

## **Annex 4 – Technical Paper on Tiering**



June 2013

Ofqual/13/5294

# Contents

Purpose of tiering .....	2
Approaches to tiering.....	2
The features of untiered forms of differentiation in exams .....	3
The features of three different tiered forms of differentiation in exams .....	3
Evidence on tiering .....	4
History of tiering .....	4
The Good and Cresswell effect and comparison between tiers .....	5
Decision making.....	6
Backwash effects .....	6

## Purpose of tiering

In a tiered system, differentiated question papers (tiers) are used for targeting different levels of achievement, so that candidates will find the exam both challenging and suitable, without being disadvantaged by questions in the papers that may be too difficult or too easy. A tiered approach can also enhance the reliability and validity of exam results by focusing assessment at the appropriate boundaries for the candidate.

It is recognised that tiering can also have negative effects, especially in an environment where school accountability measures are high stakes. The choice of which tier to enter can be influenced by a desire to achieve a “safe grade C”. Furthermore, students might not study the full range of the curriculum, only the parts related to the tier they will be entered for, and this can prevent them being able to continue their studies at A level.

## Approaches to tiering

There are generally two approaches to assessing candidates, either a broad question is set and the candidates demonstrate their ability through their answers (discrimination by outcome<sup>1</sup>), or the question is specifically designed to provide a particular evidence at a particular level of difficulty (discrimination by task).

The choice of which approach to take broadly depends on the subject. For example, mathematics questions differentiate by task whereas history questions differentiate by outcome. Other subjects fall in between these two extremes.

In very simplistic terms, assessments that differentiate by task have more need to ask the candidates different questions, depending on their ability. This tends to suggest a tiering model for these assessments to avoid unreasonably long exams with questions that cover all possible standards. Assessments that differentiate by outcome are well served by untiered exams.

---

<sup>1</sup> Note that some subject experts challenge the assertion that questions can be targeted in this way, especially when the level of difficulty is aligned with particular grades.

## The features of untiered forms of differentiation in exams

Form of differentiation	Features
Common papers (the current non-tiering model)	<ul style="list-style-type: none"> <li>■ Every paper assesses the full range of attainment, accessing the full range of available grades, A* to G.</li> <li>■ All candidates take the same papers, regardless of their ability.</li> <li>■ Differentiation occurs within, not between, the papers and can be achieved either by outcome or by task.</li> <li>■ In the case of differentiation by outcome:               <ul style="list-style-type: none"> <li>■ Questions are of neutral difficulty and accessible to candidates across the full range of ability.</li> <li>■ Questions can admit a range of possible responses, which are marked according to their quality.</li> <li>■ The mark scheme categorises responses in a number of performance levels that are hierarchical and descriptive of the type of response expected at each performance level.</li> </ul> </li> <li>■ In the case of differentiation by task, questions are set on an incline of difficulty so that less able candidates can complete early, easier questions in the paper and more able candidates can complete more questions or all of the paper.</li> </ul>

There is a range of different approaches to a tiered exam system, and these approaches can be broadly categorised in the following way.

## The features of three different tiered forms of differentiation in exams

Form of differentiation	Features
Model 1: Core plus extension paper	<ul style="list-style-type: none"> <li>■ There is a core paper and an optional extension paper.</li> <li>■ The core paper is for the lower grades.</li> <li>■ The extension paper is for the higher grades.</li> <li>■ All candidates take the core paper.</li> <li>■ The optional extension paper is available for more able candidates.</li> <li>■ The core paper and the extension paper may have a number of overlapping grades (being the top grades of the core paper and the bottom grades of the extension paper).</li> </ul>
Model 2: Tiered papers (the existing GCSE tiering model)	<ul style="list-style-type: none"> <li>■ There are two papers: a foundation tier paper and a higher tier paper.</li> <li>■ The foundation tier paper accesses the lower grades.</li> <li>■ The higher tier paper accesses the higher grades.</li> </ul>

	<ul style="list-style-type: none"> <li>■ Candidates enter just one of the tiers.</li> <li>■ Papers for different tiers are different in terms of the content and skills assessed, although they may contain common questions.</li> <li>■ There are overlapping grades between different tiers.</li> </ul>
<p>Model 3: The adjacent levels model (the Scottish Standard Grade exam – S4)</p>	<ul style="list-style-type: none"> <li>■ 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”.</li> <li>■ There are no overlapping grades between the levels (tiers).</li> <li>■ The candidates’ choice of level (tier) is based on the nature of the grade descriptors together with their performance in the school internal assessments.</li> <li>■ The syllabus content for each topic is presented in order of difficulty, showing which parts can be examined by each of the three levels of paper.</li> <li>■ Higher level papers build on and may contain content from lower levels.</li> <li>■ Most candidates enter for two adjacent levels (tiers) and retain the grade they obtain from the higher level.</li> </ul>

Please refer to section 2 of the consultation document<sup>2</sup> for details of our recommendation and the opportunity to comment on this proposal.

## Evidence on tiering

A wide body of research has been undertaken on tiering. The evidence below highlights some of the papers we have drawn upon.

### History of tiering

Baird et al.<sup>3</sup> (2001) and Hamer et al.<sup>4</sup> (2013) reviewed the history of the development of the GCSE. As noted by Baird et al., tiering was introduced in the GCSE to enhance positive achievement and effective differentiation by ensuring that, through an exam designed for most of the ability range, all candidates would have the

<sup>2</sup> Ofqual (2013) *GCSE Reform Consultation – June 2013*

See [www.ofqual.gov.uk/2013-06-11-gcse-reform-consultation-june-2013.pdf](http://www.ofqual.gov.uk/2013-06-11-gcse-reform-consultation-june-2013.pdf) (accessed 11th June 2013).

<sup>3</sup> Baird, J., Fearnley, A., Fowles, D., Jones, B., Morfidi, E. and White, D. (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.

<sup>4</sup> Hamer, J., Murphy, R., Mitchell, T., Grant, A. and Smith, J. (2013) *English Baccalaureate Certificate (EBC) Proposals: Examining with and without Tiers* (24/01/13). Confidential report to Edexcel, a Pearson company. Manchester, AlphaPlus Consultancy.

opportunity to demonstrate what they knew, understood and could do (Baird et al., 2001).

### **The Good and Cresswell effect and comparison between tiers**

When differentiated papers, such as papers for different tiers, are used, there can be different routes to the same grade (the overlapping grades between tiers). The equivalence of standards of the same grade from the different routes has always been an issue. Presently, professional judgement is the primary approach to the comparability of the standards between 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<sup>5</sup> found that examiners tend to grade work based on demanding questions more severely than that based on easier questions (the Good and Cresswell effect) (Good and Cresswell, 1988a).

There have been attempts to explore the use of statistical approaches to the equivalence of test scores from different papers, which generally involved the same candidates taking two different papers and the use of the relationship between the two sets of scores to establish a common score scale (cf. Backhouse<sup>6</sup>, 1976; Kingdon et al.<sup>7</sup>, 1983; Good and Cresswell<sup>8</sup>, 1998b). This process of establishing a common score scale onto which scores from different tests or exams are converted is termed scaling. Since the two papers are designed for assessing different attainment levels, this type of scaling is also referred to as vertical scaling or vertical equating. Vertical equating involves placing scores from two tests, which are different in difficulty and content but which are intended to measure similar constructs, on the same score scale (see Kolen and Brennan<sup>9</sup>, 2004). Test equating establishes a mathematical relationship between scores from different tests so that they can be used interchangeably, regardless of which test someone has taken.

Common items or questions are frequently used in tiered papers as reference points to examine the relative performance of candidates from different tiers, particularly on overlapping grades between the tiers. Baird et al. suggested that to make effective use of common questions they should have common mark schemes (Baird et al., 2001). Such information can be used to support judgemental comparisons of

---

<sup>5</sup> Good, F. and Cresswell, M. (1988a) *Grade Awarding Judgements in Differentiated Examinations*. British Educational Research Journal 14, 263-80.

<sup>6</sup> Backhouse, J. (1976) *Determination of Grades for Two Groups Sharing a Common Paper*. Educational Research, 18:2, 126-137.

<sup>7</sup> Kingdon, J., French, S., Pierce, G. and Woodthorpe, A. (1983) *Awarding Grades on Differentiated Papers in School Examinations at 16 plus*. Educational Research 25, 220-229.

<sup>8</sup> Good, F. and Cresswell, M. (1988b) *Placing Candidates Who Take Differentiated Papers on a Common Grade Scale*. Educational Research 30, 177-189.

<sup>9</sup> Kolen, M. and Brennan, R. (2004) *Test Equating, Scaling, and Linking: Methods and Practices*. Berlin, Springer-Verlag.

performance on the different tiers. The work by Wheadon and Bèguin<sup>10</sup> and He and Wheadon<sup>11</sup> demonstrated how common questions/items could be used to improve between-tier comparability (Wheadon and Bèguin, 2010; He and Wheadon, 2013).

## Decision making

Research suggests that some teachers can encounter some challenges when it comes to choosing the appropriate tiers of entry for their students<sup>12</sup>. Therefore, any particular choice of tier may not always lead to the optimum position when it comes to determining the exact achievement levels of individual candidates. In extreme situations, inappropriate tier entry can even cause candidates not to be awarded grades that reflect their ability as a result of the restricted range of grades available at individual tiers (Baird et al., 2001). Candidates 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 candidates who gain the highest grade on a lower tier might have received a higher grade if they had been entered for the higher tier (the ceiling effect).

## Backwash effects

Ability grouping in schools has been a subject of debate for a long time (see Ireson and Hallam<sup>13</sup>, 2009). Research suggests that schools in England show a wide range of grouping practices that vary with the age of the students<sup>14</sup>. Results from research

<sup>10</sup> Wheadon, C. and Bèguin, A. (2010a) *Fears for Tiers: Are Candidates Being Appropriately Rewarded for Their Performance in Tiered Examinations?* Assessment in Education 17, 287-300.

<sup>11</sup> He, Q. and Wheadon, C. (2013) *Using the Dichotomous Rasch Model to Analyse Polytomous Items.* Journal of Applied Measurement 14, 44-56.

<sup>12</sup> See, for example:

Elwood, J. (2005) *Gender and Achievement: What Have Exams Got to Do with It?* Oxford Review of Education 31, 373-93.

Burghes, D., Roddick, M. and Tapson, F. (1998) *Report on a Pilot Project for a Non-tiering GCSE in Mathematics.* Centre for Innovation in Mathematics Teaching, School of Education, University of Exeter. Available at: [www.cimt.plymouth.ac.uk/papers/ntgcse.pdf](http://www.cimt.plymouth.ac.uk/papers/ntgcse.pdf)

Burghes, D., Roddick, M. and Tapson, F. (2001) *Tiering at GCSE: Is There a Fairer System?* Educational Research 43, 175-187.

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

Stobart, G., White, J., Elwood, J., Hayden, M. and Mason, K. (1992). *Differential performance Performance in Examinations at 16+: English and Mathematics.* (London, SEAC).

Wheadon, C. and Bèguin, A. (2010b) *Fears for Tiers: Are Candidates Being Appropriately Rewarded for Their Performance in Tiered Examinations?* Assessment in Education 17, 287-300.

<sup>13</sup> Ireson, J. and Hallam, S. (2009) *Academic Self-concepts in Adolescence: Relations with Achievement and Ability Grouping in Schools.* Learning and Instruction 19 (2009), 201-213.

<sup>14</sup> See, for example:

Kutnick, P., Sebba, J., Blatchford, P., Galton, M. and Thorp, J. (2005) *The Effects of Pupil Grouping: Literature Review.* Department for Education and Skills, Nottingham. Available at: [www.education.gov.uk/publications/eOrderingDownload/RR688.pdf](http://www.education.gov.uk/publications/eOrderingDownload/RR688.pdf)

also suggest that no single form of organisational grouping benefits all students (see Kutnick et al., 2005; Ireson, 2008).

To an extent, ability grouping of students in schools should facilitate the entry of candidates for particular tiers in a tiered exam, especially where it supports the progression of students and where individuals who progress well still have the opportunity to learn and achieve more. There has been research investigating the link between tiering in exams and teaching practices such as ability grouping in teaching in schools<sup>15</sup>. For example, Elwood (2005) raised concerns about the inequity of tiering practices, supporting research already conducted by Gillborn and Youdell (2000) and Elwood and Murphy (2002). The observed inequity of tiering practices concerned the misrepresentation of boys' and girls' achievements through decisions surrounding allocation to particular tiers of entry. Elwood noted that more boys than girls were entered for the foundation tier in the GCSE in Mathematics exams, where the maximum available grade was D. She suggested that disaffection amongst boys in GCSE in Mathematics may well be influenced by the restricted grade range on offer at this lower tier (Elwood, 2005). Stobart et al. (1992) reported that teachers considered boys who were placed in the foundation tier to be less motivated, and as a consequence more disruptive, than girls in the same tier. Boys tended to feel that the lower tier was not worth it, whereas girls were often more content to take a lower tier (Stobart et al., 1992).

---

Ireson, J. (2008) *Learners, Learning and Educational Activity. Foundations and Futures of Education*, Abingdon, Routledge. ISBN 978-0-415-41406-7 (paperback), 978-0-415-41407-4 (hardback), 978-0-203-92909-4 (electronic).

<sup>15</sup> See, for example:

Baird, J., Ahmed, A., Hopfenbeck, T., Brown, C. and Elliott, V. (2013) *Research Evidence Relating to Proposals for Reform of the GCSE*. OUCEA, University of Oxford, Oxford.  
<http://content.yudu.com/Library/A24v28/Researchevidencerela/resources/index.htm?referrerUrl=http%3A%2F%2Ffree.yudu.com%2Fitem%2Fdetails%2F837575%2FResearch-evidence-relating-to-proposals-for-reform-of-the-GCSE>

Baird, J. and Ireson, J. (2001) *Teachers' Views on Tiering and Ability Grouping at GCSE*. AQA internal report. Guildford, AQA.

Ireson, J., Hallam, S. and Hurley, C. (2005) *What Are the Effects of Ability Grouping on GCSE Attainment?* British Educational Research Journal 31, 443-58.

Elwood, J. and Murphy, P. (2002) *Tests, Tiers and Achievement: Gender and Performance at 16 and 14 in England*. European Journal of Education 37, no. 4: 395-416.



We wish to make our publications widely accessible. Please contact us if you have any specific accessibility requirements.

First published by the Office of Qualifications and Examinations Regulation in 2013

© Crown copyright 2013

You may re-use this publication (not including logos) free of charge in any format or medium, under the terms of the [Open Government Licence](#). To view this licence, visit [The National Archives](#); or write to the Information Policy Team, The National Archives, Kew, Richmond, Surrey, TW9 4DU; or email: [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk)

This publication is also available on our website at [www.ofqual.gov.uk](http://www.ofqual.gov.uk)

Any enquiries regarding this publication should be sent to us at:

Office of Qualifications and Examinations Regulation	
Spring Place	2nd Floor
Coventry Business Park	Glendinning House
Herald Avenue	6 Murray Street
Coventry CV5 6UB	Belfast BT1 6DN

Telephone 0300 303 3344  
Textphone 0300 303 3345  
Helpline 0300 303 3346