

# **Online moderation of non-exam assessments: is Comparative Judgement a practical alternative?**

Conference Abstract

Carmen Vidal Rodeiro

Lucy Chambers

AEA-Europe conference, 10-12 November 2022

## Author contact details:

Carmen Vidal Rodeiro & Lucy Chambers  
Assessment Research and Development  
Research Division  
Shaftesbury Road  
Cambridge  
CB2 8EA  
UK

[carmen.vidalrodeiro@cambridge.org](mailto:carmen.vidalrodeiro@cambridge.org)  
[lucy.chambers@cambridge.org](mailto:lucy.chambers@cambridge.org)  
<https://www.cambridge.org/>

As a department of the university, Cambridge University Press & 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.

Cambridge University Press & Assessment is committed to making our documents accessible in accordance with the WCAG 2.1 Standard. We're always looking to improve the accessibility of our documents. If you find any problems or you think we're not meeting accessibility requirements, contact our team: [Research Division](#)  
If you need this document in a different format [contact us](#) telling us your name, email address and requirements and we will respond within 15 working days.

## How to cite this publication:

Vidal Rodeiro, C.L. and Chambers, L. (2022, November 10 - 12). *Online moderation of non-exam assessments: is Comparative Judgement a practical alternative?* [Paper presentation]. AEA-Europe Annual Conference, Dublin.

# Abstract

## Introduction

In England, many high-stakes qualifications include non-exam assessments that are marked by teachers rather than external examiners. Awarding bodies then apply a moderation process to bring the marking of these assessments to an agreed standard. This process involves awarding body trained moderators checking samples of student work to ascertain whether the rank order of the work is correct and if the marking criteria have been applied correctly.

Comparative Judgement (CJ) is a technique where multiple judges compare two (or more) pieces of work, for example pairs of portfolios, and decide which portfolio in each pair is the 'better' one (Pollitt, 2012). Analysis of the resulting data produces an overall rank order of the portfolios. As one of the main tasks in moderation is to determine whether the rank order of the portfolios is correct, CJ seems excellently placed to accomplish this.

Emerging developments in technology, allowing CJ to be implemented on digital platforms and electronic submissions of students' work has meant that moderators can perform the task online. The COVID-19 pandemic has accelerated the use of technology for teaching and assessment processes. At the same time, students' use of technology has also been steadily increasing, both at home and at school. Therefore, using online CJ for moderation could bring change to the way in which schools and students engage with non-exam assessments.

The current study explored the practical feasibility of using CJ for moderation via an experimental moderation task. This included aspects such as whether moderators could view and navigate the portfolios sufficiently to enable them to make the comparative judgements, on what basis they made their decisions, whether moderators could be confident making CJ judgements on large pieces of candidate work (e.g., portfolios), and the time taken to moderate.

## Method and analysis

The study involved making comparative judgements on authentic pairs of e-portfolios using an online platform. Judges were presented with two portfolios at a time and they had to decide which was better based on a holistic judgement of their overall quality.

After the moderation task, judges were invited to complete a short online questionnaire. This gave them the opportunity to provide feedback and enabled the researchers to gather information on their judging behaviour. The judges were also either observed by the researchers (whilst doing some of the judging) or interviewed.

The analysis comprised the evaluation of four types of data: CJ data, observation data, survey responses and interview data. The CJ data (e.g., judge fit statistics) was examined with a view to assessing the consistency of the judgements and the estimates of script quality. Analysis of the survey data, together with more detailed information gathered in the observations and interviews, provided insights about how the judges approached the CJ task and the difficulties they encountered, the usefulness of the online tool, and what features of the portfolios they attended to.

## Findings

This study provided evidence that CJ is a feasible method for moderation and should be explored further. In particular:

- The statistical analysis of the CJ data looked promising. Judge infit values were within an acceptable range, suggesting that the judges were consistent in their judgements. Student rank orders were similar for marking and CJ judgements.
- The judges were able to view and navigate the portfolios easily and found using the online platform to be a positive experience.
- The judges were confident about making comparative judgements of portfolios.
- Despite judges reporting that the process of making holistic judgements was straightforward, some did struggle with the holistic nature of the task, finding it difficult to 'let go' of their current moderation practices.
- The judges reported that they made their decisions based on features such as answer detail, use of examples, correct terminology and relationship to the mark scheme. However, during the observations, the judges made comments about context-irrelevant features (e.g., amount of text, tabulation, quality of scanning) and how these interrupted the flow and caused some frustration.
- The analysis of the CJ data showed that, when compared to traditional moderation, the CJ method was slightly faster.
- A few issues were reported concerning time taken for certain portfolios to load, time lags when scrolling or where a centre had organised the submission in a non-standard way making the evidence harder to find. While these issues are independent of the CJ method and are largely a result of local internet connection and centre submissions, they are features that should be borne in mind if the method is taken forward.

In conclusion, using CJ for moderation could transform current assessment practices and bring about change that reflects the recent and swift adoption of digital approaches in educational settings.

## References

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

## Full paper

Vidal Rodeiro, C.L. and Chambers, L. (2022). Moderation of non-exam assessments: is Comparative Judgement a practical alternative? *Research Matters: A Cambridge University Press & Assessment publication*, 33, 100-119.