



**Cambridge  
Assessment**

***Moderation of non-exam  
assessments: a novel approach  
using comparative judgement***

***Conference Abstract***

**Lucy Chambers, Sylvia Vitello & Carmen Vidal Rodeiro**

Presented at the 20<sup>th</sup> annual AEA-Europe conference  
Lisbon, Portugal  
November 2019

### **Author contact details:**

Lucy Chambers, Sylvia Vitello & Carmen Vidal Rodeiro  
Assessment Research and Development  
Research Division  
Cambridge Assessment  
The Triangle Building  
Shaftesbury Road  
Cambridge  
CB2 8EA  
UK

chambers.l@cambridgeassessment.org.uk  
vitello.s@cambridgeassessment.org.uk  
vidal.c@cambridgeassessment.org.uk

<http://www.cambridgeassessment.org.uk>

As a department of Cambridge University, Cambridge Assessment is respected and trusted worldwide, managing three world-class examination boards, and maintaining the highest standards in educational assessment and learning. We are a not-for-profit organisation.

### **How to cite this publication:**

Chambers, L., Vitello, S., and Vidal Rodeiro, C.L. (2019, November). *Moderation of non-exam assessments: a novel approach using comparative judgement*. Paper presented at the 20<sup>th</sup> annual AEA-Europe conference, Lisbon, Portugal.

## Abstract

The purpose of moderation is to bring the marking of non-exam assessments to an agreed standard in all centres. Current practice by awarding bodies in England requires moderation to be conducted at centre level, enabling moderators to build up a holistic view of a centre's approach to marking. As each centre is only viewed by a single moderator, this raises challenges with regard to holding the standard across centres.

In recent years, technological advances have allowed electronic submissions of candidates' work (e.g., portfolios). This opens the door for novel ways of moderating that can move towards a scenario in which candidates' work is distributed across multiple moderators without being bound by centre, ensuring that the marking standard is consistently applied across centres. The current study explored the use of Comparative Judgement (CJ) as a method for achieving this.

CJ is an assessment technique whereby a series of paired judgements of students' scripts is used to generate an overall rank order of scripts and a scale of script quality (Pollitt, 2012). As one of the main tasks in moderation is to determine whether the rank order of the scripts is correct, CJ seems excellently placed to accomplish this. Following recent technological advances, CJ can now be conducted online. Since this enables multiple moderators to perform the task remotely and in parallel, it now seems a good time to investigate if CJ could offer a feasible, and potentially more efficient, alternative to the current moderation process.

The CJ procedure was conducted on simulated data sets of candidates' marks. True marks were generated for each candidate and then centre marks were generated from those true marks such that six categories of marking accuracy resulted (accurate, little lenient, strong lenient, little severe, strong severe, and erratic). In essence, this entailed adding varying degrees and types of 'error' to the true marks. The CJ data was analysed using the Extended Bradley-Terry Model.

To assess the success of CJ as a moderation method, two approaches were taken. First, the CJ estimates of script quality were evaluated with regard to their reliability and how well they compared to the simulated/true data. Second, the CJ estimates were evaluated with regard to how well they could be used to assign marks to scripts and to determine the extent to which centres were lenient or severe in their marking. In this study, we used regression to predict the moderator marks based on the CJ estimates. In addition, we explored what minimum parameters of judgements per script and size of moderation sample would be needed.

Our findings show that the CJ method produced very similar estimates of script quality to the simulated/true data, although this was dependent on the number of judgements per script. The regression method of assigning moderator marks worked well maintaining the rank order of scripts within centres and correcting the introduced error. A moderation sample of 10 scripts and 20 judgements per script provided 'usable' moderator marks, with centres effectively calibrated to a common session standard.

Using CJ for moderation could transform current assessment practices by taking advantage of technological developments and new assessment platforms. CJ could also improve the reliability and fairness of moderation. Making paired comparisons should be less cognitively demanding than using a mark scheme and would require less training. As a result, the CJ moderation task could be carried out by any teacher of the qualification (following sufficient training), leading to a shared understanding across teachers and schools of what a good piece of work looks like.

## References

Pollitt, A. (2012). Comparative judgement for assessment. *International Journal of Technology and Design Education*, 22: 157-170.