assessment. Of these participants, 32 per cent said their parents 'always' helped them, 40 per cent said they did 'sometimes', and 28 per cent said they 'never' helped them.

The findings suggest that over 70 per cent of students received some form of support from their parents or carers when preparing for their controlled assessment, so the notion that controlled assessments remove the advantages associated with coursework is problematic. This is particularly the case as we know that some students are memorising entire essays in preparation for controlled assessments (see Figure 1) and so the authorship of the students' essays is not clear. We know that the advantages provided by controlled assessment cannot be as great as those conferred by coursework, where it was possible that students were submitting work produced entirely by others. Having to write out the essay in a controlled assessment at least ensures that the student has to learn the structure of the work, and deploy correct grammar and spelling to attain high marks (particularly in an English language or literature assessment). While these points must be acknowledged, it is clear that there is a perception among students that tutors can provide a significant advantage for students:

> S1: and then you've obviously got tutors and that then you cheat a bit [laughter]
> S2: tutors can just write the whole controlled assessment for them (FG, Wales).

> people have tutors outside of school who can still help them with things similar to their controlled assessment so they have more help than others so it's not as fair as it should really be (Female, FG, Wales)

Thus, in some cases the support offered outside of school may go beyond providing advice and developing understanding. The findings of this study suggest that some students share the concerns about authorship voiced by teachers in Ipsos MORI's (2011) review of controlled assessment .

There was also concern about the type and amount of support provided by teachers. Students stated that teachers had an inordinate influence upon the performance of their

students. There was a sense that rules were not always adhered to and that 'some people may get more help than others, which I see as unfair' (Male, Survey, NI). In one example in Northern Ireland, a student commented that:

> 'in different classes how the controlled assessment's carried out like varies a lot because some teachers will like help you more than others' (Female, FG, NI).

In addition, there was a perception that results do not depend so much on 'a pupil's knowledge or intelligence but the preparation a pupil has gone through or the teacher someone has' (Male, Survey, Wales). Students expressed anxiety about the level of preparation:

> S1: ...you feel kind of disadvantaged cause in another class it's easier for them...the teachers guide them through it and...basically given them a good mark whereas your teacher is being strict to the rules and saying you have to do this and learn it and you've put a lot more work in....
> S2: yeah like a lot of people say teachers don't make a big difference – they make a huge difference ... (FG, NI)

This was a source of concern for students who considered themselves unfairly disadvantaged by the amount of support provided by teachers. These findings suggest that the individual context could have a considerable impact on the way in which assessment guidelines are enacted in practice (Ball et al., 2012), and that this advantage may not be restricted to a school level, but to a class level as well.

The influence of teachers was not restricted to the amount of support provided at the preparation stage. Students provided numerous examples of the variety of practices used by teachers when administering controlled assessments:

> some classes get an unfair advantage if they have a less strict teacher, e.g. are allowed to bring in prohibited research (Male, Survey, Wales).

> some teachers do controlled assessments under different conditions, e.g. helping students when they're stuck, or under exam conditions' (Female, Survey, Wales).

23

Likewise in another school setting, students expressed the opinion that teachers were 'quite lenient' during controlled assessment and that 'you can kind of cheat [because] it's not regulated properly' (Female, FG, Wales). In Ipsos MORI's (2011) research many teachers were also concerned that other schools might be interpreting controlled assessment guidance more liberally than theirs. Generally, however, the report found that teacher suggestions for improving controlled assessment tended to relate more to the rules and guidelines put in place by regulators. However, students argued that 'it's not... the rules that are wrong' (Male, FG, Wales), but that schools are not 'enforcing them' (Female, FG, Wales).

Students recalled numerous examples of practices which did not adhere to the regulations for controlled assessment. During a focus group discussion at one school, a student reported that 'we've done one of our controlled assessments three times in our class just so everyone can bump up the marks' (Male, FG, Wales). The subject the student referred to was English language, for which redrafting is not permitted (Ofqual, 2013a). Similarly, a Welsh student reported that her teacher allowed her to take her English controlled assessment essay home with her and then assisted her in school on a 'one-to-one basis' to ensure that she completed it (Female, FG, Wales).

One of the most common examples of inconsistency in administration was related to timing. Several students protested that they were not given an equal amount of time to complete controlled assessments. For example, a student from Wales argued that:

> [n]ot everyone gets the same amount of time on controlled assessments in this school.
> Two geography classes doing the same course, and one class are getting longer...
> (Male, Survey, Wales)

Whereas the controls introduced to internal assessment in 2009 were designed to standardise assessment procedures and ensure the fairness of the assessments, the results of this study suggest that a considerable amount of variance occurs.

Fairness of internal assessment in facilitating 'best performance'

Students' responses identified a number of factors that affected their performance which were not relevant to the knowledge and skills targeted on the assessments. For many students, the format of the assessment and the feelings of stress it engendered had a strong influence on their performance:

> Some people cannot perform well under the pressure of exams. This will affect the grades they will get which is unfair as they could be better than what their grades say about them. (Female, Survey, Wales).

In particular, there was a strong sense that a number of students struggled to perform to their best on examinations under the pressure of the examination hall:

> Some people are better at doing things not under pressure in the hall, when under pressure don't perform their best. (Male, Survey, NI)

Students' accounts of struggling to perform in the pressurised setting of the examination hall reflect the findings of Putwain (2008; 2009) whose research showed that test anxiety impacted on the attainment of GCSE students. For students, the more 'relaxed environment' of the classroom in which controlled assessment was completed, contrasted with the stress of the examination hall, echoing Gipps's (1994) assertion that 'a normal classroom setting, which is not unduly threatening' (p.287) might have a positive impact on students' performance in assessments. Students believed that they were more likely to fulfil their potential in this setting as the assessment:

'isn't done under the stress of an exam room which is why I think people will perform better in these situations which will reflect their true ability' (Male, Survey, Wales).

Thus, the mental and emotional state engendered by the assessment environment was key for students. Students wanted a space without distractions or excessive pressure, and for most the classroom represented this better than the examination hall or even the home:

'it's done in class rather than at home where I wouldn't concentrate as easily in' (Male, Survey, NI).

However, other students had negative experiences of how the classroom environment affected their performance in controlled assessments. Some students reported that they found it difficult to concentrate during controlled assessments because they were distracted by the behaviour of other students:

S1: sometimes it's not as fair cause sometimes like the class is being really loud and then you can't like do what you want because the class is the rest the class is kind of
S2: disruption (FG, NI)

In this case it appears that the familiarity of the classroom can create an informal atmosphere which is not conducive to work. This can have a negative effect on the concentration and behaviour of students, with consequences for their peers. Students in three Welsh focus groups also complained about disruptive behaviour during their assessments, with one student reporting that: 'they played hell in my class' (Female, FG, Wales). While further research may be needed to confirm this, these findings indicate that it is more difficult for teachers to enforce the assessment regulations in the classroom than in an examination hall.

These examples emphasise the centrality of the school and classroom context for young people's experience of assessment and suggest that the disadvantages

experienced by some are not restricted to the preparation phase: they infringe upon the assessment itself. They also raise questions about the fairness of the assessments, since it is clear that some students are faced with additional challenges to demonstrating their knowledge and skills. Crooks et al. (1996) highlight the threat to the valid use of assessments caused by 'inappropriate assessment conditions' (p. 271). In this case, it is clear that attainment can be affected by the conditions of the classroom.

Discussion

Controlled assessment was designed to fulfil numerous facets of test fairness, ensuring that work could be confidently authenticated, whilst providing opportunities for students who struggle with exams to evidence their attainment in alternative ways. It also aimed to test a range of skills which could not be assessed through examinations. The results of this research suggest that internal assessment can enable some students who find examinations challenging to perform. However, it has also found that students' experiences of controlled assessment vary considerably depending on their school, classroom and home contexts. Student accounts show that internal assessments raise numerous questions in relation to fairness, in terms of providing students with comparable opportunities to deliver their best performance, authenticating student work, and ensuring that appropriate skills are being assessed.

Authenticating student work and testing appropriate skills

Students' accounts suggest that they and their teachers have developed 'gaming strategies' (Stobart, 2008) which resist the restrictions placed upon them by controlled assessment. The fact that students are developing such strategies suggests that the enactment of policies does not only vary according to school or class. As policy actors themselves (Elwood, 2012; Author & Other, 2018b) students enact these policies in

diverse ways and draw upon the resources available to them outside of school to maximise their attainment on controlled assessments. The skills deployed on these tests may not always be in line with those specified on the marking criteria for subjects like English, which target higher-order skills.

It is important to recognise that there will always be a variation in the amount of assistance available to students at home when preparing for an assessment; this is why it is not possible to create a completely fair test (Gipps, 1994). However, the key question in this case is whether this type of assessment permits students to submit work authored by others. This is a very important question, since the use of classroom time (which could otherwise be allocated to teaching and learning) for completing controlled assessment was largely justified on the basis that it enabled teachers to confidently authenticate students' work (Colwill, 2007). The findings suggest that placing additional controls on internal assessment may not have entirely eliminated this problem, as students believe that those with access to greater assistance from parents and tutors report being able to complete these qualifications by memorising work produced by parents or tutors. Thus, in an attempt to make assessments more rigorous, the aim of testing higher order skills has been jeopardised (Vowles, 2012).

This illustrates the difficulty of balancing different aspects of fairness within an assessment system: it is not always possible to promote one element of fairness without affecting another. Therefore, policy-makers have to make difficult decisions regarding their priorities for an assessment system. These decisions must be informed by evidence from all stakeholders.

Eliciting best performance

One element of fairness which has been under-represented in the recent discussion regarding reform is the importance of designing assessments which elicit students' best performance, a crucial element of fair testing (Gipps, 1994; Stobart, 2008). Students' accounts have made a particularly valuable contribution to this aspect of fair testing, with their perspectives offering insights on the factors which encourage and inhibit their attainment. One of the most important factors for students was the assessment environment. Students tended to evaluate assessments according to the mind-set they engendered. Not only was stress seen as an unpleasant side-effect, it was also central to their conception of a fair assessment in that a more stressful experience affected their attainment, reflecting the research of Putwain (2008; 2009) on the impact of test anxiety on performance. As a result, most participants spoke positively about the more relaxed classroom environment, contrasting this with the more pressurised setting of the examination hall.

However, student accounts' showed that the 'classroom setting' was not homogenous across schools: it was conceptualised in different ways depending on the practices of teachers and the behaviour of other students, reminding us of the need to 'take context seriously' (Braun et al. 2012). While for most students, the classroom atmosphere was relaxed, for others it was a disruptive environment within which they struggled to concentrate. This is clearly problematic in terms of fairness, since students in these settings face additional challenges to delivering their best performances.

While the assessment environment is a recognised element of test fairness (Crooks et al., 1996; Camilli, 2013), the emphasis on it in students' accounts suggests it may have a stronger impact on performance than may have been previously recognised, and it should be given greater weight in considerations of test fairness. Internal assessments can be enacted in ways that may not be anticipated by politicians,

regulators and examination boards. While 'policy-makers tend to assume 'best possible environments' for implementation' (Braun et al., 2012), the reality is often quite different. Thus, evidence from test-takers on assessment administration and environment should be used within the process of evaluation, and should be considered in any further policy decisions around testing and assessment.

The ways in which the assessment guidelines and regulations interact with the school and classroom context must be considered (Author & Other, 2018a). It is only by asking young people and their teachers about their experiences that we can begin to understand how these assessments are enacted in their contexts. Thus, the findings reinforce the argument that test-takers' experiences should be included in test validation processes (Michaelides, 2014; Ryan, 2002), and suggest that they can make particularly valuable contributions to evaluations of the fairness of internal assessments. As internal assessments are, by definition, less tightly controlled than examinations, they can lead to greater variation in practices between contexts. Students' perspectives help us understand which aspects of fairness are supported by internal assessments, and how these compare with examinations. Their accounts of the assessment preparation and taking processes can help us understand what kinds of skills they deploy and what may affect their performance, all of which have the potential to inform discussions of test fairness.

Conclusion

This study has identified numerous disadvantages associated with controlled assessment, most notably that some students have access to extensive support at home, while others have little or none, and that students can succeed on the assessments using

skills of memorisation. However, it is important to recognise that students may also depend on memorisation in examinations, and that students from more affluent and advantaged backgrounds can also be schooled in examination techniques by tutors and highly-educated parents. We do not yet know whether controlled assessment provides a greater advantage for these students than an examination-only system, and this is a topic that requires further research, including a comparison of the performance of students from different socio-economic groups on examinations and controlled assessments. Until we have such evidence, students' accounts suggest that the priority should be for schools and teachers to pay greater attention to the standardisation of administration procedures, and to the potential impact of the assessment environment on students' performance.

References

AlphaPlus. (2012). The evaluation of the impact of changes to A levels and GCSEs. (DFE-RR203). Retrieved from Gov.uk website: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/197437/DFE-RB203_1_.pdf

Author & Other (2018a). Details removed for peer review

Author & Other (2018b). Details removed for peer review

Bishop, K., Bullock, K., Martin, S., & Thompson, J. (1999). Users' perceptions of the GCSE. *Educational Research, 41*(1), 35-49. doi:10.1080/0013188990410104

Bullock, K., Bishop, K., Martin, S., & Reid, A. (2002). Learning from coursework in English and geography. *Cambridge Journal of Education, 32*(3), 325-340. doi:10.1080/0305764022000024195a

Camilli, G. (2006). Test fairness. In R. Brennan (Ed.), Educational measurement (pp. 221--256). Westport: New York, NY: American Council on Education & Praeger publishers.

Camilli, G. (2013). Ongoing issues in test fairness. *Educational Research and Evaluation, 19*(2-3), 104-120. doi:10.1080/13803611.2013.767602

CCEA (2017a). *GCSE Specification: English Language.* Retrieved from CCEA:

https://www.rewardinglearning.org.uk/common/includes/microsite_doc_link.aspx?docid=21001-1

CCEA (2017b). *GCSE Specification: English Literature.* Retrieved from CCEA:

http://ccea.org.uk/english/

CCEA (2017c). *GCSE Specification: French.* Retrieved from CCEA:

https://www.rewardinglearning.org.uk/common/includes/microsite_doc_link.aspx?docid=20920-2

Chung Wei, R., Pecheone, R. & Wilczak, K. (2014) *Performance Assessment 2.0. Lessons from Large-Scale Policy & Practice.* Retrieved from Stanford Center for Assessment, Learning, and Equity website: https://scale.stanford.edu/sites/default/files/Performance%20Assessment%202.0%20Final_Digital.pdf

Cole, N. S., & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), *Educational Measurement* (pp. 201-220). London: Macmillan.

Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. Journal of Educational Measurement, 38(4), 369-382. doi:10.1111/j.1745-3984.2001.tb01132.x

Colwill, I. (2007). *Improving GCSE: Internal and controlled assessment recommendations on the nature of controls needed for internal and controlled assessment in future GCSE specifications* (QCA/07/3207). Retrieved from CORE website: https://core.ac.uk/download/pdf/4156733.pdf

Crisp, V., & Green, S. (2013). Teacher views on the effects of the change from coursework to controlled assessment in GCSEs. *Educational Research and Evaluation, 19*(8), 680-699. doi:10.1080/13803611.2013.840244

Crooks, T., Kane, M., & Cohen, A. (1996). Threats to the valid use of assessment. *Assessment in Education, 3*(3), 265-286.

Darling-Hammond, L. & Adamson, F. (2010) *Beyond Basic Skills: The Role of Performance Assessment in Achieving 21st Century Standards of Learning.*. Retrieved from Stanford Center for Opportunity Policy in Education website: https://edpolicy.stanford.edu/sites/default/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-report_0.pdf

Elwood, J., & Lundy, L. (2010). Revisioning assessment through a children's rights approach: Implications for policy, process and practice. *Research Papers in Education, 25*(3), 335-353.

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), 321-344. doi:10.1076/edre.5.4.321.6937

Elwood, J. (2012). Qualifications, examinations and assessment: Views and perspectives of students in the 14–19 phase on policy and practice. *Cambridge Journal of Education, 42*(4), 497-512. doi:10.1080/0305764X.2012.733347

Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*.
    Buckingham: Open University Press.

Gipps, C., & Stobart, G. (2009). Fairness in assessment. In C. Wyatt-Smith, & J. J.
    Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory
    and practice* (pp. 105-118). Dordrecht: Springer Netherlands. doi:10.1007/978-
    1-4020-9964-9_6

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), 245-270. doi:
    10.1080/02671520500193744

Ipsos MORI (2006). *Teachers' views on GCSE coursework: Research study conducted
    for the QCA*. Retrieved from the Gov.uk website:
    https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6
    05733/0506_QCA_teachers-views-coursework.pdf

Ipsos Mori (2011). *Evaluation of the introduction of controlled assessment: Report on
    qualitative and quantitative research* (Ofqual/11/5049). Retrieved from Gov.uk
    website:
    https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6
    05706/1011_IpsosMori_final_evaluation_of_the_introduction_of_controlled_as
    sessment.pdf

Ipsos MORI (2012). *Controlled assessment in modern foreign languages: Final report*
    (Ofqual/12/5182). Retrieved from:
    https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6
    05751/Controlled_Assessment_in_MFL.pdf

Kunnan, A. J. (2004). Test fairness. In M. Milanovic, & C. Weir (Eds.), *European language testing in a global context* (pp. 27-48). Cambridge: Cambridge University Press.

Lundy, L., & McEvoy, L. (2012). Childhood, the United Nations Convention on the Rights of the Child and research: What constitutes a 'rights-based' approach? . In M. Freeman (Ed.), *Law and childhood* (pp. 75-91). Oxford University Press: Oxford.

Messick, S. (1990). *Validity of test interpretation and use.* Princeton: Educational Testing Service.

Messick, S. (1996). *Validity and washback in language testing*. Princeton: Educational Testing Service.

Michaelides, M. P. (2014). Validity considerations ensuing from examinees' perceptions about high-stakes national examinations in Cyprus. *Assessment in Education: Principles, Policy & Practice, 21*(4), 427-441. doi:10.1080/0969594X.2014.916655

Mislevy, R. J., Haertel, G., Cheng, B. H., Ructtinger, L., DeBarger, A., Murray, E., . . . Vendlinski, T. (2013). A "conditional" sense of fairness in assessment. *Educational Research and Evaluation, 19*(2-3), 121-140. doi:10.1080/13803611.2013.767614

Murphy, P. (1995). Sources of inequity: Understanding students' responses to assessment. *Assessment in Education: Principles, Policy & Practice, 2*(3), 249-270. doi:10.1080/0969595950020302

Ofqual. (2013a). *Design details of new GCSEs in England.* Retrieved from National Archives website: http://webarchive.nationalarchives.gov.uk/20141031163546/http://ofqual.gov.uk /news/design-details-of-new-GCSEs-in-England/

Ofqual. (2013b). *Review of controlled assessment in GCSEs*. Retrieved from Gov.uk
    website: https://www.gov.uk/government/publications/review-of-controlled-
    assessment-in-gcses (Accessed: 8th June 2016).

Ofqual (2018). *Future assessment arrangements for GCSE (9 to 1) computer science.*

Retrieved from gov.uk:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment
_data/file/771555/Consultation_Future_assessment_arrangements_for_GCSE_computer
_science_Ofqual_18_6439.pdf

Stobart, G., Elwood, J. and Quinlan, M. (1992) 'Gender bias in examinations: how

equal are the opportunities?', *British Educational Research Journal*, 18(3), pp. 261-276,

doi:10.1080/0141192920180304

Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in
    Education: Principles, Policy & Practice,* 12(3), 275-287.
    doi:10.1080/09695940500337249

Stobart, G. (2008) *Testing times: the uses and abuses of assessment*. London:
    Routledge.

Tattersall, K. (1994). The role and functions of public examinations. *Assessment in
    Education: Principles, Policy & Practice*, 1(3), 293-304.
    doi:10.1080/0969594940010305

Tremain, K. (2011). *Findings from focus groups: Is controlled assessment working?*
    Manchester, UK: AQA Centre for Education Research and Policy.

Vitello, S. & 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. doi:10.18546/LRE.15.3.14

Vowles, C. G. (2012). Variability in GCSE controlled assessments subject to high levels of control: Ipsos MORI's 'Evaluation of the introduction of controlled assessment' (2011) and its implications for controlled assessments in English literature. *Changing English, 19*(4), 439-449. doi:10.1080/1358684X.2012.736749

Wilson, F. (2013). *A study of stakeholders' views on science courses for 16 year olds*. Cambridge Assessment.

WJEC (2015). *GCSE Specification: English Language.* Retrieved from WJEC: https://www.wjec.co.uk/qualifications/english/r-english-language-gcse-wales/WJEC%20GCSE%20English%20Language%20Specification%202015%20(24-10-14)%20-%20Branded.pdf.pdf

WJEC (2015). *GCSE Specification: English Literature.* Retrieved from WJEC: https://www.wjec.co.uk/qualifications/english/english-literature-gcse/WJEC%20GCSE%20English%20Literature%20-%20Specification%20(For%20teaching%20from%202015%20-%20WALES%20ONLY).pdf

WJEC (2016). *GCSE Specification: French.* Retrieved from WJEC:

https://www.wjec.co.uk/qualifications/french/r-french-gcse-from-2016/wjec-gcse-french-spec-from-2016-e.pdf?language_id=1

WJEC (2017). *GCSE Specification: History.* Retrieved from WJEC: https://www.wjec.co.uk/wjec-gcse-history-spec-from-2017-e.pdf

Qin Xie (2011). Is Test Taker Perception of Assessment Related to Construct Validity?, *International Journal of Testing, 11*(4), 324-348, doi: 10.1080/15305058.2011.589018

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27 (2), 147-170. doi: 10.1177/0265532209349465