

Assessing pupils at the age of 16 in England – what is the best approach for effective examinations?

Dennis Opposs, Qingping He, Matt Glanville and Angela Deavall

Office of Qualifications and Examinations Regulation, Coventry CV5 6UB, UK

Dennis.Opposs@ofqual.gov.uk; Qingping.He@ofqual.gov.uk

Matthew.Glanville@ofqual.gov.uk; Angela.Deavall@ofqual.gov.uk

Abstract

In England, pupils aged 16 take the General Certificate of Secondary Education (GCSE) examinations for a range of subjects. The current assessment models for GCSE include a two-tier structure for some subjects and a non-tier model for the others. The tiered subjects have a higher tier designed for high achieving pupils and a lower tier for low achieving pupils. The higher tier paper is targeted at grades A*-D (with A* the highest grade available), while the lower tier paper at grades C-G (with G the lowest grade). In its proposed GCSE reforms, the government suggested that, with tiered papers, pupils are forced to choose between higher and lower tier papers, which will place a cap on the ambition of those entering for the lower tier. It therefore suggests avoiding tiering in the reformed GCSEs when possible. This paper discusses the technical and equity issues with the use of tiered examinations in current GCSEs and explores alternative assessment approaches for effective differentiation between pupils for the reformed GCSEs.

Keywords: General Certificate of Secondary Education, tiering, differentiated papers, grading

1 Introduction

The national assessment systems in England

In England, the education system is divided into four key stages (KS) for pupils aged 5-16 (KS1-KS4) and a further two-year post-compulsory education for pupils aged 16-18. The majority of pupils are assessed formally at the end of each stage of education (except for KS3 at age 14, where the tests were abolished in 2009):

- National Curriculum end of key stage tests and teacher assessments in mathematics, English and science at KS1 and KS2 (ages 7 and 11 respectively).
- General Certificate of Secondary Education examinations, end of KS4 (age 16).
- General Certificate of Education Advanced levels (GCE A levels), end of post-compulsory education (age 18).

The tiered assessment structure in current GCSEs

Baird et al. (2001) and Hamer et al. (2013) reviewed the history of the development of GCSEs which were introduced in the late 1980s. Tiering was introduced in the GCSE to enhance positive achievement and effective differentiation by ensuring that, through an examination designed for most of the ability range, all pupils would have the opportunity to demonstrate what they knew, understood and could do. In a tiered system, differentiated question papers are used for tiers targeting different levels of achievement, and the aim is that pupils will find the examination both challenging and suitable, without being frustrated or bored by the questions in the papers which may be too difficult or too easy. The current assessment models for GCSE include a two-tier structure for the majority of the subjects and a non-tier model for some of the subjects (e.g. history). Basic characteristics of tiered examinations include:

- The higher tier is targeted at grades A*-D (with ‘an allowed grade’ E, or a U - unclassified), and the lower tier (the Foundation tier) grades C-G. Grade C is the judgemental grade achievable through both tiers. Table 1 provides a graphic representation of the two-tiered structure of the system.
- Candidates are entered just for one of the two tiers (at the overall qualification level for linear GCSEs and at unit level for modular GCSEs) and can therefore only access the grades restricted to the tier concerned.

Table 1 Structure of the current tiered GCSE subjects

<i>Higher Tier</i>	A*	A	B	C	D	E			U
<i>Foundation Tier</i>				C	D	E	F	G	U

The Government proposed reform of current GCSEs

The government has proposed a comprehensive reform of the Key Stage 4 qualifications. In his policy steer letter to the regulator Office of Qualifications and Examinations Regulation (Ofqual), the Secretary of State for Education has expressed his concerns about the use of tiered examinations in subjects to accommodate the needs of pupils of different levels of attainment and the need to have a new grading scale (Gove, 2013). He argues that with tiered papers, pupils are forced to choose between higher and lower tier papers, and this will place a cap on the ambition of those entering for the lower tier. He expressed his desire to avoid tiering while enabling high quality assessment for pupils at all ability levels and proposed the use of alternative assessment approaches. He also stressed that the reformed GCSEs should prioritise stretching assessment that truly tests the depth and breadth of pupils’ knowledge and ability, with clearer differentiation in performance, particularly for more able pupils.

The purposes of reformed GCSEs are set as follows:

- To evidence pupils’ achievement against demanding and fulfilling content
- To provide a strong foundation for further academic and vocational study
- To provide a basis upon which schools will be held accountable for the performance of all their pupils

The design and development of the assessment will need to take these purposes and objectives of GCSEs into consideration.

Aim of the study

This paper provides a review of the technical and equity issues associated with the use of tiered examinations in current GCSEs and explores alternative assessment approaches for effective differentiation between pupils for the reformed GCSEs.

2 Technical issues with the current tiered GCSEs

The technical issues of grading in tiered GCSE examinations have been investigated extensively by researchers (for example, Good and Cresswell, 1988a, b; Baird et al., 2001).

The floor and ceiling effects associated with tiered examinations

Research suggests that some teachers can encounter some challenges when it comes to choosing the appropriate tier of entry for their pupils (Elwood, 2005). Any particular choice of tiers may not always therefore lead to the optimum position when it comes to determining the exact achievement levels of individual pupils. In extreme situations inappropriate tier entry can even cause pupils not to be awarded grades that reflect their ability as a result of the

restricted range of grades available (Baird et al, 2001). Pupils can suffer by doing badly on a higher tier and not receive a grade at all even though they might have achieved one if they had been entered for the lower tier (the floor effect). Alternatively, other pupils can only gain the highest grade on a lower tier even though they might have received a higher grade if they had been entered for the higher tier (the ceiling effect). Baird et al. (2001) and Wheadon and Bèguin (2010) provided evidence of the ceiling and floor effects in current GCSEs.

Between-tier comparability

Presently, professional judgement has been the primary approach to the comparability of the standards for the overlapping grades between the tiers, aided by some statistical information. However, research has cast doubt about the accuracy of the results from the judgemental approach. For example, Good and Cresswell (1988a) found that examiners tend to grade work based on demanding questions more severely than that based on easier questions (the Good and Cresswell effect).

Baird et al. (2001) found that the performances of candidates on the same grade but in different tiers on common components or questions were not consistent, which might suggest a mismatch of standards between tiers. Such a mismatch could lead to candidates being penalised or advantaged on the basis of which tier they were entered for. Through item response theory (IRT) equating, involving the use of items/questions common to both tiers, Wheadon and Beguin (2010) found evidence which indicated that pupils on the foundation tier of GCSE science papers were being over-rewarded, whereas those on the higher tier were being under-rewarded at the overlapping grades.

3 Impact of ability grouping in schools and tiering in examinations

Ability grouping in schools and its impact

Ability grouping in schools has been a subject of debate for a long time (see Ireson and Hallam, 2009). Research suggests that schools in England show a wide range of grouping practices that vary with age of pupils (Kutnick et al., 2005; Ireson, 2008). Results from research also suggest that no one form of organisational grouping benefits all pupils (see Kutnick et al., 2005; Ireson, 2008). Kutnick et al. (2005) indicated that in ability-based grouping, pupils in lower groups could be vulnerable to making less progress, becoming demotivated and developing anti-school attitudes. They suggested that these pupils could experience poorer quality of teaching and a limited range of curricular and assessment opportunities likely to have an impact on later life chances.

Potential impact of tiering on teaching, learning and the curriculum

Oates (2013) discusses how the behaviour of schools adapts when the structure of qualifications change. To an extent, ability grouping of pupils in school should facilitate the entry of pupils for particular tiers in a tiered examination. There has been research investigating the link between tiering in examinations and teaching practice such as ability grouping in teaching in schools (e.g. Baird et al., 2013; Baird & Ireson, 2001; Ireson et al., 2005; Elwood and Murphy, 2002). Elwood (2005) raised concerns about the inequity of tiering practices supporting research already conducted by Gillborn and Youdell (2000) and Elwood and Murphy (2002) related to the misrepresentation of boys' and girls' achievements through decisions surrounding allocation to particular tiers of entry.

Wheadon and Bèguin (2010) noted that the tiering process itself could be characterised as unjust. Pupils labelled as foundation level can be demotivated by the labelling process and placed in lower ability sets with restricted access to the curriculum (Baird et al. 2001; Baird and Ireson 2001; Elwood 2005; Elwood and Murphy 2002; Gillborn and Youdell 2000).

Wheadon and Bèguin (2010) suggested that such inequity can be compounded for foundation tier pupils whose achievement is capped at grade C, no matter how well they perform on the examination. Ireson et al. (2005) suggested that pupils of similar ability achieve higher GCSE grades when they are placed in higher ability sets.

Baird and Ireson (2001) found that although tiering does not drive ability grouping in schools, ability grouping and tiering are associated as many teachers allocate whole teaching groups to specific tiers, particularly in mathematics. Further, they found that few pupils changed ability groups over the course of their GCSE study. When this happens, candidates' grade expectations are largely determined early on in the course of study.

Other factors that could affect pupils' learning and development

It is to be noticed that GCSE results are also used for school accountability purposes, and this has implications for school practices as revealed by the evidence from the high stakes assessment impact study conducted by Ofqual (He et al., 2013), the 2012 GCSE English awarding (Ofqual, 2012), and the research carried out by Ofsted (Ofsted, 2012a,b). These studies suggest that some schools deploy certain strategies to improve exam performances and in many cases focus too much on topics likely to be tested. Some schools also provide additional teaching time and other resources to help pupils who are just below the grade C (conceived to be the pass grade) boundary and encourage pupils to study subjects believed to be easy to get a C grade. Oates (2013) discusses similar issues.

4 Considerations for appropriate assessment approaches for reformed GCSE

Assessment plays an important role in the education system. Effective assessment should facilitate teaching and learning and enhance the school curriculum. The following major factors need to be taken into consideration when designing and developing assessment for reformed GCSEs. A balance has to be reached between the various competing elements.

Purposes of qualifications

In line with the purposes of the reformed GCSEs set by the government, the aim of a course leading to a qualification should be to help learners acquire the required knowledge and skills within a specified domain of content and skills, and the purpose of the assessment itself is to provide an accurate measurement of the level of attainment or proficiency that a learner has achieved at the end of the course of study. To allow pupils to fully demonstrate what they know, understand and can do, the tasks that the pupils meet should be appropriately challenging and demanding for their level of achievement.

The technical quality of the assessment and practical constraints

The assessment should possess high technical quality, including validity, reliability, comparability, and minimal bias. The assessment should generate results that provide a valid and reliable measure of the required knowledge, skills and understanding as specified by the assessment objectives. Results from different test sessions should be comparable in standards. If grades are defined for the curriculum, then the standard should be the same for the same grade regardless of the route which is taken to achieve it. The assessment should minimise bias, differentiating learners only on the basis of their ability to meet the relevant learning outcomes. The assessment should also be manageable.

The nature of the subject

Baird et al. (2001) suggested that the nature of a subject's content and skills is a fundamental consideration in ensuring how differentiation can be most effectively realised. Different approaches to differentiation for different subjects may be required.

The impact of assessment approaches on teaching, learning and the curriculum

As discussed previously, as an important component of the education system, assessment can impact upon the way in which teaching and learning are carried out in schools. Effective assessment should promote effective teaching and learning and minimise any potential negative consequences.

5 A review of different forms of differentiation in examinations

Baird et al. (2001) discussed four major forms of differentiation that could be used to provide opportunities for pupils across a range of attainment to demonstrate what they know, understand and can do. These include:

- *Common papers*: All pupils take the same papers which are targeted at the full range of grades available, regardless of their ability. Pupils can therefore access the full range of grades with equal opportunities.
- *Core plus extension paper*: All pupils take the core paper which is targeted at the lower grades and more able pupils take the extension paper which is targeted at the higher grades. Only pupils who take the extension paper can access the higher grades.
- *Tiered papers (the current tiered GCSE model)*: Pupils can enter for one of the tiers which are targeted at different ranges of available grades, and different tier pupils take different papers with access to the corresponding range of grades.
- *The adjacent levels model (the Scottish Standard Grade examinations)*: Pupils can enter for two adjacent levels (or tiers) and papers differ for different levels. Each level is targeted at two adjacent grades and there are no overlapping grades between levels. Pupils retain the grades obtained from the higher level of the two levels entered for.

Table 2 summarises the advantages and disadvantages of the different models, in terms of effectiveness in differentiation, technical quality (including validity, reliability and comparability), curriculum backwash effects, accessibility and inclusion, and efficiency.

Table 2 The features, advantages and disadvantages of different models of differentiation in examinations

<i>Model</i>	<i>Features</i>	<i>Advantages and disadvantages</i>
Model 1: <i>Common papers</i>	<ul style="list-style-type: none"> • All pupils take the same papers, regardless of their abilities. • Differentiation occurs within, not between, the papers and can be achieved either by outcome or by task. • In the case of differentiation based on outcome: <ul style="list-style-type: none"> ○ Questions are of neutral difficulty and accessible to pupils across the range of abilities ○ Questions can admit a range of possible responses which are marked according to their quality ○ Mark scheme categorises responses in a number of performance levels which 	<p><i>Advantages</i></p> <ul style="list-style-type: none"> • It can embody the aspiration of rewarding positive achievement for all pupils. • There is no need to ensure comparability of standards of the same grade awarded via differentiated papers. • All grades are available to all pupils • There is no need to characterise pupils by ability for teaching purposes early in the course of study. • Manageability and cost: this approach is effective <p><i>Disadvantages</i></p> <ul style="list-style-type: none"> • It is difficult to design papers that cover the syllabus adequately in terms of breadth of content and depth of demand, such that pupils with different abilities are able to fully demonstrate their achievements, particularly those at the extremes of the ability spectrum, given that there are practical constraints such as limited testing time. • It is difficult to write clear and unambiguous questions with appropriate wording so that pupils across the ability range can access and respond to the highest level they are

	<p>are hierarchical and descriptive of the type of response expected at each performance level</p> <ul style="list-style-type: none"> • In the case of differentiation by tasks: Questions are set on an incline of difficulty so that less able pupils can complete early easier questions in the paper and more able pupils can complete more questions or all of the paper. • This approach is generally more appropriate for subjects that place emphasis on the development of particular skills that pupils are expect to demonstrate on the basis of content or stimuli which are equally accessible to pupils across the range of abilities. 	<p>capable of.</p> <ul style="list-style-type: none"> • In the case of differentiation by outcome: It is difficult to design and apply a mark scheme which can both generate reliable marking and also signal to pupils what is rewarded and by how much. • In the case of differentiation by task: <ul style="list-style-type: none"> ○ Use of examination time is inefficient, - pupils are expected to answer questions which are either too difficult (for less able pupils) or too easy (for more able pupils) ○ Given the practical constraints such as limited testing time and the use of evidence drawn from only parts of the overall assessment for differentiating pupils, the reliability and validity could be reduced. • Both low ability and high ability pupils may have a demoralising and demotivating experience.
<p>Model 2: Core plus extension paper</p>	<ul style="list-style-type: none"> • In this approach, all pupils take a core paper for which there is a limited range of grades available (typically C-G in GCSEs). • An optional extension paper, which gives access to the higher grades, is available for more able pupils. • The core paper and the extension paper may have a number of overlapping grades (being the top grades of the core paper and bottom grades of the extension paper). 	<p><i>Advantages</i></p> <ul style="list-style-type: none"> • This model embodies the advantages of differentiated papers with respect to increased reliability and validity. • It removes the floor effects and minimises ceiling effects associated with the tiered approach. • Extension paper pupils are not unduly penalised for unexpected poor performance on the paper. • It enables more able pupils to be effectively differentiated through the extension paper. • It has the potential for resolving the comparability of any overlapping grades. • If there are no overlapping grades, the issue of comparability of identical grades from different routes will not exist. • It can retain a positive motivational effect on pupils throughout the two year of the course because the possibility of being entered for the higher level paper remains open for longer than a tiered system. <p><i>Disadvantages</i></p> <ul style="list-style-type: none"> • Similar to the common papers approach, it could represent inefficient use of scarce examination time, particularly for many of the able pupils who will also have to complete the core paper which could be too easy for them. • It is difficult to develop the core paper that is appropriately challenging to pupils across the full ability range. • It encourages many inappropriate entries as pupils have nothing to lose by entering for the extension paper • An appropriate set of rules that are used to grade the extension paper pupils is needed. • Manageability and cost: the administrative and financial burden to centres and mental burden on pupils are substantially more than the common papers and tiered papers approaches.
<p>Model 3:</p>	<ul style="list-style-type: none"> • Pupils enter one of the tiers. 	<p><i>Advantages</i></p>

<p>Tiered papers</p>	<p>Each tier targets a specific range of ability and has a restricted range of grades available to pupils who are entered for it.</p> <ul style="list-style-type: none"> • Papers for different tiers are different in terms of content and skills assessed, although they may contain common questions. • There are overlapping grades between different tiers. • This approach is more suitable for subjects for which the content and skills that pupils need to learn can be delineated and constructed in a structured way. 	<ul style="list-style-type: none"> • The tiered approach is efficient in terms of use of examination time and discrimination of pupils across the ability range. • Tiered papers are targeted at the appropriate level of demand and difficulty so that all pupils should have a satisfying experience. • This approach can maximise the opportunity for positive achievement for all pupils. Pupils at different ability levels are given a reasonable chance to demonstrate what they know, understand and can do. • This approach can improve reliability and validity <p><i>Disadvantages</i></p> <ul style="list-style-type: none"> • This approach relies upon teachers' ability to accurately predict pupils' potential examination performance. • This approach can result in demotivation of pupils who are entered for lower tiers • This approach, coupled with the centre teaching practice in grouping pupils into different ability groups, could produce negative effect on learning opportunities for some pupils (negative curriculum backwash effects) • There is potential for entry decisions to be made not based solely on pupils' abilities • The ceiling and floor effects as a result of restricted range of grades available for individual tiers and inaccurate tier entry decisions could prevent pupils achieving the results they deserve. • As tiered papers are likely to assess different aspects of the syllabus content and weigh the assessment objectives differently, the independent grading of the tiers can make the meaning of the grades awarded more difficult to interpret, particularly the overlapping grades. • It is difficult to maintain and interpret the comparability of standards across tiers for the overlapping grades. The judgement approach used for setting comparable standards for the overlapping grades between tiers has been found to produce the Good and Cresswell effect. Although common questions between tiered papers are currently being used for some subjects by some awarding bodies, a satisfactory statistical approach for maintaining the between-tier comparability is yet to be established.
<p>Model 4: The adjacent levels model</p>	<ul style="list-style-type: none"> • There are three levels (tiers) for most subjects with restricted grade ranges: Credit (grades 1 and 2), General (grades 3 and 4) and Foundation (grades 5 and 6); grade 7 represents "no pass". • There are no overlapping grades between levels/tiers. • Pupils' choice of level is based on the nature of the grade descriptors together with their performance in the school internal assessments. • The syllabus content for each topic is presented in order of difficulty showing which parts can be examined on the 	<p><i>Advantages</i></p> <ul style="list-style-type: none"> • This approach appears to embody the advantages associated with the full tiering model and the core plus extension paper model. • Because there are no overlapping grades between levels, there is only a single route to each grade and therefore there is no issue of ensuring comparability between grades awarded from different routes. • Because each grade has unique meaning, papers can be designed more accurately to reflect the grade descriptors or criteria, resulting in enhanced validity. • The ceiling and floor effects are minimised as pupils can enter for two adjacent levels since performance on a lower level is discounted if they gain a grade on a higher level. • This approach can minimise the potential negative curriculum backwash effects • The targeting of content and skills to specific levels

	<p>three levels of paper.</p> <ul style="list-style-type: none"> • Higher level papers build on and may contain content from lower levels. • Summary and detailed grade related criteria are constructed. • Most pupils enter for two adjacent levels and results from these pupils are also used to align standards between the levels (particularly to ensure the borderlines at grades 2 and 4 are of a higher standard than those at grades 3 and 5). 	<p>enhances positive achievement by allowing pupils from across the ability range an equal opportunity to fully demonstrate their ability, representing more effective use of examination time.</p> <ul style="list-style-type: none"> • The validity of the use of the equi-percentile method or other means of equating for aligning standards of adjacent grades between levels is enhanced because the samples of pupils on which the mark equivalences are made are more typical of the population. • There are potential savings in terms of the number of question papers that need to be prepared. • There is potential for making good use of the available test data for constructing common score scales for grading and enhancing the validity and reliability of results. <p><i>Disadvantages</i></p> <ul style="list-style-type: none"> • Manageability/administrative implications: As candidates tend to enter for two adjacent levels or tiers, only the Foundation and Credit levels can be timetabled concurrently, suggesting that the examination schedule could be twice that of the current GCSEs. • The development of detailed grade related criteria is required, which would be challenging. • This approach represents a three-tier system and the standards setting procedures could be more complicated than other models if all data were to be used. • Cost associated with this model could be higher than the common papers model and the tiered papers model
--	--	--

6 Discussion

As discussed above, the different models of differentiation have different strengths and weaknesses and may be suitable for particular subjects of the reformed GCSEs.

Common papers

The main advantage of this model is that candidates are not required to choose between tiers and can access all the grades available. This model appears to be appropriate for subjects where the differences in demand and content and skills across the full attainment range are relatively small and differentiation between pupils can be achieved by outcomes. This model has been used successfully for a range of subjects, including History and Art and Design.

Core plus extension paper

The main advantages of this model include minimising the floor and ceiling effects associated with the current GCSE tiering model, avoiding the problem of between-tier comparability, allowing adequate discrimination at the top of the ability range, and minimising any potential negative curriculum backwash effects. The core can be targeted at pupils of middle and low levels of attainment, while the extension paper at the top end pupils.

Tiered papers (the current tiering approach)

The main advantages of the existing tiering approach are the potential for enhanced reliability and validity. Given the issues of this approach discussed previously, particularly the requirement that candidates can only enter for one of the two tiers, the floor and ceiling effect due to inappropriate tier entry decisions, and the difficulty in maintaining between-tier comparability for the overlapping grades, this model needs to be revised in terms of the

number of overlapping grades and their position in the overlap should it be used for the reformed GCSEs. The use of grade C as the bottom grade for the higher tier and the top grade for the foundation tier that contributes towards school performance measures could potentially cap the ambition of those prepared and entered for the foundation tier.

If we accept that there can be questions aimed specifically at the overlapping grades, these can form some kind of “core” with lower grades representing an extension for the less able pupils and upper grades an extension for the more able pupils. Such a structure would also allow the common questions to be used to make the alignment of standards between the two tiers more reliable and valid (through some kind of equating which could eliminate the Good and Cresswell effect, although this would require less emphasis on examiner judgement and more on statistical approaches).

The adjacent levels model

The adjacent levels model could be viewed as one variant of the core plus extension paper model, but it extends both upwards for high achieving pupils and downwards for low achieving pupils. Assuming there are three levels, the middle level will be taken by most pupils (except the highest and lowest achieving pupils) and therefore acts as the core. The top level would be an extension for higher achieving pupils, while the bottom level for low achieving pupils. A lower level acts as a 'safety net'.

The main advantages of the adjacent levels model include minimising the floor and ceiling effects associated with the current GCSE tiering approach, no need to ensure the comparability of overlapping grades, and minimising any potential negative curriculum backwash effects. This model, or a variant of it, would seem particularly suitable for subjects where the differences in demand and content and skills across the full attainment range of the pupils are large and differentiation between pupils though outcomes will not be effective.

A variant of the adjacent levels model with a different number of levels or different number of grades in different levels may be also explored. A two levels model would however in effect reduce to the common papers model.

Exploring alternative assessment models

With the rapid development of computer-related technologies, alternative assessment models such as multi-stage testing, on-demand testing, and adaptive testing for some subjects could also be explored (Mead, 2006; Luecht et al., 2006; He, 2012)

7 Conclusions

Given the wide range of achievement in the target population, the requirement of challenging and demanding questions for all pupils (those at the top end of the achievement range in particular), the reliability and validity of results, the potential backwash effects on teaching and learning, and the practical constraints, the design and development of an effective assessment for each subject that can provide an accurate measure of the achievement of individual pupils remains a great challenge.

References

Baird, J., A. Ahmed, T. Hopfenbeck, C. Brown and V. Elliott (2013). Research evidence relating to proposals for reform of the GCSE. OUCEA, University of Oxford. Available online at: <http://oucea.education.ox.ac.uk/wordpress/wp-content/uploads/2013/04/WCQ-report-final.pdf>.

- Baird, J.-A., A. Fearnley, D. Fowles, B. Jones, E. Morfidi, and D. White (2001). *Tiering in the GCSE: A study undertaken by AQA on behalf of the Joint Council for General Qualifications*. London: Joint Council for General Qualifications.
- Baird, J.-A., and J. Ireson (2001). *Teachers' views on tiering and ability grouping at GCSE*. AQA Internal Report. Guildford: AQA.
- Burghes, D., M. Roddick and F. Tapson (2001). Tiering at GCSE: is there a fairer system? *Educational Research* 43, 175-187.
- Elwood, J. (2005). Gender and achievement: What have exams got to do with it? *Oxford Review of Education* 31, 373-93.
- Elwood, J. and P. Murphy (2002). Tests, tiers and achievement: Gender and performance at 16 and 14 in England. *European Journal of Education* 37, 395-416.
- Gillborn, D. and D. Youdell (2000). *Rationing education: Policy, practice, reform and equity*. Buckingham: Open University Press.
- Good, F. and M. Cresswell (1988a). Grade awarding judgements in differentiated examinations. *British Educational Research Journal* 14, 263-80.
- Good, F. and M. Cresswell (1988b). Placing candidates who take differentiated papers on a common grade scale. *Educational Research* 30, 177-189.
- Gove, M. (2013). Ofqual policy steer letter: reforming Key Stage 4 qualifications. Letter to Ofqual on February 06 2013.
- Hamer, J., R. Murphy, T. Mitchell, A. Grant and J. Smith (2013). English Baccalaureate Certificate (EBC) proposals: examining with and without tiers (24/01/13). Confidential report to Pearson/Edexcel. AlphaPlus Consultancy.
- He, Q. (2012). On-demand testing and maintaining standards for general qualifications in the UK using item response theory: possibilities and challenges. *Educational Research* 54, 89-112.
- He, Q., M. Glanville, L. Rodriguez and D. Opposs (2013). Perceived impact of high stakes national tests and public examinations in England. Coventry, England: Ofqual.
- Ireson, J. (2008). *Learners, learning and educational activity*. Foundations and Futures of Education. Abingdon, England: Routledge.
- Ireson, J. and S. Hallam (2009) Academic self-concepts in adolescence: Relations with achievement and ability grouping in schools. *Learning and Instruction* 19, 201-213.
- Ireson, J., S. Hallam and C. Hurley. (2005). What are the effects of ability grouping on GCSE attainment? *British Educational Research Journal* 31, 443-58.
- Kutnick, P., J. Sebba, P. Blatchford, M. Galton and J. Thorp (2005). The effects of pupil grouping: literature review. Available at: <https://www.education.gov.uk/publications/eOrderingDownload/RR688.pdf>
- Luecht, R., T. Brumfield, & K. Breithaupt (2006). A Testlet Assembly Design for Adaptive Multistage Tests. *Applied Measurement in Education* 19, 189-202.
- Mead, A. (2006). An Introduction to Multistage Testing. *Applied Measurement in Education* 19, 185-187.
- Oates, T. (2013). Tiering in GCSE – which structure holds most promise. Available online at: http://www.cambridgeassessment.org.uk/ca/digitalAssets/207172_TieringInGCSE_TimOates_250413_FINAL.pdf
- Office for Standards in Education (Ofsted) (2012a). Mathematics: made to measure. Available online at: <http://www.ofsted.gov.uk/resources/mathematics-made-measure>.
- Office for Standards in Education (Ofsted) (2012b). Moving English forward. Available online at: <http://www.ofsted.gov.uk/resources/moving-english-forward>.
- Office of Qualifications and Examinations Regulation (Ofqual) (2012). GCSE English 2012. Available online at: <http://www.ofqual.gov.uk/files/2012-11-02-gcse-english-final-report-and-appendices.pdf>.
- Wheadon, C. & A. Bèguin (2010). Fears for tiers: are candidates being appropriately rewarded for their performance in tiered examinations? *Assessment in Education* 17, 287-300.