

Institution: Loughborough University		
Unit of Assessment: C23 Education		
Title of case study: Transforming educational assessment with comparative judgement		
Period when the underpinning research was undertaken: 2009-2019		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Lara Alcock	Reader in Mathematics Education	2007 – date
Colin Foster	Reader in Mathematics Education	2019 – date
Camilla Gilmore	Professor of Mathematical Cognition	2011 – date
Matthew Inglis	Professor of Mathematical Cognition	2008 – date
Ian Jones	Reader in Educational Assessment	2011 – date
Period when the claimed impact occurred: November 2013 to October 2020		
Is this case study continued from a case study submitted in 2014? No		
1. Summary of the impact (indicative maximum 100 words)		
<p>The marking of student work is plagued by human error. Loughborough University research into applying comparative judgement to educational assessment minimised marking error and resulted in three impacts. First, the non-ministerial government department Ofqual implemented the technique to regulate qualifications, which has increased the fairness of examinations for around 5.5m candidates to date in England and Wales. Second, our research was commercialised by the testing agency No More Marking Ltd., which generated sales of almost £1.5m, and created nine new jobs. Third, our research transformed assessment practice in 2,227 schools in 27 countries, which enabled the novel assessment of learning progression for around 579,400 students and enhanced the professional development of over 10,000 teachers.</p>		
2. Underpinning research (indicative maximum 500 words)		
<p>Marking errors create three systemic problems. First, they lead to ‘tick-box’ assessments comprising short, piecemeal questions that reduce marking errors at the cost of thoughtful, lengthier responses, and that fragment curricular knowledge into isolated facts for rote learning. Second, marking errors mean that students’ grades are unreliable; for some students this unfairly hinders life chances. Third, marking errors mean that standards over time and across different awarding bodies cannot be accurately compared, or equivalence maintained. These latter two problems reduce public confidence in qualifications.</p> <p>Our programme of research began when Jones joined Loughborough in 2011 with a five-year Royal Society Fellowship awarded to investigate why GCSE mathematics examinations were not fit for purpose. Jones found that the requirement to rapidly mark hundreds of thousands of exam scripts resulted in question papers almost entirely made from short, piecemeal questions that minimise marking errors but also reduce validity [R1]. To address this problem Jones led a programme of research that developed a comparative judgement (CJ) technique for assessing mathematical knowledge. The CJ technique takes a radical approach to tackling marking errors by eliminating marking itself. Instead, subject experts decide which of two presented scripts is ‘better’ in terms of a high-level criterion such as ‘conceptual understanding’. Many such binary decisions are collected from several assessors and then fitted to the Bradley-Terry model to produce a score for each student.</p>		

This work involved Alcock (2012-2014), Foster (2017-ongoing), Gilmore (2014-ongoing) and Inglis (2013-ongoing) and led to three key research findings:

1. CJ is reliable (funders: HEFCE, Royal Society ED090015).

To demonstrate the reliability of CJ we developed a statistical measure based on a split-halves technique [R2]. This measure enabled us to demonstrate that CJ produces reliable outcomes when used to score open-ended mathematics assessments, showing that assessments need no longer resort to piecemeal questions to minimise marking error [R1]. Moreover, the reliability of CJ assessment outcomes is robust across age groups from primary [R5] to undergraduate level [R3], and across topics, institutions [R2] and jurisdictions [R5].

2. CJ is a valid assessment tool (funders: HEFCE, No More Marking Ltd., Nuffield Foundation EDU/40389 & EDU/38927, Royal Society ED090015).

Given the dominance of piecemeal test questions in high-stakes assessments, we theorised validity as the reliable assessment of nebulous but important learning outcomes such as problem-solving and conceptual understanding. We investigated the validity of CJ using numerous techniques including expert review [R3], interviews and surveys with students and assessors [R1], comparing the outcomes of different groups of assessors [R4], content analysis of test responses [R5], and investigating the convergence and divergence of CJ outcomes with standardised measures and existing achievement data [R3].

3. CJ can help measure standards over time (funder: AQA).

We developed a CJ-based technique for comparing assessment standards across institutions and time. To demonstrate the technique, we showed that standards in A-level Mathematics have declined since the 1960s [R6].

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

R1 Jones, I. & Inglis, M. (2015). The problem of assessing problem solving: can comparative judgement help? *Educational Studies in Mathematics*, 89, 337-355.

DOI: 10.1007/s10649-015-9607-1

R2 Bisson, M.-J., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education*, 2, 141-164.

DOI: 10.1007/s40753-016-0024-3

R3 Jones, I., Bisson, M.-J., Gilmore, C., & Inglis, M. (2019). Measuring conceptual understanding in randomised control trial studies: can comparative judgement help? *British Educational Research Journal*, 45, 662-680.

DOI: 10.1002/berj.3519

R4 Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education*, 39, 1774-1787.

DOI: 10.1080/03075079.2013.821974

R5 Hunter, J., & Jones, I. (2018). Free-response tasks in primary mathematics: a window on students' thinking. In Hunter, J., Darragh, L. & Perger, P. (Eds), *Proceedings of the 41st Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 41, pp. 400-407). Auckland, New Zealand: MERGA.

<https://eric.ed.gov/?id=ED592426>

R6 Jones, I., Wheadon, C., Humphries, S., & Inglis, M. (2016). Fifty years of A-level mathematics: have standards changed? *British Educational Research Journal*, 42, 543-560.

DOI: 10.1002/berj.3224

Quality. The quality of the research is demonstrated by published papers in leading education journals and competitively awarded grants in excess of £380,000. This income included a prestigious five-year Royal Society educational research fellowship and two grants from the Nuffield Foundation. Output 6 was awarded the *British Educational*

Research Journal's Editor's Choice Award. This award honoured the article that the editors considered to be the strongest paper published in the journal in 2016.

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

Our research into applying comparative judgement to educational assessment resulted in three impacts via pathways including the dissemination of research findings to awarding bodies (Assessment and Qualifications Alliance (AQA); Council for the Curriculum, Examinations and Assessment (CCEA); Edexcel; New Zealand Qualifications Authority (NZQA); Oxford, Cambridge and RSA (OCR)) and invited seminars and international assessment conferences (e.g. AEA-Europe). Another important pathway to impact was high-profile national media coverage (*BBC, Daily Mail, Daily Telegraph, ITV, TES, The Times*), and receiving the *British Educational Research Journal* Editor's Choice Award for the best paper published in BERJ during 2016.

We describe the resulting three impacts in turn.

1. Fairer examinations have benefitted 5.5m candidates over 5 years.

Our novel technique for comparing assessment standards changed the examination system in England and Wales, having been adopted by the Office of Qualifications and Examinations Regulation (Ofqual) since 2015 to increase the fairness of UK examinations for around 1.1m candidates per year [S1]. Ofqual used outcomes from our novel technique to instruct awarding bodies to account for relative difficulty when writing subsequent examination papers, improving fairness both across awarding bodies and over time. Dr Michelle Meadows, the Deputy Chief Regulator at Ofqual, said

"Research conducted at Loughborough has directly impacted the examination system in England and Wales, making it fairer and impacting around 1.1 million candidates per year. Our comparability and awarding work based on Loughborough's research is crucial for ensuring public acceptance of the examination system" [S1].

2. SME No More Marking Ltd. has benefited commercially to the value of £1.5m.

Our research was commercialised in November 2013 when Dr Chris Wheadon, former Chief Examiner of awarding body AQA, set up the small/medium enterprise No More Marking Ltd. (NMM). The purpose of NMM was to develop an online comparative judgement service based on findings from our research [S2]. NMM had created 9 new jobs and generated almost £1.5m in sales as of October 2020 [S3]. Wheadon, CEO of NMM, wrote that

"Loughborough's research on comparative judgement directly led to the development of No More Marking's online assessment platform. ... The development and validation of an inter-rater reliability measure enabled us to convince potential clients that our approach eliminated marker error; this was particularly influential in the growth of our customer base" [S2].

3. Transformed practice in schools has improved the assessment of 579,400 students and enhanced the professional development of over 10,000 teachers.

NMM launched six subscription services for schools that enabled the novel assessment of learning progression in mathematics and English. Within each service, NMM administered tests and provided schools with nationally benchmarked data on learning progression at the student, class and school level. To date, 2,227 schools in 27 countries have subscribed to NMM's services, and these schools received reliable data on the learning progression of around 579,400 students [S3]. Wheadon wrote that Loughborough's research

“enabled the design of our ‘Proof of Progress’ [mathematics] subscription service to schools” [S2].

‘Proof of Progress’ was NMM’s first service, launched in 2015 [S3]. Following this initial success, further services were developed. Wheadon wrote that

“Loughborough’s research into the reliability of comparative judgement within the context of comparing standards across institutions led to the development of further subscription services for assessing writing (Assessing GCSE English; Assessing Primary Writing; Assessing America’s Writing)” [S2].

NMM’s services received high-level endorsement in 2016 when Nick Gibb, the Schools Minister, said that NMM has

“improved the assessment of written work in English” [S4].

Our research into improving the validity of mathematics assessments has particularly benefited students who were previously unfairly disadvantaged by more traditional assessments. For example, one mathematics lead teacher in New Zealand expressed concern about the effect that traditional tests have on

“children of Pasifika and Māori heritage who perform disproportionately poorly due to the cultural inappropriateness of such tests” [S5].

She went on to explain that, in contrast

“comparative judgement enables the use of more open-ended and therefore more valid mathematics assessments, as demonstrated by research conducted at Loughborough University” [S5].

A unique benefit of NMM’s services is that data on the learning progression are inherently moderated and directly comparable across schools. To achieve this, teachers assess the work of students from their own schools and from different schools. This moderation and comparability across schools increased confidence in the assessment data that NMM provided to schools. For example, one Vice Principal wrote that students’

“reaction to the judging has been very positive [and students] are reassured and motivated by the fact their work is not just seen by the class teacher but by teachers across the school and the country” [S6].

To date, NMM’s services have involved over 10,000 teachers making holistic comparisons of the quality of students’ work from across different schools and different year groups. This contrasts with common school practice whereby each teacher assesses their own students using traditional marking. Consequently, NMM’s services have enhanced teachers’ professional development as explained by a prominent Deputy Headteacher, who wrote that the comparative judgement approach

“is helping me to be much better informed about how writing is looking in each cohort compared to traditional assessment methods. ... I also love that I have every teacher looking at writing outside of their year group” [S7].

A Vice Principal wrote that

“the teachers ... were very positive on how the switch to judging had impacted on their working life. They felt better supported, less alone, more reflective of their teaching practice and enjoyed the opportunity to learn from teachers not just in their own school but across the country. One teacher remarked that it is really interesting

when you see that a teacher in another school has taught the pupils a certain approach and you can see it has really worked" [S6].

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

S1 Meadows, M. (2020). Testimonial. Deputy Chief Regulator, Ofqual.

S2 Wheadon, C. (2019). Testimonial. CEO, No More Marking Ltd.

S3 No More Marking Ltd. (2020). Technical report.

S4 Gibb, N. (2016). Speech Delivered to Freedom and Autonomy for Schools - National Association (FASNA). Speech available at

<https://www.gov.uk/government/speeches/the-role-freedom-and-autonomy-has-played-in-school-improvement>

S5 Testimonial from mathematics lead teacher Megan Kanz (2020).

S6 Interview with Vice Principal Emma Hockey (2019). Blog post available at

<https://blog.nomoremarking.com/judging-gcse-english-and-believe-it-or-not-pe-at-farnham-heath-end-school-b406952a33d3>

S7 Endorsement from podcaster and Deputy Principal DynamicDeps (2020). Tweet available at <https://twitter.com/DynamicDeps/status/1237838064124219392>