Oxford University Centre
for Educational Assessment

UNIVERSITY OF
OXFORD

Research evidence relating to proposals for reform of the GCSE

Jo-Anne Baird, Ayesha Ahmed,
Therese Hopfenbeck, Carol Brown
and Victoria Elliott

Oxford University Centre for Educational Assessment Report OUCEA/13/1

Contents

Overview

Overview

The September 2012 consultation document, Reforming Key Stage 4 Qualifications (DfE, 2012a), set out the government's proposals for changes to the examination system at age 16. The Secretary of State said in the House of Commons that these reforms were radical,[1] and others commented that they composed the biggest change in the examination system in a generation. This warrants an analysis of the research literature relevant to the proposals. We address each of the following issues raised in the consultation document in turn, and present research evidence in each case.

- How England competes in international test scores
- The claim that there has been grade inflation
- The proposals to raise the level of challenge in examinations
- Familiarity with examination materials
- The desire to increase students' motivation
- The plan for students only to be tested at the end of the course, rather than within a modular system
- The proposal that the use of controlled assessment should be restricted
- The proposal that examinations should not be tiered

On 7 February 2013, the Secretary of State announced the government's position in response to the consultation.[2] Contrary to the original proposals, rather than the 'English Baccalaureate Certificate', the new qualifications will continue to be named GCSEs and there will be multiple examination boards for each subject. The details of the reforms have yet to unfold, but the matters listed above appear to remain pertinent to the current reform of GCSE and are, of course, more broadly relevant to examination reforms in general. Interest in many of these issues is perennial. We summarise the findings of this report overleaf.

[1] Hansard, 17 September 2012, column 655.

[2] Hansard, 7 February 2013, column 441.

1. International test scores for England show no decline

a) England's scores on the Programme for International Student Assessment (PISA) are generally close to the Organisation for Economic Co-operation and Development (OECD) countries' average. England also has an average level of spending on education compared with OECD or EU countries.

b) The Trends in Mathematics and Science Study (TIMSS) shows no decline in England's test scores in the four yearly cycles of testing (1995, 1999, 2003, 2007, 2011).

c) England is among the top ten countries for TIMSS 2011 for mathematics (year 5 and year 9 pupils) and for science (year 9 pupils). For year 5 science pupils, England is among the top fifteen countries.

d) TIMSS data show that England has a wide spread between the lowest and highest achievers compared with other countries.

2. Evidence for grade inflation is mixed

a) Good data on the causes of rises over time in GCSE grades are not available. Some of the possible causes, like better teaching and more student effort, could be legitimate, but others, such as easier examinations and teaching to the test, would undermine the currency of the examinations.

b) There is mixed evidence about whether IQ test scores themselves are rising for the population in England. As with examinations, whether IQ tests have become easier is also debated in the literature.

c) There is some evidence that pupils with the same scores on general ability tests got higher GCSE grades in 2010 compared with 1996. This could be caused by falling examination standards, improvements in educational standards or a mixture of each.

d) Overall, research evidence does not point to a general pattern of decline in cognitive demand of examination questions.

3. Raising standards

a) Raising the level of demand in examinations will not in itself raise standards of achievement.

b) A sizeable proportion of pupils (41%) currently do not gain five A*-C grades at GCSE including English and mathematics. If demand is increased, and this causes an increase in difficulty, then there are likely to be more students failing to reach the five A*-C target.

c) Reducing the attainment gap between lowest and highest performers is one way in which countries have improved their overall performances on international surveys.

d) Access to past papers and mark schemes ensures that students will be familiar with the format of the assessments, so that examination demands can focus upon what is understood about the subject. However, a balance must be sought to avoid teaching to the test.

e) Self-assessment is important for effective learning and should, therefore, offer a way to raise standards in our classrooms. Understanding examination standards is part of students' self-assessment skills.

4. Knowing what is expected

a) Without access to examination materials, students have to guess what is required of them. What is tested is then less relevant to the subjects of interest, as good guesswork comes into play.

b) Ofqual concluded that unpredictable assessments are just as poor as over-predictable assessments.

5. Aspiration and motivation

a) Most young people in England have high aspirations.

b) A fixed view of intelligence is unhelpful in motivating students to do well. Research shows that emphasizing the role of effort rather than natural ability improves performance.

c) Proposals for the reforms to GCSE present a combination of policies that are jointly unlikely to foster a growth mindset: raising demands, fewer opportunities to try again, and fewer routes to success in the form of different subjects or qualifications.

6. Research on modular examinations

a) Modular assessment has not been found to be consistently easier than end of course examinations.

b) Appropriateness of modular assessment could vary by subject.

c) Students find the feedback from modular examination results useful.

7. End of course assessments

a) High stakes end of course examinations have been found to produce backwash effects on teaching and learning. These include narrowing of the curriculum and drilling of students. Additionally, rote learning rather than broad and deep approaches has been fostered by raising the stakes of tests.

b) Anxiety associated with high stakes single examinations could mean that we do not get a true picture of students' knowledge, understanding and capabilities.

c) Public examinations involve a range of assessment formats and styles. At school level, skills, such as those required to do practical work in science, are important for progression in and beyond education, and require assessment formats that go beyond the standard end-of-course written examinations.

8. Teacher assessment

a) Assessment for learning practices have been adopted by governments internationally, due to research findings showing that the principles can have a large impact upon learning.

b) Concerns about teacher assessment include unreliability, bias and cheating. Systems have been implemented in high stakes assessments to address these issues.

9. Assessing the whole ability range

a) A great deal of research has been conducted on methods of assessing the whole ability range, both prior to the introduction of GCSE and later. Possible approaches include:

i. Tiering
ii. Multiple papers
iii. General and extended papers
iv. Computer adapted testing
v. Multistage testing
vi. Separate modules

b) The research on tiering at GCSE indicates that it is considered more appropriate in some subjects than others and that entry decisions can be unfair to some students.

Research evidence

Research evidence

This report summarises research evidence relating to the Department for Education proposals for English Baccalaureate Certificates (DfE, 2012a).

We principally focus upon research on GCSEs or research conducted in the UK, with occasional references to research from other countries where appropriate. Many of the issues raised are central to assessment debates, wherever and whenever they occur, and the review therefore goes beyond the current contexts of country, time, policy and stage of assessment in a number of places.

As the report is intended for a broad audience, we present key texts rather than every study. We have also favoured empirical research evidence articles over critical evaluations, and peer-reviewed journal articles rather than the grey literature. To present a balanced picture, we have indicated dissension in the literature in places.

The particular issues mentioned in the consultation document (DfE, 2012a) that we address are those relating to: international comparisons, grade inflation, raising demands in order to raise standards, teaching to the test and familiarity with examination materials, students' motivation, modular and end of course assessments, assessment of deep and broad understanding, teacher assessment and differentiation across the whole cohort.

We have focused on these as the core issues and we present research evidence that relates to each of these. We do not discuss the more administrative issues of: accountability systems, the changes to the marketplace of Awarding Organisations, the title of the qualifications, the implementation period, grading structures, study post-16, or Statements of Achievement. The government has decided not to change the names of the qualifications or to franchise the provision of GCSEs in core subjects.[3]

In the preparation of this report, it became clear that we could not rely solely upon articles published in peer-reviewed journals.

Much of the important work that was closely related to current assessment policy issues had been produced by examination boards or government agencies. Also apparent was the fact that many of these policy issues had been studied in the past, in different societal, political and policy contexts.

Examination policy is underpinned by political values, and tensions play out through assessment reform cyclically. A major tension running through the issues that arise in this report is that between the inclusion and selection agendas. For example, do rising grades indicate higher levels of achievement for all, or do they signal devaluation of educational standards?

Science has a role to play here. However, we have found that the relevant research, although often meticulously conducted, was sometimes lacking to address the underlying issues. Instead, it may be directed at the questions of the moment, with answers satisfying the immediate needs of those who have sponsored the work. Without such research, even short-term questions would involve decision-making in an evidence vacuum, so the examination boards and government agencies are providing a public good by conducting this work. There are also examples of research going beyond this general pattern. As a field though, there is a need to address the fundamental questions, lest we do education the disservice of failing to build knowledge that moves the wider societal debates forward. Our purpose here is to contribute to this wider agenda by summarising, however briefly, some relevant research evidence.

The authors of this report are committed to raising educational standards. The purpose of this report is to investigate the evidence base for the current proposals, with a view to ensuring that decisions taken have the desired impact. To plan the way ahead, we need to be clear about not only the destination, but also where we are now.

[3] Ibid.

Section 1: International test scores show no decline

England is average in PISA

England's PISA scores are, in fact, close to the OECD mean (OECD, 2010e) in both reading (rank 20th) and mathematics (rank 22nd), but above average in science (rank 11th). This is commensurate with the average level of economic investment in England (OECD, 2010a). We appeared to perform well in the first two rounds of PISA surveys, but the UK Statistics Authority (UKSA, 2012) referred to the following caveat in the PISA 2009 OECD report for the UK, writing that,

"As the PISA 2000 and PISA 2003 samples for the United Kingdom did not meet the PISA response-rate standards, no trend comparisons are possible with these years."

OECD (2010f, p.1)

The Trends in Mathematics and Science Study (TIMSS), shows no decline in England's test scores

England is placed in the top ten for mathematics for year 5 and year 9 pupils in TIMSS 2011, and in the top ten for science for year 9 pupils. For year 5 pupils, England is in the top 15 countries for science (see Figure 1 to Figure 4). Sturman et al. (2012) tell us that England's mathematics performance has remained stable since 2007, and that year 9 pupils are about average, while in science, England is above the international average. None of the participating European countries performed significantly better than England.

Trends in TIMSS

Figure 1. Year 5 mathematics

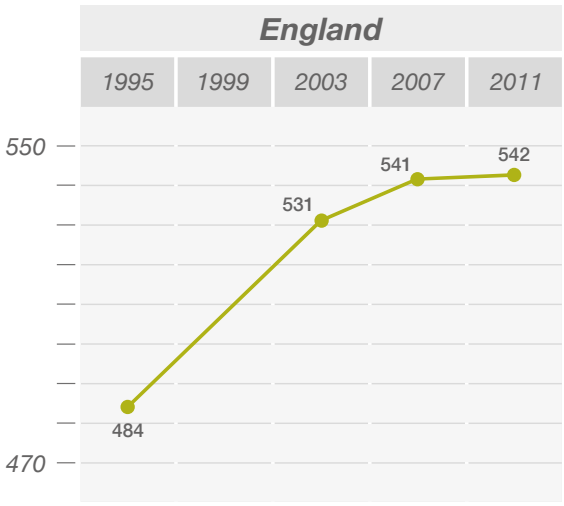
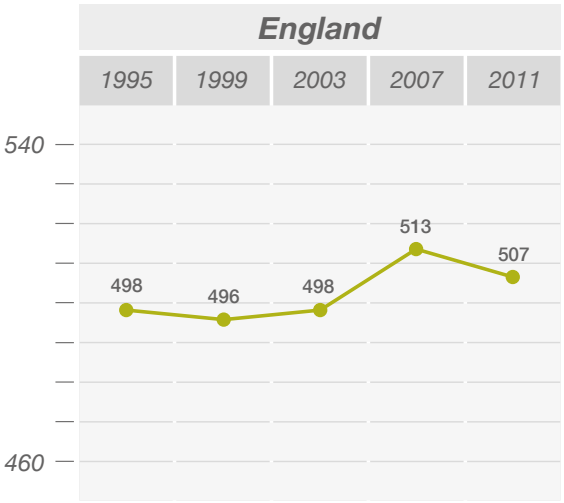


Figure 2. Year 9 mathematics



Trends in TIMSS

Figure 3. Year 5 science

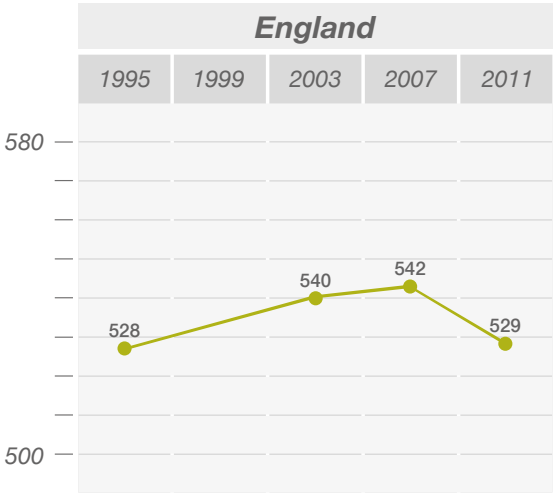
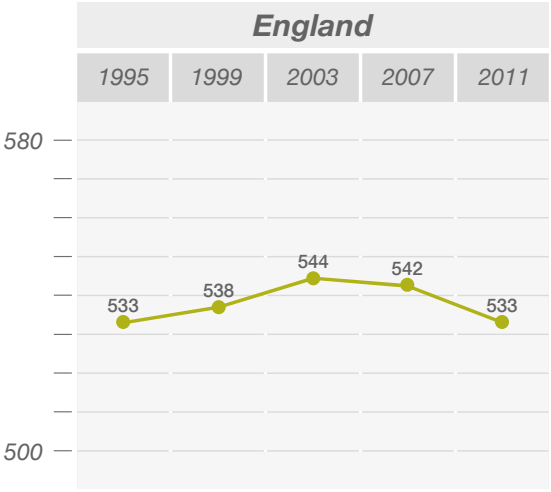


Figure 4. Year 9 science



There could be a number of reasons for the discrepancy between England's relative performance in PISA and TIMSS - student motivation, age of test-takers, countries involved, population representation, translation, relationship with curriculum, and familiarity of test format - meaning that there are limitations to what we can conclude.

There is wide variation between England's highest and lowest scoring pupils in TIMSS

TIMSS data show that England has a wide spread of achievement. In TIMSS, the difference between the highest and lowest scoring pupils is just under 300 TIMSS scale points for mathematics 4th grade (Sturman et al., 2012), where the scale runs from 0 to 1000. This is a spread greater than the difference between the highest and lowest international benchmarks. In comparison, Finland and Hong Kong had a spread closer to 200.

Among top performing countries in PISA, Canada, Finland, Shanghai-China, Hong Kong-China, Japan and Korea all have above–average performance in reading and they also have higher equity levels (OECD, 2013, p.2). In PISA 2009, 70% of England's between-school performance variance was explained by the socio-economic intake of schools (OECD, 2010c).

Nearly all of the countries that improved their PISA scores between 2000 and 2009 showed a decrease in the percentage of low performers. For these countries, the number of students scoring below the baseline reading level decreased significantly (OECD, 2010b). This highlights the need for provision for low achieving students from disadvantaged socio-economic backgrounds in England to be improved. Reducing the tail of underachievement can raise the overall mean in the international test scores.

Governments justify policies using international test data

Governments have been found to use international test data as a governing device and for policy legitimation (Lawn and Grek, 2012). For example, pre-existing policies were justified by reference to PISA results in

- Japan (Takayama, 2008)
- Portugal (Afonso and Costa, 2009)
- Turkey (Gür et al., 2012)

International test data do not produce ready-made solutions. Many different policies are proposed with the aim of raising standards in countries worldwide.

Criticisms of international test data

International tests have been criticised as having limitations as measures of a country's educational standards, in a number of ways:

- the curriculum tested might differ from that taught (Nardi, 2008)
- the age at which pupils are tested is not consistent across countries (Prais, 2003, 2007) and pupils can be at different stages of their education even if they are the same age (Wagemaker, 2008)
- there are too few test questions for reliable and valid measures (Ruddock et al., 2006; Hutchison and Schagen, 2007)
- students could be more motivated to do well in some countries (Eklöf, 2010)
- test translation could affect student performances (Hilton, 2006)
- the psychometric models are contested (Goldstein, 2004).

Overall, the international test data cannot simply be taken at face value and need to be interpreted carefully. Secondary data analyses, which use further information about students and schools, can be useful ways to use the data to investigate relationships and potential causes of test outcomes (e.g. see Olsen, Prentzel and Martin 2011).

"Valid use of the data and outcomes can be achieved by treating the international findings not as end-points, but as useful indicators and starting points for further investigation."

Sturman (2012, p.17)

Section 2: Grade inflation at GCSE has not been established

Research on grade inflation at GCSE reaches variable conclusions – there is not a general trend

- In 1986, 27% of 16 year olds achieved 5 o-level or CSE passes at grade C or above (Bolton, 2012). By 2011, 81% of 16 year olds achieved at least 5 A*-C GCSEs (DfE, 2012b).
- The proportion of students entered for no GCSE examinations has gone down from 6.5% in 1990 (DfEE, 1988) to 0.5% in 2012 (DfE, 2012b; figure includes equivalents, such as GNVQ).
- We do not know the extent to which different factors have caused these rises. Key questions remain unanswered. For example,
  - » Are students working harder?
  - » What effects has commercial competition between examination boards had upon grading?
  - » Are higher grades due to teaching to the test?
  - » How much of an increase in grades could legitimately be due to changes in teaching and other improvements in the learning environment?
- Grade inflation is a concern in many countries, particularly in higher education (HE).[4]

Pupils might be getting brighter

Lynn (2009) reported that fluid intelligence (logical and analytical reasoning) had risen between 1979 and 2008 in primary and secondary school pupils, whilst vocabulary levels had stayed static. Increases in fluid intelligence were most notable at the lower end of the ability range. Lynn (2009) also reported improvements in fluid intelligence in Britain between 1938 and 2008 for children aged 5–12. In ages 12–15, he reported small declines in IQ since 1979.

There are of course problems in making claims about changes in intelligence over these periods because the tests themselves have changed. (This parallels the difficulty in interpreting examination result rises, as the content of syllabuses changes over time too.) Shayer et al. (2007) compared year 7 pupils on the same test of scientific thinking. They found a sizeable drop in performance between 1976 and 2003. Although evidence is mixed, there is some indication of rising fluid intelligence (see also Folger, 2012). Just as with examinations, there are debates about whether the causes are artefacts of the tests or due to more legitimate reasons, such as improvements in diet, schooling or reactions to the complexity of modern society (see Neisser, 1998; Nisbet et al., 2012).

Equally bright students get one grade higher

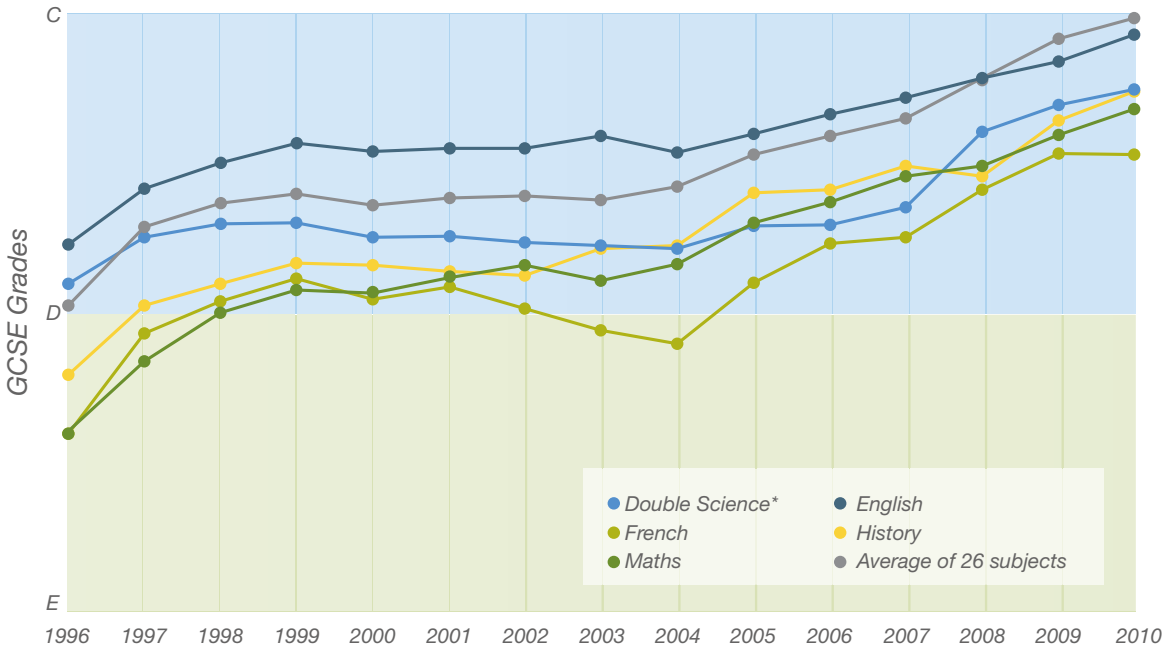
Taking students with the same score on an ability test (Yellis[5]), Coe (2007) showed that the average grade went up from D to C between 1996 and 2006 (Figure 5). Results rose for the first two years, for five years there was a flat profile and from 2003 results began to rise again. Yellis is a sensible comparison, as the ability test has a good relationship with average GCSE score (correlation approximately 0.8[6]), although the correlations with individual subjects will be lower. To interpret the findings, we must assume that the ability test has remained equally relevant over time and that the samples of students included in each year were representative.

[4] e.g. Canada – Anglin and Meng, 2000; New Zealand – Hickson and Agnew, 2012; Sweden – Wikström, 2005; US - Bejar and Blew, 1981; Johnson, 2003

[5] http://www.cem.org/yellis/introduction

[6] http://www.cemcentre.org/yellis/research-yellis-2008-gcse-analysis

Figure 5. Average grade achieved in GCSE subjects by students with Yellis score of 45

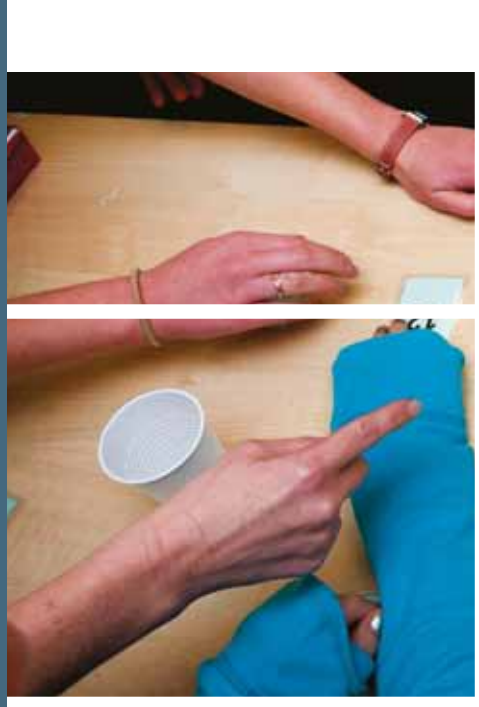


These rises have occurred in the context of schools having to achieve targets for the proportion of pupils attaining 5 A* to C grades. A great deal of attention has therefore gone into getting students who were on a trajectory for a grade D to get over the hurdle for a grade C, so-called 'bubble students'.

Rises in grades might be explained by harder work, better teaching, more support for learning and higher aspirations. But examinations could be easier, or they could be testing the wrong kind of learning in a target-driven education system.

Demand of tests – findings are inconsistent

Government agency reports (QCA, Ofsted, SCAA, Ofqual) on this topic typically involve senior examiners and other experts coming together to compare the demands of syllabuses, question papers and marking schemes, as well as looking at the quality of students' performances at the key grades. Judgments of comparisons over time are generally produced, although there are some exceptions. Each row in Table 1 represents a study on English, mathematics or science GCSE examinations. Before 1988 the examinations were o-level and CSE. There is no overall trend in standards from these studies.



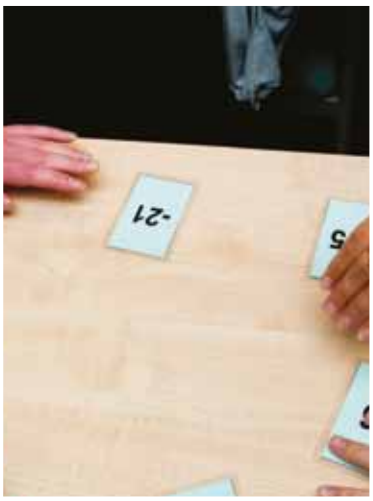


Table 1. Comparison of GCSE standards over time

SUBJECT		75	77	80	85	90	95	97	98	99	00	02	03	04	05	07	08	09
English Literature	a			●	●	●	●				●							
	b										●					●		
English	c	●		●	●	●	●											
	d									●		●						
	e											●			●			
	f														●			●
Maths	g	●		●	●	●	●											
	h						●			●								
	i									●				●				
	j													●			●	
Science	k						●				●							
	l										●				●			
Biology	m								●				●					
	n												●				●	
Chemistry	o	●		●	●	●	●											
	p						●		?									
	q								●				●					
	r												●				●	
Physics	s		●					?										
	t							●				?						
	u											●				●		

Key

Legend:

● = first year of study
● = standards declined
? = no conclusion
● = no change
● = improved

Studies:

[a,g,o]QCA, 2004b
[b]Ofqual, 2009a
[c]SCAA/Ofsted, 1996
[d]QCA,2004a
[e]QCA, 2007b
[f]Ofqual, 2011a
[h]QCA, 2001b
[i]QCA,2006a,
[j]Ofqual,2012c
[k]QCA, 2004c
[l]QCA, 2007c
[m]QCA, 2005a
[n]Ofqual, 2012a
[p]QCA, 2001a
[q]QCA, 2005b
[r]Ofqual, 2012b
[s]QCA, 2001c

These reports rely upon qualitative, expert judgments of very large volumes of paperwork that represent the examination standards. Often, the study investigated standards between examination boards as well as over time. Rarely has inter-judge reliability been investigated and reported. The volume of student work available also varied and sometimes only syllabuses, questions papers and marking schemes could be studied.

The Royal Society of Chemistry (2008) took a different approach, and administered an examination composed of chemistry questions sampled from o-level and GCSE examination papers between the 1960s and 2000s (six from each decade). The 1,301 volunteer students who took the examination scored more highly on the questions from the 2000s question papers (average of 35%) than in the 1960s questions (average 15%). Every year's examination paper will have consisted of easier and harder questions.

Only a small number of questions were sampled from each decade, so it is possible that they were unrepresentative of the time. Notwithstanding, the report also shows that there was variability of performance on questions within a decade. One of the 1960s questions was 6th easiest of the 40 questions, with the top five easiest questions coming from question papers in the 2000s and 1990s. There have been policy efforts to make question papers more accessible, and these findings might reflect those changes. Another issue is that current students are prepared to answer the question styles reflected in the most recent years, rather than the o-level styles from the 1960s.

A great deal of research has been conducted on the comparability of examinations, and there is a QCA book on Techniques for Monitoring the Comparability of Examination Standards (Newton et al., 2007), which raises methodological issues that we have not covered here.

Section 3: Raising standards

The consultation document (DfE, 2012a) refers to a desire to 'drive up standards', and one of the mechanisms suggested for this is to raise the level of challenge in our exams. Raising the challenge means increasing the demands of our exams, but unless we equip students with the means to meet those demands, it will not have the desired effect. It is also important that any extra challenge is relevant to what we are trying to test. We do not want to introduce additional reading demands to a science exam for example. This will not raise standards in science (Pollitt and Ahmed, 2001).

It's not easy to tell if we have raised the level of challenge

It is not easy to judge in advance whether the challenge is higher in one written exam than in another. Crisp and Novakovic (2009) report that examiners are not accustomed to thinking about demands in this way, and that they may be influenced by other aspects of the exam questions such as the style of writing, rather than the cognitive demands underpinning the tasks.

The requirements that we add to the examination may or may not make it more difficult – we will not know this until students have completed the exam. Even then, we cannot be certain that the exam was more difficult for the right reasons. If students score lower marks, is this because we have raised the challenge in the right way? Have we increased the demands in what is important in this subject? If not, then we will not raise standards (Pollitt et al., 2007).

If we raise the challenge too much, we may find that we are testing factors other than students' knowledge, skills and understanding. Talbi (2003) carried out an experimental study, manipulating the demands in mathematics questions in Scotland. Higher-grade pupils and first-year university students (n=223) attempted these questions. Talibi found a sharp drop in performance when demand was increased, which he interpreted as a point of overload, after which students' working memory capacity, rather than their mathematics skill was being tested.

At A-level, extended projects were introduced to raise the level of challenge, by adding depth and breadth and in order to test a wider range of skills (Daly and Pinot de Moira, 2010). It is not so easy to stretch students within the confines of an end of course written exam. Students perceive case studies and application of knowledge tasks to be more challenging than recall questions, requiring them to think more deeply (Daly and Pinot de Moira, 2010).

Is increasing the challenge appropriate for the whole cohort?

Consider the students who currently don't reach five grades A* to C at GCSE or equivalent: 18.9% of the cohort in England in June 2012 (DfE, 2012b). This figure is 41.4% if GCSE English and GCSE mathematics are required. Is it appropriate to raise the level of challenge in our exams for these pupils?

Increasing the challenge may also be inappropriate for higher achieving students. A study by QCA (2008) included a comparison of demands in geography and history qualifications and found that

> *"In every case except geography at AS, reviewers rated both subjects as slightly too demanding, with GCSE history being seen as significantly demanding."*
>
> **QCA (2008, p.3)**

Daly et al. (2012) looked at the responses of teachers and students to the introduction of stretch and challenge at A-level. Most students who participated in the study thought that it was demoralising to increase challenge at A-level, and some students and teachers felt it would undermine confidence.

Ways of raising standards

"School systems considered successful spend large amounts of money on education, and tend to prioritise teachers' pay"

OECD (2010d, p.14)

Top PISA performers, such as Finland, give teachers high professional status and ongoing high quality training, thereby encouraging high achieving individuals to become teachers (OECD, 2010d). They also help children to apply knowledge rather than just recall and repeat it.

Improving the provision for low performing students and increasing equality of learning outcomes is another way to improve countries' overall performances (OECD, 2010b).

We can make sure our exams are assessing what is important in a specific subject, via better questions and mark schemes. Valuing our examiners as professionals and investing in training for them will help with this.

There are methods for ensuring quality at the design stage in large-scale assessments, such as Evidence Centred Design (ECD) (Mislevy et al., 2003) or Outcome Space Control and Assessment (OSCA) (Pollitt et al., 2008). These methods aim to keep task demands relevant to what we are striving to test. The result should be a fairer measure of the students, so that those who are better at what we are testing will be the ones who get the most marks.

Section 4: Knowing what is expected

Access to past papers and mark schemes allows students to concentrate on showing what they know and can do in an exam, without being confused or distracted by the way in which a task is presented. Students' expectations of what an exam paper will look like can have an effect on performance (Crisp et al., 2008). Andrade and Valtcheva (2009) say that understanding a mark scheme

"gives students valuable information about the task they are about to undertake and takes the guess-work out of understanding their learning targets, or what counts as high quality work."

(p.13)

Ofqual (2008) also reported that students feel it is important to know how their papers are marked. If we do not give teachers and students access to these assessment materials, we are in danger of assessing their ability to guess what is expected by examiners, rather than assessing their knowledge, skills and understanding of the subject. This is particularly true of less structured exam tasks such as essays: how can students' performance be judged fairly if they do not know what the demands of the task are? (Pollitt et al., 2008).

Students must be familiar with the format an exam will take, so that we are assessing their knowledge, skills and understanding rather than their ability to interpret the questions.

Ofqual (2008) states that

"It is important that question papers do not contain major surprises from one series to the next, otherwise far too much of what you will be measuring is how well candidates cope with those surprises rather than knowledge and understanding of the material."

(p.28)

They also make the point that unpredictability will reduce the reliability and validity of the assessment:

"unpredictable question papers are just as poor in assessment terms as overpredictable ones."

Ofqual (2008, p.3)

Self-assessment is essential for effective learning (Black and Wiliam, 1998), and is therefore a way to raise standards in our classrooms. This cannot easily be done without access to past papers and mark schemes, as these allow students and teachers to understand the criteria by which they will be assessed.

In Daly et al.'s (2012) study, many students commented that they recognised the importance of understanding the mark scheme. Andrade and Valtcheva (2009) argued that when students use a mark scheme themselves this can affect their performance positively. In another study, McDonald and Boud (2003) trained students in self-assessment, and this had a significant impact on their performance in conventional exams.

Section 5: Aspiration and motivation

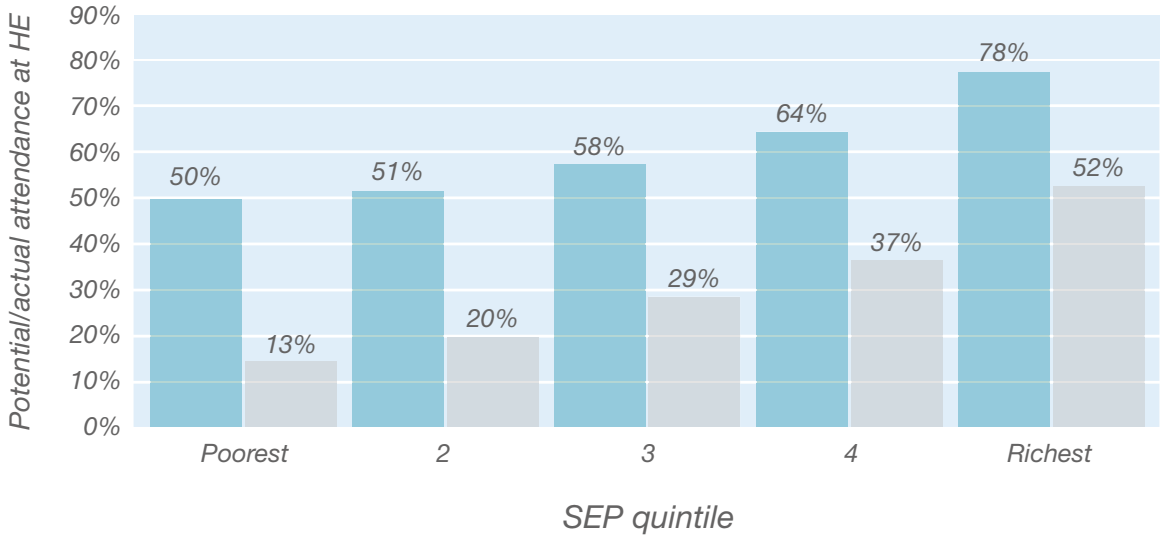
The youth of today are aspirational

In an analysis of the Longitudinal Study of Young People in England dataset, Chowdry, Crawford and Goodman (2011, p.74) reported that half of young people from the poorest backgrounds aspired to HE (Figure 6). We can see that socio-economic status had a relationship with aspiration, as was the case in the past (Schoon and Parsons, 2002), but there are now large groups from all sections of society who aspire to go on to HE. Actual attendance at HE institutions is also associated with socio-economic status.

We see from Chowdry et al.'s research that approximately two thirds of the richest young people who aspired to go on to HE fulfil this ambition; whilst only a quarter of the poorest students who aspired to HE at age 14 will fulfil that aspiration at age 18. Clearly, a number of factors caused this disparity, including differentials in opportunity.

Good GCSE grades, especially in mathematics and English, are required not only for progression through the academic track. They are also valued by employers, requirements for apprenticeships, and conditions for vocational course entry, such as nursing (Wolf, 2011).

Figure 6. Aspiration to attend HE at age 14 and attendance at HE at age 18 in socio-economic quintiles (from Chowdry, Crawford and Goodman, 2011)



Three principles of motivation

Ryan and Deci's (2001) self-determination theory of motivation indicates that there are three fundamental psychological needs:

- Competence – master new challenges and perfect new skills
- Autonomy – regulate themselves rather than be externally controlled
- Relatedness – meaningful bonds with others

Research indicates that these factors affect people's well-being, happiness, work satisfaction, positive social relationships and a sense of having meaning in life. Assessment and qualification systems can contribute to motivation by providing opportunities for mastery and allowing students to formulate their own learning goals to a larger extent than is currently possible.

Effort matters

According to Dweck (1986), there are two theories that people hold about their own intelligence, with associated mindsets relating to learning (Table 2). One theory is that intelligence is a fixed property that might even be physically endowed. These ideas about intelligence can lead to lack of effort in the classroom and concerns about the ego-threatening nature of failure.

Research on intelligence itself has shown that intelligence is a product of genes and the environment (see Nisbet et al., 2012, for a review). Intelligence test scores account for one fifth of variability in GCSE grades (Furnham and Monsen, 2009).

Dweck's second self-intelligence theory is that intelligence can be improved through study. If people hold this theory, they are more resilient in the face of inevitable difficulties in learning. Dweck calls this the growth mindset. How do we encourage this through our assessments?

Table 2. Dweck's self-theories of intelligence

Fixed self-theory of intelligence	***Incremental self-theory of intelligence***
Ability, personality and other characteristics are part of your physical make-up that cannot be changed.	Personal qualities are things that can be cultivated through effort.
Low-effort mindset	***Growth mindset***
Make as little effort as possible. Attribute setbacks to low ability rather than effort. Withdraw effort or persevere with the same strategy in response to failure.	Creates a passion for learning, for stretching yourself and sticking to it even when it is not going well.
Study techniques	
Read textbooks and class notes. Tried to memorise. Set performance goals that documented their abilities.	Looked for underlying themes and principles. Went over their mistakes until they understood the issues. Created their own motivation for learning the materials. Set themselves learning goals.

Section 6: Modular examinations

Modular courses are not necessarily easier

One of the common beliefs about modular courses is that they lead to improved grades. The evidence is mixed, depending on subject and level, but the differences are small. Investigating GCSE English and mathematics, Vidal Rodeiro and Nádas (2012) found that, taking prior attainment into account, modular courses were associated with slightly higher grades in mathematics but slightly lower grades in English.

"The fact that students obtain higher grades from a modular scheme does not necessarily mean standards have dropped …which would have meant that the qualification had become easier for students. It could be the case, for example, that with setting targets throughout the course, having ongoing feedback and allowing a certain amount of re-taking within the course, students are learning more."

Vidal Rodeiro and Nádas (2012, p.426)

Taverner and Wright (1997) found students with the same GCSE attainment got, on average, half a grade higher at A-level mathematics if they took a modular course instead of end of course assessments (with marginally more impact for boys than girls).

Ofqual (2012d) found no consistent pattern regarding whether modular or end of course GCSE produced higher results, and where there were differences, they tended to be small (0.2 of a grade). In two subjects (English and geography), the modular assessment route appeared to advantage students, after taking into account prior attainment, gender and centre type. Similar analyses showed the opposite pattern, with end of course assessments being advantageous in English literature, ICT, mathematics and religious education.

Re-sitting modules leads to improved scores (Vidal Rodeiro and Nádas, 2012; McClune, 2001; Taverner and Wright, 1997). Re-sitting a module at the end of a course leads to the greatest improvement (Taverner and Wright, 1997). However, improved marks in re-sits are commonly linked to a maturation effect. In other words, although students originally got a lower mark, their improved mark is the one we would expect them to have achieved if they were following a linear course.

On the whole, the amount of re-sitting at GCSE is relatively low. The regulations for the modular GCSEs (Ofqual, 2012e) limit candidates to one re-sit of each unit but the evidence from A-level suggests that there are very few multiple re-sits (QCA, 2007a). In Vidal Rodeiro and Nádas's (2012) study, 90% of English GCSE students and just below 50% in mathematics did not re-sit any units.

In general at GCSE it is the weaker students who re-sit. The probability of a good grade is significantly decreased if students re-sit more than one module (Vidal Rodeiro and Nádas, 2010).

Modular syllabuses and subjects

Ofsted (1999) suggested that modular syllabuses were more likely to be appropriate for subjects such as mathematics or physics, and would be less suitable for subjects such as English or languages, where the nature of the learning process is more cumulative. Most A-level teachers agree with this assessment (Hodgson and Spours, 2003). The pattern of GCSE entries in mathematics (more modular) and English (more linear) supports the distinction (Vidal Rodeiro and Nádas, 2012).

However, the appropriateness of modular approaches to mathematics and science has been challenged elsewhere, on the basis that it makes it difficult for students to acquire a synoptic and thorough understanding of material (Hayward and McNicholl, 2007; McClune, 2001).

There are significantly different patterns of attainment related to modular and linear course choices in different subjects (Vidal Rodeiro and Nádas, 2010; 2012).

Is modular assessment effective preparation for advanced study?

One of the concerns raised about modular curricula is that they lead to the fragmentation and trivialisation of subject knowledge (Raffe, 1994). HE staff have expressed concern that modular assessment of 14-19 qualifications is responsible for a number of problems including compartmentalisation of learning, a poorly developed overview of subjects and an inability to connect discrete areas of knowledge (Wilde et al., 2006).

Vidal Rodeiro (2012) investigated whether following a modular or end of course assessment at GCSE had an effect on A-level performance. Given the same GCSE grade performance, there was no effect for English or ICT, but there was a slightly higher probability of passing and gaining high grades at A-level if students had taken an end of course assessment at GCSE in mathematics.

Module results are useful feedback

Students like modular courses, and use feedback from exams to adjust their behaviour in later modules (Hodgson and Spours, 2003). Teachers find modular courses contribute to assessment for learning and enable better focus in planning (Vidal Rodeiro and Nadas, 2010).

"Students of English and mathematics found feedback (positive and negative) useful and motivating and encouraged them to do better."

Vidal Rodeiro and Nadas (2010, p.viii)

This finding is in keeping with Hattie's (2009, p.178) international meta-analysis of 569 studies on the effects of frequent testing. He concluded that it had a positive, medium-sized effect.

Section 7: Ensuring the validity of end of course assessments

The consultation document suggests that the new examinations will be end-of-course, single, written examinations without the facility for a re-sit. Many of the issues below would also apply to modular examinations, but may not apply to more authentic assessments where the examined tasks are closer to the real-life tasks (Gipps, 1994).

High stakes testing needs to tell us what we want to know

GCSE grades determine progression to post-16 study, either widening or limiting the choice of courses, and in particular A-level subjects. Grades can also influence entry to HE and future career options. The stakes for schools, too, can be high. The percentage of students receiving 5 A*-C grades, including GCSE English and mathematics, is published in league tables, with under-achieving schools liable to receive additional inspections and funding implications as a consequence (Stobart, 2008).

With such high stakes involved, examinations need to be as reliable and valid as possible. The main challenge remains how to develop a valid and reliable exam that measures the subject domain as accurately as possible – and also predict what students are able to do after they have taken the exam. Examinations have been criticised for being unreliable (e.g. Wiliam, 2001) and inauthentic (e.g. Gipps, 1994) measures of what students know, as well as poor predictors of future performances (e.g. James and Chilvers, 2001). Whilst not universally accepted, these criticisms are likely to re-surface if assessments are introduced comprising entirely end of course examinations.

End of course assessments serve different purposes. The questions are whether we ask too much of a single assessment, and whether there are too many and conflicting purposes for the examinations.

Summative assessment, such as grades at both GCSE and A-level, is supposed to inform HE about what students are able to do, as well as predict their level of success. End of course exams are also important for future employers, who need to be able to rely on certification to provide an accurate reflection of knowledge and skills. In the current job market, 21st century skills and creativity are among those strongest in demand (Autor et al., 2003). The question remains as to what degree exams assess such skills.

Accountability purposes of examinations

End of course assessments are a key component of the accountability system, with GCSE attainment in particular used to rank schools in England. This ranking system has unintended consequences for teachers and students as the focus switches to how to get enough students over the C-grade threshold, rather than learning more broadly (Stobart, 2009).

Pressure from accountability systems where teachers are judged by students' test scores has also incentivised teachers to 'teach to the test', narrowing the curriculum to ensure students know what will be included (Hamilton et al., 2008). Accountability pressures raise the risk of narrowing the teaching material and focusing too much upon rote learning, instead of developing critical thinking and meta-cognitive awareness among students.

On the one hand, it is important for teachers to prepare their students for the end of course examinations and support their learning. On the other, teaching to the test and too much transparency about what is tested on the final exam can limit students' learning to lower-order skills. Even if some research studies have provided evidence for increases in scores when teachers teach to the test, it is not a given that students are actually learning more:

"We need to better understand the extent to which performance gains on assessments reflect improved instruction and student learning or are a result of a more superficial intervention such as narrow test preparation activities."

Lane (2004)

In a review by the US National Research Council, it was found that test-based incentive programmes had not increased student achievement enough compared to other high achieving countries. In fact, some of the high-school exit programmes in the US decreased the rate of high school graduation without increasing achievement (National Research Council, 2011).

On the other hand, a recent review on the effects of testing on student achievement between 1910 and 2010, including both quantitative and qualitative studies on almost 750,000 students, found positive effects of testing. Moderate to fairly large effect sizes across several hundred studies were found and 93% of the qualitative studies also reported positive effects of testing (Phelps, 2012).

Higher order thinking skills, not just rote learning

Ideally, learners should be supported to approach knowledge with the goal of deep learning and developing understanding. Students should thus be taught to use deep-learning strategies, and not surface strategies when preparing for a test. Therefore, it is crucially important that exam tasks are rich and complex, so that students are required to engage in critical thinking, reflection, analysis and reasoning (Brookhart and Nitko, 2011).

Developing a single GCSE exam to test students' knowledge and skills would raise the stakes, since students would only sit one exam. It would also increase the likelihood of teachers tightly focusing lessons on exam content and strategies. Hodgson and Spours (2003, p.93) and Priestley (2003) found that teachers reported teaching to the test when Curriculum 2000 A-levels were first introduced. In another study, students from the first cohort to sit the newest syllabus A-level examinations in summer 2008 reported being drilled in examination techniques, which interfered with the breadth and depth of their studies (Baird et al., 2009).

What is tested gets taught

There is a tendency, then, for end of course examinations to drive the curriculum or "what you test is what you get" (Madaus, Russell and Higgins, 2009). Ideally, the end of course examination should be complex, demanding and stimulating, so that the backwash from the examination encourages teachers to develop complex and varied teaching material to support students' critical thinking, participation and motivation. However, this needs to be monitored.

Impact of examinations on students

Research has shown that girls tend to score more highly than boys on test questions that require a longer written response, whereas boys tend to score higher than girls on multiple choice questions (DeMars, 2000). Different test-formats should therefore be discussed in relation to fairness of high stakes tests.

There is evidence that learners face test anxiety around GCSEs (Putwain, 2008) and it has a negative impact upon scores - an effect which is greater for lower socio-economic group students. Anxiety can influence high ability students to the extent that even this group can fail to show what they really know on the test (Pintrich and DeGroot, 1990; Putwain, 2009).

More than a written examination

More than 50 years ago, Bruner described how the curriculum of a subject should be determined by "the most fundamental understanding that can be achieved of the underlying principles that give structure to that subject" (Bruner, 1960, p.31). A modern assessment should be able to assess such knowledge and skills, but the question is whether a written examination alone would be able to do that. In times where pilots are learning to fly in simulators and doctors conduct operations in teams across borders, students are still expected to demonstrate their knowledge through pen and paper and are seldom assessed by computer-based simulations.

Authentic assessments, introduced into GCSE examinations over 20 years ago, might be better indicators of skill in sport, art, science practicals, cooking, geography fieldwork, dance, presentation skills, speaking in a foreign language and so on.

Section 8: Teacher assessment

Teachers' judgments are often used to assess aspects of pupils' work that cannot be validly assessed by end of course examinations. Such assessments are in use in many well-established high stakes systems internationally.[7] Teacher assessment can take many forms, depending upon, for example, the proportion of the assessment that is teacher-assessed, the domains assessed, the number of assessment occasions, the role it plays in the final grade, the source and duration of the tasks, the task type and the ways in which evidence is collated (Wilmut, 2005).

Positive effects upon teaching and learning

The interaction between students and teachers during the production of work that is teacher assessed can involve positive, formative features. In their influential work, Black and Wiliam (1998, 2009) claimed very large effect sizes for formative assessment that would produce improvements of one to two grades at GCSE if implemented. Formative assessment has therefore been adopted by many governments internationally as a way to boost achievement (e.g. Chile, Hong Kong, New Zealand, Norway and Scotland). Some authors have disputed the evidence for these effect sizes in the Black and Wiliam research (Elwood, 2007; Bennett, 2011). Separately, Hattie (2009, p.173) published a meta-analysis of 1,287 studies specifically on feedback and found that it had a high, positive effect upon learning.

Problems with teacher assessment

When reviewing the evidence contained within a number of studies in this area, Johnson (2013) found that the reliability of teacher assessment in some systems is limited. Teachers may be influenced by a number of irrelevant factors, including gender, socio-economic background, effort and behaviour, that risk biasing their judgments where assessment requires degree of subjective interpretation.

Cheating in coursework has also been an ongoing concern at GCSE. Bishop, Bullock, Martin and Thompson (1997) found that although students thought coursework was important to GCSEs, many thought that it was easy to cheat and that it was not possible to be sure that the work had been conducted by the student. Overly supportive parents were considered to be part of the problem.

A review of GCSE coursework was carried out by the Qualifications and Curriculum Authority (QCA, 2006b) and one of the recommendations was that coursework set and marked by the teacher should be replaced by external exams or by controlled assessments in the majority of subjects. The widespread introduction of controlled assessments followed.

Ofqual's (2011b) evaluation of controlled assessment indicated that teachers were broadly supportive of the system, although many had concerns about its ease of implementation in their schools. There were also concerns from teachers about the amount of teaching time taken up by controlled assessment, resulting in a narrowing of their teaching and reducing the time they can take for extra-curricular activities. Suggestions for improvement from teachers and stakeholders most commonly included having fewer and shorter tasks that are less strictly controlled.

Quality assurance for teacher assessment

Black et al. (2011) summarised international approaches to teacher assessment moderation systems, focusing on those developed in the Australian states of New South Wales and Queensland (see also Stanley et al., 2009; Klenowski, 2013). These systems assessed a limited number of tasks from each pupil (i.e. portfolios) and involved extensive resources for professional development. Moderation procedures involved either scaling teacher judgments against external tests in New South Wales, or joint work across schools in Queensland (where outcomes are based entirely on teachers' assessments). Wilmut and Tuson (2004) provided a review of forms of statistical moderation of teacher assessment.

[7] e.g. Australia (Stanley et al., 2009), Canada (Aitkin, 2009), US (Brookhart, 2013), Portugal (Fernandes, 2009) and Sweden (Wikström , 2005)

Section 9: Assessing the whole ability range

What are the challenges for assessing the whole ability range?

One qualification intended to cover the whole ability range must be both accessible to the bottom end and provide suitable challenge for the top end. Differentiation between students can be achieved by task demand, in which case lower ability students are not able to tackle all questions well. Alternatively, differentiation can be achieved by outcome: the same task is given to all, but students are graded differently depending upon the quality of their responses.

"Examinations papers and other methods of assessment which are used should be such that they enable candidates to demonstrate what they do know rather than what they do not know. …examinations should not undermine the confidence of those who attempt them."

Cockcroft (1982, p.158)

Accessibility is a key issue both for accurate assessment and for motivation. Demands that are introduced because of the way the question is worded, for example, will result in difficulty that is not relevant to the construct being tested, and will unfairly exclude some students from gaining marks (Pollitt at al, 2008).

When differentiation is by task, assessing accurately across the full range requires a very large number of questions and a high level of testing time (Wiliam, 2001). Such a system would be prohibitively expensive, time-consuming and also has repercussions on motivation. This is an obstacle to setting a single examination for the entire population. The solution commonly adopted is to require students to only take only a subset of questions from the whole.

Methods for assessing the whole ability range accurately

A great deal of research on differentiation was conducted prior to the introduction of GCSE examinations (see Appendix A). Methods that have been investigated in the literature include the following.

Tiers: Students can be entered for one of two or more question papers that vary in difficulty.

Multiple papers: A number of tiered papers, in which students can sit two adjacent tiers. They can be overlapping or not. The current Scottish Standard Grade examinations have three tiers which do not overlap and for which the majority of students sit two tiers.

General and extended papers: All candidates take a general paper; some take an additional paper to access higher grades.

Computer adaptive testing: One question is posed at a time. The computer uses an iterative algorithm starting with an initial estimate of a student's knowledge level, then chooses the best question out of all the ones that student hasn't yet been asked. That question is asked, and the response gives a new estimate of the student's knowledge level (Van der Linden and Glas, 2000). These steps are repeated until the criteria for being certain of the student's level of knowledge are met (Guzmán and Conejo, 2004; Guzmán et al., 2007). This requires a high level of pre-testing and calibration. It uses a very large bank of items. It is appropriate for testing on demand, when the student is ready.

Multistage testing: In this form of adaptive assessment, students follow a predefined path between testlets, depending upon their performance on each one. This method does not involve the complex algorithms that standard computer adaptive testing can involve (Mead, 2006).

Individual modules: Within a unified system, students could be differentiated for at the individual level as they took the module appropriate to them (Raffe, 1994).

What we know about tiering at GCSE

A high proportion of French, mathematics and science teachers surveyed were in favour of tiering of some sort, believing that it stretches the most able (Baird et al., 2001). Only mathematics and modern foreign languages are currently differentiated by different content for different tiers; other subjects rely on question difficulty (Wheadon, 2011).

At least 50% of teachers in all subjects reported de-motivation of students entered for lower tiers (Baird et al., 2001). The use of different content for mathematics and modern foreign languages means that students are likely to have decisions made about their future very early, in that lower ability sets will not be taught the necessary material and high-performing students within those groups cannot move into a higher set (Elwood and Murphy, 2002; Gillborn and Youdell, 2000; Baird et al., 2001).

There is evidence that setting or streaming by ability is influenced by other factors, such as socio-economic status, gender and race (Gillborn and Youdell, 2000). This is a problem if it prevents students from attaining as highly as they could.

Concerns that teachers could predict grade performance inappropriately and enter students for the wrong tier (Good and Cresswell, 1988a) are balanced by evidence suggesting that they make appropriate choices (Murphy, 1979; Good and Cresswell, 1988c). However, the increasing issue of accountability for grades, particularly the 'magic' C, is believed to make it more likely that teachers will 'play it safe' in choosing the lower of two possible tiers (Gillbourn and Youdell, 2000; Burghes, Roddick and Tapson, 2001; Baird et al., 2001; Elwood & Murphy, 2002; Stobart et al., 2005). In this case, tiering would cap achievement. There was evidence from Baird et al. (2001) to show that a small, but significant, proportion of students' grades may have been capped by their entry for a lower tier of assessment. However, a statistical analysis by Wheadon and Béguin (2010) suggested that candidates entered for lower tier science papers were slightly over-rewarded.






Acknowledgements

Acknowledgements

We would like to acknowledge the people below who have generously provided us with insightful comments on a draft of this report. The content of the report and any remaining errors are of course the responsibility of the authors.

- Professor Patricia Broadfoot, University of Bristol
- Professor Richard Daugherty, Oxford University Centre for Educational Assessment
- Professor Jannette Elwood, Queen's University, Belfast
- Professor John Gardner, University of Stirling
- Professor Caroline Gipps, Institute of Education, University of London
- Professor Harvey Goldstein, University of Bristol
- Dr Tina Isaacs, Institute of Education, University of London
- Dr Sandra Johnson, Assessment Europe
- Dr Michelle Meadows, Assessment and Qualifications Alliance
- Professor Roger Murphy, University of Nottingham
- Professor Paul Newton, Institute of Education, University of London
- Professor Gordon Stobart, Institute of Education, University of London
- Kath Thomas, Pearson UK
- Professor Jim Tognolini, Pearson International
- Professor Lorna Unwin, Institute of Education, University of London

We are grateful to Pearson UK for the support towards research assistants and dissemination of this work. We hope it will provide a useful resource for their own thinking and that of others working in qualifications development.

References

Afonso, N. & Costa, E. (2009) The influence of the Programme for International Student Assessment (PISA) on policy decision in Portugal: the education policies of the 17th Portuguese Constitutional Government, ***Educational Sciences Journal***, 10, 53-64.

Aitkin, N. (2009) Country Case Study: Canada (Alberta), Chapter 5 in: B. Vlaardingerbroek & N. Taylor (Eds) ***Secondary School External Examination Systems.*** Cambria Press, NY.

Anglin, P.M. & Meng, R. (2000) Evidence on grades and grade inflation at Ontario's universities, ***Analyse de Politiques,*** 26 (3), 361-368, September.

Andrade, H. & Valtcheva, A. (2009) Promoting learning and achievement through self-assessment, ***Theory into Practice,*** 48, 12-19.

Autor, D.H., Levy, F. & Murnane, R.J. (2003) The Skill Content of Recent Technological Change: An Empirical Exploration, ***The Quarterly Journal of Economics,*** 118(4),1279-1333.

Baird, J., Chamberlain, S., Meadows, M., Royal-Dawson, L. & Taylor, R. (2009) ***Students' views of stretch and challenge in A-level examinations.*** Qualifications and Curriculum Authority Report, March.

Baird, J., Fearnley, A., Fowles, D., Jones,B., Morfidi, E. & White, D. (2001) ***Tiering in the GCSE: A study undertaken by AQA on behalf of the Joint Council for General Qualifications.*** Joint Council for General Qualifications.

Baird, J., Isaacs, T., Johnson, S., Stobart, G.,Yu, G., Sprague, T. & Daugherty, R. (2011) ***Policy effects of PISA.*** Oxford University Centre for Educational Assessment, Report commissioned by Pearson UK.

Bejar, I.I. & Blew, E.O. (1981) Grade inflation and the validity of the scholastic aptitude test. ***American Educational Research Journal,*** 18, 143-156.

Bennett, R.E. (2011) Formative Assessment: a critical review, ***Assessment in Education: Principles, Policy & Practice,*** 18(1), 5-25.

Bishop, K.N., Bullock, K., Martin, S. & Thompson, J.J. (1997) Students' Perceptions of Coursework in the GCSE: the effects of gender and levels of attainment, ***Educational Studies,*** 23(2), 295-310.

Black, P., Harrison, C., Hodgen, J., Marshall, B.,& Serret, N. (2011) Can teachers' summative assessments produce dependable results and also enhance classroom learning?, ***Assessment in Education: Principles, Policy & Practice,*** 18(4), 451-469.

Black, P. & Wiliam, D. (1998) Assessment and Classroom Learning, ***Assessment in Education: Principles, Policy & Practice,*** 5(1), 7-74.

Black, P. & Wiliam, D. (2009) Developing the theory of formative assessment, ***Educational Assessment, Evaluation and Accountability,*** 21 (1), 5-31.

Bolton, P. (2012) ***Education: Historical statistics.*** Standard Note: SN/SG/4252. London, House of Commons Library.

Brookhart, S.M. & Nitko, A.J. (2011) Strategies for constructing assessments of higher order thinking skills, in: G. Schraw & D.R.Robinson (Eds) ***Assessment of higher order thinking skills. A volume in Current Perspectives on Cognition, Learning and Instruction.*** Information Age Publishing, Inc.

Brookhart, S.M. (2013) The use of teacher judgement for summative assessment in the USA, ***Assessment in Education: Principles, Policy & Practice,*** 201(1), 69 -90.

Bruner, J. (1960) ***The Process of Education.*** Harvard University Press.

Burghes, D. Roddick, M. & Tapson, F. (2001) Tiering at GCSE: Is there a fairer system?, ***Educational Research,*** 43(2), 175-187.

Chowdry, H., Crawford, C. & Goodman, A. (2011) The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 16, ***Longitudinal and Life Course Studies,*** 2(1), 59-76. Special Issue: The socio-economic gradient in cognitive and educational achievement.

Coe, R. (2007) ***Changes in standards at GCSE and A-Level: Evidence from ALIS and YELLIS,*** Report for the Office for National Statistics.

Cockcroft, W.H. (1982) ***Mathematics Counts.*** Report of the Committee of Inquiry into the Teaching of Mathematics in Schools under the Chairmanship of Dr WH Cockcroft. The Cockcroft Report. London, HMSO.

Crisp, V. & Novakovic, N. (2009) Is this year's exam as demanding as last year's? Using a pilot method to evaluate the consistency of examination demands over time, ***Evaluation & Research in Education,*** 22(1), 3-15.

Crisp, V., Sweiry, E., Ahmed, A. & Pollitt, A. (2008) Tales of the expected: the influence of students' expectations on question validity and implications for writing exam questions, ***Educational Research,*** 50(1), 95-115.

Daly, A.L., Baird, J., Chamberlain, S. & Meadows, M. (2012) Assessment reform: students' and teachers' responses to the introduction of stretch and challenge at A-level, ***The Curriculum Journal,*** 23(2), 139-155.

Daly, A.L. & Pinot de Moira, A. (2010) Students' approaches to learning and their performance in the Extended Project pilot, ***Curriculum Journal,*** 21(2), 179-200.

DeMars, C. (2000) Test Stakes and Item Format Interactions, ***Applied Measurement in Education,*** 13(1), 55-77.

Department for Education (DfE) (2012a) ***Reforming Key Stage 4 Qualifications.*** Consultation Document.

Department for Education (2012b) ***Statistical First Release. GCSE and Equivalent Results in England 2011/12 (Provisional).*** SFR 25/2012, 18 October 2012. Available online at: http://www.education.gov.uk/rsgateway/DB/SFR/s001094/sfr25-2012.pdf, last accessed January 2013.

Department for Education and Employment (1998) ***Statistics of Education. Public examinations GCSE/GNVQ and GCE/AGNVQ in England 1988.*** The Stationery Office. ISBN 0 11 271068 9.

Fernandes, D. (2009) Educational assessment in Portugal, ***Assessment in Education: Principles, Policy & Practice,*** 16(2), 227-247.

Dweck, C. (1986) Motivational processes affecting learning, ***American Psychologist,*** 41(10), 1040-1048. Special Issue: Psychological Science and Education.

Eklöf, H. (2010) Skill and will: test-taking motivation and assessment quality, ***Assessment in Education: Principles, Policy & Practice,*** 17(4), 345-356.

Elwood, J. (2007) Formative assessment: possibilities, boundaries and limitations, ***Assessment in Education, Principles, Policy & Practice,*** 13(2), 215-232.

Elwood, J. & Murphy, P. (2002) Tests, tiers and achievement: gender and performance at 16 and 14 in England, ***European Journal of Education,*** 37(4), 395-416.

Fisher-Hoch, H. & Hughes, S. (1996) ***What makes mathematics exam questions difficult?*** Paper presented at British Educational Research Association conference.

Folger, T. (2012) Can we keep getting smarter? Ever rising IQ scores suggest that future generations will make us seem like dimwits in comparison, ***Scientific American,*** 307(3), 44-47, September.

Furnham, A. & Monsen, J. (2009) Personality traits and intelligence predict academic school grades, ***Learning and Individual Differences,*** 19(1), 28-33.

Gillborn, D. & Youdell, D. (2000) ***Rationing Education: Policy, Practice, Reform and Equity.*** Buckingham, Open University Press.

Gipps, C. (1994) Developments in Educational Assessment: what makes a good test?, ***Assessment in Education: Principles, Policy & Practice,*** 1(3), 283-292.

Goldstein, H. (2004) International comparisons of student attainment: some issues arising from the PISA study, ***Assessment in Education: Principles, Policy & Practice,*** 11(3), 319-330.

Good, F.J. & Cresswell, M.J (1988a) Can Teachers Enter Candidates Appropriately for Examinations Involving Differentiated Papers?, ***Educational Studies,*** 14(3), 289-297.

Good, F.J. & Cresswell, M.J. (1988b) Grade Awarding Judgements in Differentiated Examinations, ***British Educational Research Journal,*** 14(3), 263-281.

Good, F.J. & Cresswell, M.J. (1988c) Placing candidates who take differentiated papers on a common grade scale, ***Educational Research,*** 30(3), 177-189.

Gür, B.S., Çelik, Z. & Özoğlu, M. (2012) Policy options for Turkey: a critique of the interpretation and utilization of PISA results in Turkey, ***Journal of Education Policy,*** 27(1),1-21.

Guzmán, E. & Conejo, R. (2004) A Model for ***Student Knowledge Diagnosis Through Adaptive Testing,*** in: Proceedings of Intelligent Tutoring Systems 7th International Conference, ITS 2004, Brazil.

Guzmán, E., Conejo, R. & Pérez-de-la-Cruz, J.-L. (2007) Adaptive testing for hierarchical student models, ***User Modeling and User-Adapted Interaction,*** 17(1-2), 119-157, March.

Hamilton, L.S., Stecher, B.S. & Yuan, K. (2008) ***Standards-Based Reform in the United States: History, Research, and Future directions.*** Paper commissioned for the Center on Education Policy, Washington, D.C. For its project on Rethinking the Federal Role in Education. Available online at: http://www.cep-dc.org/displayDocument.cfm?DocumentID=332, last accessed January 2013.

Hattie, J. (2009) ***Visible Learning: A synthesis of over 800 meta-analyses relating to achievement.*** Routledge, New York.

Hayward, G. & McNicholl, J. (2007) Modular Mayhem? A case study of the development of the A-level science curriculum in England, ***Assessment in Education: Principles, Policy & Practice,*** 14(3), 335-351.

Hickson, S. & Agnew, S. (2012) Assigning grades during an earthquake – Shaken or stirred? ***New Zealand Economic Papers,*** DOI:10.1080/00779954.2012.715825.

Hilton, M. (2006) Measuring standards in primary English: issues of validity and accountability with respect to PIRLS and National Curriculum test scores, ***British Educational Research Journal,*** 32(6), 817-37.

Hodgson, A. & Spours, K. (2003) ***Beyond A Levels: Curriculum 2000 and the Reform of 14-19 Qualifications.*** London, Kogan Page.

Hutchison, D. & Schagen, I. (2007) Comparison between PISA and TIMSS – are we the man with two watches? In: T. Loveless (Ed) ***Lessons Learned: What International Assessments Tell Us About Math Achievement.*** Washington DC, Brookings Institution Press.

James, D. & Chilvers, C. (2001) Academic and non-academic predictors of success on the Nottingham undergraduate medical course 1970–1995, ***Medical Education,*** 35(11), 1056-1064.

Johnson, S. (2013) On the reliability of high-stakes teacher assessment, ***Research Papers in Education,*** 28(1), 91-105.

Johnson, V.E. (2003) ***Grade inflation: a crisis in college education.*** Springer: New York.

Klenowski, V. (2013) Special Issue: Moderation practice and teacher judgement. ***Assessment in Education: Principles, Policy & Practice,*** 201(1).

Lane, S. (2004) Validity of High-Stakes Assessment: Are Students Engaged in Complex Thinking?, ***Educational Measurement:*** Issues and Practice, 23(3), 6-14.

Lawn, M. & Grek, S. (2012) ***Europeanizing Education: governing a new policy space.*** Oxford, Symposium Books.

Lynn, R. (2009) Fluid intelligence but not vocabulary has increased in Britain, 1979–2008, ***Intelligence,*** 37(3), 249-255.

Madaus, G.F., Russell, M. & Higgins, J. (2009) ***The paradoxes of high stakes testing: How they affect students, their parents, teachers, principals, schools, and society.*** Charlotte, NC: Information Age Publishing.

McClune, B. (2001) Modular A-levels – who are the winners and losers? A comparison of lower-sixth and upper-sixth students' performance in linear and modular A-level physics examinations, ***Educational Research,*** 43(1), 79-89.

McDonald, B. & Boud, D. (2003) The Impact of Self-assessment on Achievement: The effects of self-assessment training on performance in external examinations, ***Assessment in Education: Principles, Policy & Practice,*** 10(2), 209-220.

Mead, A. D. (2006) An Introduction to Multistage Testing ***Applied Measurement in Education,*** 19(3), 185-187.

Mislevy, R.J., Steinberg, L.S. &Almond, R.G. (2003) On the structure of educational assessments, ***Measurement: Interdisciplinary Research and Perspectives,*** 1(1), 3-62.

Murphy, R.J.L. (1979) Teachers' Assessments and GCE Results Compared, ***Educational Research,*** 22(1), 54-59.

Nardi, E. (2008) Cultural biases: a non-Anglophone perspective, ***Assessment in Education: Principles, Policy & Practice,*** 15(3), 259-266.

National Research Council (2011) ***Incentives and Test-Based Accountability in Education.*** Committee on Incentives and Test-Based Accountability in Public Education.

Neisser, U. (1998) (Ed) ***The Rising Curve.*** Washington DC, American Psychological Association.

Nisbet, R.E., Aronson, J., Blair, C., Dickens,W., Flynn, J., Halpern, D.F. & Turkheimer,E. (2012) Intelligence: New Findings and Theoretical Developments, ***American Psychologist,*** 67(2), 130-159, February-March.

OECD (Organisation for Economic Cooperation and Development) (2010a) ***Education at a Glance 2010: OECD Indicators.*** Paris, OECD.

OECD (2010b) ***PISA 2009 Results: Learning Trends: Changes in Student Performance Since 2000 (Volume V).*** Available online at: http://dx.doi.org/10.1787/9789264091580-en, last accessed January 2013.

OECD (2010c) ***PISA 2009 Results: Overcoming Social Background: Equity in Learning Opportunities and Outcomes (Volume II).*** Available online at: http://dx.doi.org/10.1787/9789264091504-en, last accessed January 2013.

OECD (2010d) ***PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Volume IV).*** Available online at: http://dx.doi.org/10.1787/9789264091559-en, last accessed January 2013.

OECD (2010e) ***PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I).*** Available online at: http://dx.doi.org/10.1787/9789264091450-en, last accessed January 2013.

OECD (2010f) ***Viewing the United Kingdom School System Through the Prism of PISA.*** Available online at: http://www.oecd.org/pisa/46624007.pdf, last accessed January 2013.

OECD (2013) ***Are countries moving towards more equitable education systems?*** PISA in Focus, 25, 2013/02 (February). Available online at: http://www.oecd.org/pisa/pisainfocus/pisa%20in%20focus%20n25%20(eng)--FINAL.pdf, last accessed January 2013.

Ofqual (Office of Qualifications and Examinations Regulator) (2008) ***Predictability studies report. A study of GCSE and GCE level examinations.*** Ofqual/08/3866.

Ofqual (2009a) ***Review of standards in GCSE English literature. 2000 and 2007.*** Ofqual/09/4154.

Ofqual (2009b) ***Review of standards in physics. GCSE 2002-2007.*** GCE 2001–2007. Ofqual/09/4156.

Ofqual (2011a) ***Review of standards in GCSE English. 2005 and 2009.*** Ofqual/11/4846.

Ofqual (2011b) ***Evaluation of the introduction of controlled assessment.*** Ofqual/11/5049.

Ofqual (2012a) ***Review of standards in GCSE Biology.*** 2003 and 2008. Ofqual/12/5151.

Ofqual (2012b) ***Review of standards in GCSE Chemistry.*** 2003 and 2008. Ofqual/12/5152.

Ofqual (2012c) ***Review of standards in GCSE Mathematics***. 2004 and 2008. Ofqual/12/5154.

Ofqual (2012d) ***Effects of unitization in 2009 GCSE Assessments. Comparison on Candidate Achievement in Modular and Linear Assessments.*** Ofqual/12/5137.

Ofqual (2012e) ***Criteria for GCSE Qualifications.*** Ofqual/12/5124.

Ofsted (1999) ***Modular GCE AS and A-level examinations 1996–1998.***

Olsen, R.V., Prenzel, M. & Martin, R. (2011) Interest in Science: A many-faceted picture painted by data from the OECD PISA study, Editorial, ***International Journal Of Science Education,*** 33(1), 1-6.

Phelps, R.P. (2012) The Effect of Testing on Student Achievement, 1910–2010, ***International Journal of Testing,*** 12(1), 21-43.

Pintrich, P.R. & De Groot, E.V. (1990) Motivational and self-regulated learning components of classroom academic performance, ***Journal of Educational Psychology,*** 82(1), 33-40.

Pollitt, A. & Ahmed, A. (2001) ***Science or reading?: How students think when answering TIMSS questions.*** Paper presented at International Association for Educational Assessment conference, May.

Pollitt, A., Ahmed, A. & Crisp, V. (2007) The demands of examination syllabuses and question papers, in: P. Newton, J.Baird, H.Goldstein, H. Patrick & P. Tymms (Eds) ***Techniques for monitoring the comparability of examination standards.*** London, Qualifications and Curriculum Authority.

Pollitt, A., Ahmed, A., Baird, J., Tognolini, J. &Davidson, M. (2008) ***Improving the quality of GCSE Assessment.*** Report commissioned by Qualifications and Curriculum Authority.

Prais, S.J. (2003) Cautions on OECD's recent educational survey (PISA), ***Oxford Review of Education,*** 29(2), 139-163.

Prais, S.J. (2007) Two recent (2003) international surveys of schooling attainments in mathematics: England's problems, ***Oxford Review of Education,*** 33(1), 33-46.

Priestley, M. (2003) Curriculum 2000: A broader view of A levels, ***Cambridge Journal of Education,*** 33(2), 237-255.

Putwain, D.W. (2008) Test anxiety and GCSE performance: the effect of gender and socio-economic background, ***Educational Psychology in Practice: theory, research and practice in educational psychology,*** 24(4), 319-334.

Putwain, D.W. (2009) Assessment and examination stress in Key Stage 4, ***British Educational Research Journal,*** 35(3), 391-411.

QCA (Qualifications and Curriculum Authority) (2001a) ***Five year review of standards. GCSE chemistry.*** QCA/01/771.

QCA (2001b) ***Five year review of standards. GCSE mathematics.*** QCA/01/772.

QCA (2001c) ***Five year review of standards. GCSE physics.*** QCA/01/769.

QCA (2004a) ***GCSE English. Review of standards 1999–2002,*** March.

QCA (2004b) ***GCSE English literature. Review of standards 1980–2000 (includes GCE O level),*** March.

QCA (2004c) ***GCSE science: double award. Review of standards 1995– 2000,*** March.

QCA (2005a) ***Review of standards in biology. GCSE 1998 and 2003; A level 1999 and 2003***. QCA/05/1571.

QCA (2005b) ***Review of standards in chemistry. GCSE 1998 and 2003; A level 1999 and 2003.*** QCA/05/1572.

QCA (2005c) ***Review of standards in physics. GCSE 1997 and 2002; A level 1996 and 2001.*** QCA/05/1574.

QCA (2006a) ***Review of standards in mathematics: GCSE 1999–2004 and A level 1998–2004.*** QCA/06/2348.

QCA (2006b) ***A review of GCSE coursework.*** QCA/06/2736.

QCA (2007a) ***A-level re-sitting: summary of research findings.*** QCA/07/3387.

QCA (2007b) ***Review of standards in GCSE English 2002–5.*** QCA/07/3102.

QCA (2007c) ***Review of standards in GCSE Science (double award).*** QCA/07/3032.

QCA (2008) ***Inter-subject comparability studies.*** QCA/08/3568.

Raffe, D. (1994) Modular strategies for overcoming academic/vocational divisions: issues arising from the Scottish experience, ***Journal of Education Policy,*** 9(2), 141-154.

Royal Society of Chemistry (2008) ***The Five-Decade Challenge. A wake-up call for UK science education?*** November. Available online at: http://www.rsc.org/images/ExamReport_tcm18-139067.pdf, last accessed January 2013.

Ruddock, G., Clausen-May, T., Purple, C. & Ager, R. (2006) ***Validation Study of the PISA 2000, PISA 2003 and TIMSS-2003 International Studies of Pupil Attainment.*** DfES Research Report 772. Slough, NFER.

Ryan, R.M. & Deci, E.L. (2001) On happiness and human potentials: A Review of Research in Hedonic and Eudaimonic Well-Being, ***Annual Review of Psychology,*** 52, 141-166.

School Curriculum and Assessment Authority (SCAA)/Ofsted (1996) ***Standards in Public Examinations 1975–1995. A Report on English, Mathematics and Chemistry Examinations Over Time.*** School Curriculum and Assessment Authority Ref KS4/96/639. ISBN 1 85838 215 7.

Schoon, I. & Parsons, S. (2002) Teenage aspirations for future careers and occupational outcomes, ***Journal of Vocational Behavior,*** 60, 262-288.

Shayer, M., Ginsburg, D. & Coe, R. (2007) Thirty years on – a large anti-Flynn effect? The Piagetian test *Volume & Heaviness* norms 1975–2003, ***British Journal of Educational Psychology,*** 77(1), 25-41.

Stanley, G., MacCann, R., Gardner, J.,Reynolds, L. & Wild, I. (2009) ***Review of teacher assessment: Evidence of what works best and issues for development***. Report commissioned by Qualifications and Curriculum Authority. Available online at: http://oucea.education.ox.ac.uk/research/publications/last accessed January 2013.

Stobart, G. (2008) ***Testing Times.*** Abingdon, Routledge.

Stobart, G. (2009) Determining validity in national curriculum assessments, ***Educational Research,*** 51(2), 161-173.

Stobart, G., Bibby, T., Goldstein, H., Schagen, I. & Treadaway, M. (2005) ***Moving to two-tier GCSE mathematics examinations: An independent evaluation of the 2005 GCSE Pilot and Trial.*** Report commissioned by Qualifications and Curriculum Authority.

Sturman, L. (2012) Making best use ofinternational comparison data. ***British Educational Research Association Research Intelligence,*** Issue 119, Autumn/Winter, 16-17.

Sturman, L., Burge, B., Cook, R. & Weaving,H. (2012) ***TIMSS 2011: Mathematics and science achievement in England.*** Slough, National Foundation for Educational Research.

Takayama, K. (2008) The politics of international league tables: PISA in Japan's achievement crisis debate, ***Comparative Education,*** 44(4), 387-407.

Talbi, M.T. (2003) The demand of a task, ***International Journal of Mathematical Education in Science and Technology,*** 34(4), 501-526.

Taverner, S. & Wright, M. (1997) Why go modular? A review of modular A-level Mathematics, ***Educational Research,*** 39(1), 104-112.

UK Statistics Authority (2012) ***Letter from Andrew Dilnot to David Miliband.*** Available online at: http://www.statisticsauthority.gov.uk/reports---correspondence/correspondence/index.html, last accessed January 2013.

Van der Linden, W.J., Glas, C.A.W. (2000) (Eds) ***Computerized Adaptive Testing: Theory and Practice.*** Netherlands, Kluwer Academic Publishers.

Vidal Rodeiro, C.L. (2012) An investigation on the impact of GCSE modularisation on A level uptake and performance, ***Cambridge Assessment Research Matters,*** 14, June.

Vidal Rodeiro, C.L. & Nádas, R. (2010) ***Effects of modularisation.*** Cambridge, Cambridge Assessment.

Vidal Rodeiro C.L & Nádas R. (2012) Effects of modularity, certification session and re-sits on examination performance, ***Assessment in Education: Principles, Policy & Practice,*** 19(4), 411-430.

Wagemaker, H. (2008) Choices and tradeoffs: reply to McGaw. ***Assessment in Education: Principles, Policy & Practice,*** 15(3), 267-278. Special Issue: International Comparative Studies in Achievement.

Wheadon, C.B. (2011) ***An Item Response Theory Approach to the Maintenance of Standards in Public Examinations in England.*** Doctoral thesis, Durham University, available online at: http://etheses.dur.ac.uk/615/1/Chris_Wheadon_PhD.pdf?DDD29+.

Wheadon, C. & Béguin, A. (2010) Fears for tiers: are candidates being appropriately rewarded for their performance in tiered examinations?, ***Assessment in Education: Principles, Policy and Practice,*** 17(3), 287-300.

Wikström, C. (2005) Grade stability in a criterion-referenced grading system: the Swedish example, ***Assessment in Education: Principles, Policy & Practice,*** 12(2), 125-144.

Wilde, S., Wright, S., Hayward, G., Johnson, J. & Skerrett, R. (2006) ***Nuffield Review Higher Education Focus Groups: Preliminary Report.*** Available online at: http://docstore.mch-net.info/nuffield.pdf, last accessed January 2013.

Wiliam, D. (2001) Reliability, Validity and All That Jazz, ***Education*** 3-13, 29(3), 17-21.

Wilmut, J. (2005) ***Experiences of summative teacher assessment in the UK***. A review conducted for the Qualifications and Curriculum Authority.

Wilmut, J. & Tuson, J. (2004) ***Statistical moderation of teacher assessments.*** A report to the Qualifications and Curriculum Authority. London: QCA.

Wolf, A. (2011) ***Review of vocational education***. The Wolf Report. Department for Education. DFE-00031-2011, March. Available at: https://www.education.gov.uk/publications/standard/publicationDetail/Page1/DFE-00031-2011.

Appendix A: Exam board studies on differentiation [8]

A. Joint 16+ pilot studies to first GCSE examinations

Schools Council
1975. Examinations at 16+: proposals for the future, report of the Joint Examinations Sub-Committee of the School Councils

Backhouse, J.K.
1976. Determination of grades for two groups sharing a common paper, Educational Research, 18, pp126-133

Wood, R
1978. Placing candidates who take different papers on the same mark scale, Educational Research, 20, pp210-215

Cresswell, M.J.
1982. Some Possible Approaches to the Problem of Examining Across a Wide Range of Ability: A Discussion of the New 16+ Examinations, Curriculum, 4, 2, pp38-44

Kingdon, J.M. et al.
1983. Awarding grades on differentiated papers in school examinations at 16+, Educational Research, 25, 3, 220-229

Tattersall, K.
1983. Differentiated examinations a strategy for assessment at 16+, Schools Council Examinations Bulletin 43

Cresswell, M.J.
1986. Differentiated papers – an analysis of some grading issues, AEB, NOT PUBLISHED

Good, F.J. & Cresswell, M.J.
1988. Grade Awarding Judgements in Differentiated Examinations, British Educational Research Journal, 14, 3, pp263-281

Good, F.J. & Cresswell, M.J.
1988. Can Teachers Enter Candidates Appropriately for Examinations Involving Differentiated Papers? Educational Studies, 14, 3, pp289-297

Good, F.J. & Cresswell, M.J.
1988. Placing candidates who take differentiated papers on a common grade scale, Educational Research, 30, 3, pp177-189
Good, F.J. & Cresswell, M.J.
1988. Grading the GCSE, SEC

NEA
1988. Approaches to differentiation in the GCSE, based upon data from four subjects from the NEA Joint GCE/CSE examinations, June 1987, NOT PUBLISHED

Good, F.J.
1989. Setting Common Examination Papers that Differentiate, Educational Studies, 15, 1, pp67-82

B. From first GCSE examinations to revised examinations in 1994

SEAC funded projects on setting effective examination papers (SEEP):

WJEC
1992. Setting GCSE English papers which differentiate effectively, a study organised by WJEC on behalf of IGRC

MEG
1992. Setting GCSE Science papers which differentiate effectively, a study organised by MEG on behalf of IGRC

NEA
1992. Setting GCSE Mathematics papers which differentiate effectively, a study organised by NEA on behalf of IGRC
Other:

IGRC / SEAC (Elwood, J. et al.)
1992. Differentiation in GCSE Mathematics. Part A and Part B Centres' entry decision-making policy, UCLES for IGRC

SEAC
(Stobart, G. et al.) 1992. Differential performance in examinations at 16+: English and Mathematics, ULEAC/NFER for SEAC

MEG
1993. Grading MEG Extended level candidates under the SEG scheme (Mike Lewis with explanatory note from Helen Patrick)

C. From 1994 to revised GCSE examinations in 1998

NEAB
1997. The effectiveness of (written) Paper 3 and Coursework in differentiating among male and female candidates in GCSE Business Studies, 1996 examination, NOT PUBLISHED

NEAB
1997. The effect of tiering on awards at A*; an illustration using artificial data, NOT PUBLISHED

NEAB
1997. The possible effect of tiering on awards at A* in GCSE Science, NOT PUBLISHED

Massey, A.J.
1997. The feasibility of equating national test standards in science between Key Stages 2 & 3 and from year to year, Educational Review, 49, 1, pp29-45

Pollitt, A.
1997. Analyses of MEG Salter's Double Science, June 1995, MEG, NOT PUBLISHED

Taylor, M.
1997. Aggregation of marks in tiered subjects, SEG, NOT PUBLISHED

Ganson, H. and Collins, J.
1997. Standard Grade 'Near Miss' Analyses, 1997, SQA, NOT PUBLISHED

Cresswell, M.J.
1997. Tiers at Grade C, SEG, NOT PUBLISHED

[8] Adapted from a list produced by Dee Fowles in 1998 for the Northern Examinations Board.

Oxford University Centre
for Educational Assessment

Oxford University Centre for Educational Assessment
Department of Education
University of Oxford
15 Norham Gardens
Oxford
OX2 6PY

Phone: +44 (0)1865 284098/274002
Fax: +44 (0)1865 274027
Email: admin.oucea@education.ox.ac.uk

http://oucea.education.ox.ac.uk