

<b>Institution:</b> University of Oxford		
<b>Unit of Assessment:</b> 23 Education		
<b>Title of case study:</b> Improving the Quality of Educational Assessments		
<b>Period when the underpinning research was undertaken:</b> 2012-2019		
<b>Details of staff conducting the underpinning research from the submitting unit:</b>		
<b>Name(s):</b>	<b>Role(s) (e.g. job title):</b>	<b>Period(s) employed by submitting HEI:</b>
Jo-Anne Baird	Professor	01/10/2011 - Present
Therese Hopfenbeck	Professor	01/01/2016 – Present
Victoria Elliott	Associate Professor	01/09/2014 – Present
Jennifer Ingram	Associate Professor	01/09/2013 – Present
Ian Thompson	Associate Professor	01/09/2011 – Present
Natalie Usher	Research Assistant	17/06/2015 – 01/06/2018
Daniel Caro	Research Fellow	01/11/2012 – 28/02/2018
<b>Period when the claimed impact occurred:</b> 2014 – 31 July 2020		
<b>Is this case study continued from a case study submitted in 2014? N</b>		
<b>1. Summary of the impact</b> (indicative maximum 100 words)		
<p>The research from the University of Oxford Centre for Educational Assessment (OUCEA) has been influential on examination systems, particularly in England, Wales and Ireland. This programme builds upon the work of the centre researchers, which dates back for two decades. The impacts comprise government policy, examination board practices and classroom practices. Specific impacts include changes to how examination standards are set and defined, monitoring of educational standards, marking policies and practices, teacher workload management and the structure of examinations. These changes improved the quality of assessments, and the education systems more broadly, which impacted, and continues to impact, millions of young people taking examinations.</p>		
<b>2. Underpinning research</b> (indicative maximum 500 words)		
<p>This case is an example of embedded impact research practices, in which a combination of research, policy and practice has been a collaborative agenda with national bodies such as the examinations regulator (Ofqual) and the examination boards. A distinctive component of the OUCEA research programme is the focus upon system-wide issues, rather than test-specific topics and methods. The research has influenced examination policy and practice. Additionally, the research programmes of these organisations have been impacted, demonstrating agenda-setting effects of the research upon government and industry thinking. Four aspects of the assessment systems research have had the most impact in this period:</p>		
<p><b>Strand A - Examination standards</b> – this research involved experiments on examiners' judgments, advanced large-scale quantitative modelling of outcomes, qualitative research with teachers and pupils, and policy research. The research (e.g. <b>R1</b>):</p> <ul style="list-style-type: none"> <li>• Introduced new definitions and classification systems for standards definitions (based upon an ecological model) and assessments ('assessment paradigms'),</li> <li>• Demonstrated that quantitative information is essential to the setting of fair standards,</li> <li>• Examined the application of advanced methodological techniques such as multilevel modelling and comparative judgment,</li> <li>• Outlined policy issues in standard setting.</li> </ul>		

**Strand B – Marking** – this research ranges across the selection of examiners, training processes, monitoring processes and comparisons of applications of advanced analytical techniques. The research showed that:

- School marking policies did not often have an evidence base and created a great deal of unnecessary workload (**R2**),
- A wider pool of examiners (such as students with a degree in the subject) could be trained to mark consistently than had previously been thought possible (e.g. **R3**),
- Examiner training was affected by the group in which they were trained (**R3**),
- Instability in examiner severity was demonstrated,
- As part of the Ofqual Reliability Programme (Chaired by Baird) (**R4**), research on the definitions of true scores and reliability showed the complexity of the issues, the need for research on qualifications of the kind used by examination boards internationally and better availability of benchmarks for reliability data.

**Strand C - Examination structures** –this research addressed reform issues such as tiered examinations at GCSE, the number of examination boards and governance structures. Key insights included that the change from modular to linear GCSEs did not impact upon standards or inequalities and that teachers adapted their practices readily (**R5**).

**Strand D - Test preparation** – effects of test preparation for school-leaving and university level examinations have been investigated in mixed method research studies. This research showed:

- Levels of predictability of the Irish Leaving Certificate examinations and ways to introduce higher order thinking skills assessments to avoid predictability and any consequent undermining of the integrity of exam results (**R6**).
- Test preparation influences on the Thinking Skills Assessment (TSA) and British Medical Assessment Test (BMAT).

### 3. References to the research (indicative maximum of six references)

- R1.** Baird, J., Isaacs, T., Opposs, D. & Gray, L. (Editors) (2018). *Examination standards: how measures & meanings differ around the world*. UCL Institute of Education Press. [output type: B – available on request]
- R2.** Elliott, V., Baird, J., Hopfenbeck, T.N., Ingram, J., Thompson, I., Usher, N., Zantout, M., Richardson, J. & Coleman, R. (2016) *A marked improvement? A review of the evidence on written marking*. Education Endowment Foundation.  
[https://educationendowmentfoundation.org.uk/public/files/Presentations/Publications/EEF\\_Marking\\_Review\\_April\\_2016.pdf](https://educationendowmentfoundation.org.uk/public/files/Presentations/Publications/EEF_Marking_Review_April_2016.pdf) [output type: N]
- R3.** Baird, J., Meadows, M., Leckie, G. & Caro, D. (2017) Rater accuracy and training group effects in Expert- and Supervisor-based monitoring systems, *Assessment in Education: Principles, Policy & Practice*, 24, 1, 44-59.  
<https://doi.org/10.1080/0969594X.2015.1108283> [output type: D]
- R4.** Baird, J., Beguin, A., Black, P., Pollitt, A. and Stanley, G. (2012). The Reliability Programme: Final Report of the Technical Advisory Group. Chapter 20. In D. Opposs and Q. He (Editors). *Ofqual's Reliability Compendium*. 771 – 838. Office of Qualifications and Examinations Regulation. Coventry.  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/578895/2011-03-16-the-reliability-programme-final-report-of-the-technical-advisory-group.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/578895/2011-03-16-the-reliability-programme-final-report-of-the-technical-advisory-group.pdf) [output type: N]
- R5.** Baird, J., Caro, D., Elliott, V., El Masri, Y., Ingram, J., Isaacs, T., Pinot de Moira, A., Randhawa, A., Stobart, G., Meadows, M., Morin, C. & Taylor, R. (2019) *Examination Reform: Impact of Linear and Modular Examinations at GCSE* (Oxford, University of Oxford Department of Education). Ofqual/19/6506/2.

<https://www.gov.uk/government/publications/examination-reform-impact-of-linear-and-modular-examinations-at-gcse>. [output type: N]

**R6.** Baird, J., Hopfenbeck, T., Elwood, J., Caro, D. & Ahmed, A. (2015) Predictability in the Irish Leaving Certificate, Report commissioned by the State Examinations Commission, Ireland. <https://www.examinations.ie/about-us/Predictability-Overall-Report.pdf> [output type: N]

#### 4. Details of the impact (indicative maximum 750 words)

##### **Strand A – Impact on Examination Standards in England, Scotland and Wales**

The research has influenced how examination boards and the regulator (Ofqual) conceptualise standards, and been embedded into their practice (**R1**). Ofqual introduced an examination standards monitoring system (comparable outcomes) which has affected GCSE (approximately 5,500,000 per annum) and A-level (over 800,000 per annum) outcomes since 2010. Problems uncovered by Baird's research regarding examiners' judgements of standards led Ofqual to give greater weight to quantitative evidence in standard setting which in turn improved control of grade inflation and competition on examination outcomes between exam boards (**E1**). Ofqual referenced our research stating that "*Experience and research evidence shows it is difficult for awarders to make grade boundary judgements accurately and consistently simply by looking at students' work. Statistical information is crucial to the awarding process.*" (**E2**). Independent evaluation of the impact comes from Cambridge Assessment, who stated that "*Recent explanations by Ofqual of their approach to maintaining standards ... draw heavily upon (and to a large extent directly repeat) a section from a book chapter written by Baird.*" (**E3**). Ofqual's conclusion, that England's standard setting system was world standard and 'as good as any' to be found internationally, was based upon our 2018 collaborative international research programme on standard setting (**R1**) and hence they retained the systems and policies which built upon this research programme (**E4**). Examination boards have based their training materials for examiners upon the research. The Assistant Director of Standards and Processing at WJEC, Wales' largest awarding body, stated that **R1** had been used in the '*training of all Chairs of Examiners on themes relating to awards and standards, and the re-development of Chair's report templates and guidance to ensure better articulation of the reasons behind their grade boundary recommendations*' and for the '*ongoing development of a framework for standards...which will underpin our policies relating to standards in the years ahead*' (**E5**). The theoretical framework of Baird's research is also used in a recent World Bank book aimed at practitioners and researchers (**E6**).

##### **Strand B – Marking: Impact on Teacher Workloads and Marking Standardisation**

The evidence review on marking policies (**R2**) was the Education Endowment Foundation's (EEF) most downloaded evidence report for some time, with 37,285 unique downloads, indicating its influence upon teachers' desire for an evidence-based approach to marking policy in their schools, and fed into an influential study on the effects of different policies upon teachers' workload (**E7, p.7**). Teacher unions have developed policies to tackle teacher workload based upon the review, advocating lighter-touch marking in schools, citing the observation within **R2** that '*the introduction of deep marking as an intervention to raise standards of attainment is unsupported by any meaningful evidence*' (**E8**). DfE and Ofsted have used the report in formulating their approaches to teacher workload (**E9**). Schools and school chains have implemented policies and initiatives that build on the review findings, such as the Southwark Teaching School Alliance project '*Mark Less, Mark Better!*' (**E10a**) and the Flying High Partnership research report '*Reducing teacher workload*' (2018) (**E10b**). The Chartered Teaching College uses it as a reference in their Teaching Programme and have also written a feature on the review in their journal, *Impact* (**E11**). Teach First have used **R2** '*as the foundation of several sessions on written feedback and the implications of research on classroom practice*' (**E12**). It has been cited in practice journals and twice in *The Economist* (**E13**), demonstrating recognition of the wider implications of the work for the teacher workforce.

Ofqual decided that online standardisation of marking was important to the quality of marking, drawing directly on research in **R3** which showed that markers trained face-to-face in teams were influenced by the team leaders' standards (**E2b**). With online standardisation, all markers are trained using the same materials, with much less input from team leaders in the training process. Eradicating the unwanted variation produced by team leaders in this process has been important to improving marking reliability. Ofqual also introduced new procedures for regulation of reliability based upon the reliability programme (**R4**). Examination boards are now required to produce statistics on reliability and produce a rationale for any assessments with low levels of reliability (**E2c**).

### Strand C - Examination structures: Impact on Exams Policy

The Chief Regulator of examinations concluded that linear examinations were more suitable at GCSE than modular structures (**R5 - Foreword written by chief regulator, p2**) and maintained the linear examinations policy based upon the research in **R5**. Benefits of this decision include the avoidance of costs associated with examination reform (this has never been costed but would run into many millions of pounds), as well as the considerable upheaval in the system which can impact examination standards and teacher workload.

### Strand D - Test preparation: Impact on Irish National Leaving Certificate Assessments

The Minister for Education and Skills in Ireland committed to tackling any problematic predictability in the national Leaving Certificate examinations in response to **R6**. Perceived predictability in the examinations had undermined the credibility of the results. Following the research (**R6**), trialling of the introduction of more coursework in the biology, chemistry and physics qualifications as conducted, to allow the demonstration of higher order thinking and developed arguments (documented by the Houses of the Oireachtas Joint Committee on Education and Skills (**E14a**)). A decision on the future of science practical assessment is forthcoming, but the State Examinations Commission research shows beneficial effects of the teaching of practical science upon students' understanding (**E14a**). Ongoing syllabus reviews are introducing more credit for higher order thinking skills (e.g. in in the classics, economics and geography syllabuses) in response to this research (**E14b**).

### Ongoing Impact

Impact is ongoing. For example, Teach First will roll out their Teach First Early Career Framework to 2,500 newly qualified teachers and their mentors in 2021, including the sessions on written feedback developed using **R2 (E12)**.

## 5. Sources to corroborate the impact (indicative maximum of 10 references)

**E1.** House of Commons Education Committee (2013). *The administration of examinations for 15 – 19 year olds in England*. First Report of Session 2012 -13. Volume I.

<https://publications.parliament.uk/pa/cm201213/cmselect/cmeduc/141/141.pdf>

**E2.** Ofqual:

- a) 'Consultation on setting the grade standards of new GCSEs in England' (2014)  
Ofqual/14/5401. Paragraph 2.3, page 11  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/630155/2014-04-03-consultation-on-setting-the-grade-standards-of-new-gcses-in-england.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/630155/2014-04-03-consultation-on-setting-the-grade-standards-of-new-gcses-in-england.pdf)
- b) Ofqual (2018), *Online standardisation: Observations, interviews and survey*,
- c) These details can be corroborated by the Chair of Standards, Ofqual. Details entered into the submission system.

**E3.** Benton, T and Bramley, T. (2015). *The use of evidence in setting and maintaining standards in GCSEs and A levels*. Discussion paper. Assessment Research and Development (ARD). Cambridge Assessment. <https://www.cambridgeassessment.org.uk/Images/459318-the-use-of-evidence-in-setting-and-maintaining-standards-in-gcses-and-a-levels.pdf>

**E4.** TES (September 2018), *Ofqual: 'Our exams system is as good as any'*, <https://www.tes.com/news/ofqual-our-exams-system-good-any>

**E5.** Testimony from the Assistant Director (Standards, Processing and Research), WJEC

**E6.** World Bank (2019), *Public Examinations Examined (English)*, <http://documents.worldbank.org/curated/en/208231572358210940/Public-Examinations-Examined>

**E7.** Kime, S (2018), *Reducing teacher workload: the 'Rebalancing Feedback' trial*, <https://dera.ioe.ac.uk/31210/>

**E8.** NASUWT (2018), *Effective Interventions: Promoting Learning, Tackling Workload*, <https://www.nasuwt.org.uk/uploads/assets/uploaded/68ddda7f-afca-420c-a18798ea9271695f.pdf>

**E9.** Referenced by DfE on <https://www.gov.uk/government/publications/feedback-and-marking-reducing-teacher-workload> and <https://www.gov.uk/government/publications/reducing-teachers-workload/reducing-teachers-workload>

**E10.** Schools and School consortia initiatives:

- a) Southwark Teaching School Alliance, *Mark Less, Mark Better: A How-to Guide to Live Marking*, 2<sup>nd</sup> ed. <https://www.londonsouthtsa.org.uk/perch/resources/mark-less-mark-better.pdf>
- b) Flying High Partnership report, 'Reducing Teacher Workload' (Nov. 2018) [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/687198/Flying\\_High\\_-\\_Reducing\\_teacher\\_workload.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687198/Flying_High_-_Reducing_teacher_workload.pdf)

**E11.** Taylor, M (2017), 'A Marked Improvement: Where Are We Now?', *Impact*, Chartered College of Teaching, <https://impact.chartered.college/article/taylor-marked-improvement-where-now/>

**E12.** Email from the Curriculum Design Manager at Teach First

**E13.** The Economist (2018), *England has become one of the world's biggest education laboratories*, <https://www.economist.com/britain/2018/03/31/england-has-become-one-of-the-worlds-biggest-education-laboratories>

**E14.** Irish National Leaving Certificate Assessments:

- a) Houses of the Oireachtas Joint Committee on Education and Skills (2018), *Report on the ongoing reform of the Leaving Certificate curriculum*, [https://data.oireachtas.ie/ie/oireachtas/committee/dail/32/joint\\_committee\\_on\\_education\\_and\\_skills/reports/2018/2018-10-25\\_report-on-the-ongoing-reform-of-the-leaving-certificate-curriculum\\_en.pdf](https://data.oireachtas.ie/ie/oireachtas/committee/dail/32/joint_committee_on_education_and_skills/reports/2018/2018-10-25_report-on-the-ongoing-reform-of-the-leaving-certificate-curriculum_en.pdf)
- b) These details can be corroborated by the Assistant Head of Examinations and Assessment Division, State Examinations Commission, Ireland. Details entered into the submission system.