

Assessment moderation: Is it fit for purpose?

Associate Professor Lenore Adie

Australian Catholic University

<https://doi.org/10.37517/978-1-74286-685-7-2>

Lenore Adie is Associate Professor of Teacher Education and Assessment, and Senior Research Fellow at the Institute for Learning Sciences and Teacher Education, Australian Catholic University. Her research focuses on assessment and moderation processes, which contribute to quality assurance and improvement purposes. Lenore's research has generated new knowledge in the field of assessment, focusing on quality in assessment practices and processes, especially within systems of standards-referenced assessment. She currently leads an Australian Research Council project investigating the application of standards in promoting learner growth. Lenore has over 30 years of professional experience working in schools as a teacher and in leadership positions, and in teacher education.

Abstract

This presentation outlines the different practices and purposes of assessment moderation and analyses the benefits and issues of these. Two projects are presented; each seeks to improve teacher judgement. Both attempt to redefine what we mean by moderation and how moderation can be conducted in ways to reach diverse and dispersed groups. The first draws on a current Australian Research Council project that is exploring the development of scaled exemplars, the use of commentaries of judgement decisions, and the use of digital platforms to support teachers grading student work. The second draws on the work of the Graduate Teacher Performance Assessment (GTPA) group of 19 universities to demonstrate how large-scale moderation of complex performance assessments can be conducted across states and territories. The layers of processes, including the use of technology, required to ensure judgement reliability are illustrated. In both contexts, the ways in which moderation is understood and practised are being reimaged.

Introduction

Assessment has long been a contentious practice. What to test, when and whose role it is, are embedded in levels of trust and beliefs in teacher professionalism. Moderation is one of the quality assurance processes we look to for reliability and consistency in judgement decisions. But this stance assumes that aspects such as consistency in judgement are desirable. Some teachers would view these practices as a form of top-down accountability that constrain their professionalism by imposing a structure that hinders their decision-making about how to best progress a student's learning. The position taken in this paper, however, is that assessment and moderation can be designed and used in ways that enhance the agency of teachers to improve their teaching and the learning of their students. Of importance here is how these practices contribute to, rather than distract from, good teaching that progresses student learning.

What is understood as assessment moderation?

Moderation, as a process, is broadly understood as a form of quality assurance and an 'approach to agreeing, assuring and checking standards' (Bloxham et al., 2016, p. 638). Moderation of judgement decisions can be undertaken in various ways, dependent on purpose. For example, consensus or social moderation is used to reach agreement on grades awarded. Expert moderation defers to the evaluative experience of someone such as a curriculum leader to make a final judgement decision, particularly in cases of disagreement about the awarded grade. Statistical moderation, often used in high-stakes examinations, has a focus on reliability and involves the scaling of scored performances (Crisp, 2017).

Social moderation meetings involve teachers in discussion and negotiation of the meaning of criteria and standards to reach agreement of an awarded grade. Through these discussions, teachers can develop shared understandings of the qualities of a standard and the different ways these can combine while still representing the same standard. This process of discussion and negotiation is essential since standards are text-based, so various interpretations of terms can exist. Individual histories impact on the reading of any text, and research has shown that teachers have different understandings of standards (Wyatt-Smith & Gunn, 2009). Through social moderation, teachers can reach agreement on the qualities that represent a standard. The issue is that these shared understandings of an expected standard of performance can remain idiosyncratic to a school, or to a district, and in fact, vary widely across schools and districts.

Statistical moderation is claimed to mitigate the subjectivity of human judgement by calibrating scores against other assessments or judgement patterns. In this way, comparability across dispersed populations can be achieved. However, a recent example illustrates the challenges that are made to the claim of objectivity. During the COVID-19 pandemic, the cancellation of end-of-course examinations in the United Kingdom turned the gaze to teacher knowledge of student performance. A decision was made to use teachers' estimation of grades based on student classroom performance with an algorithm to adjust these scores based on the historical examination performance of the school (Kippin & Cairney, 2022; Porter, 2020). In a turn of events, the algorithm was charged with introducing bias as it caused the socio-economic inequalities among schools to be perpetuated through its use of historical performance data. This example illustrates how statistical models may not, of themselves, remove bias, and that data should always be interrogated for what it can tell us.

How can moderation processes be strengthened to meet different purposes and provide reliable outcomes?

In our research, we have been investigating how statistical and social moderation can be utilised to improve consistency and demonstrate reliability. To reach across dispersed populations, we have used and developed customised online infrastructures. To address diversity within these populations, several key practices and artefacts have been identified. We present two current projects to illustrate how moderation can be implemented in ways that connect teacher judgement with improved teaching and learning. Both have involved us in rethinking our understanding of moderation.

Project 1: Improving teacher assessment capabilities using exemplars and online moderation¹

This project, funded by the Australian Research Council, is investigating the extent to which the development of scaled exemplars of A–E standards with commentaries of how the judgement was made supports comparability of judgement. The research focuses on the middle years of schooling, specifically Years 4, 6 and 8, in the areas of English narrative, maths and science investigations and religious education.

The project involves teachers working online in:

- a process of pairwise comparison of submitted samples of student assessments.
- a standard-setting process to identify samples that best represent a standard descriptor.
- writing cognitive commentaries for selected exemplars. The commentaries describe the strengths and weaknesses of a work sample and how these combine to give an overall judgement, as well as identifying next steps for teaching (Wyatt-Smith & Adie, 2021). Teachers have told us that writing the commentaries focused their attention on specific features of assessments and the application to next steps for teaching. This has given them the knowledge to progress learning from one assessment to the next unit of work. Commentaries also provide a means for this thinking to be used by others.
- meeting online with other year-level and discipline teachers to discuss and refine the cognitive commentaries.
- using the exemplars with the accompanying cognitive commentary to moderate their own judgement decisions. A selected set of A–E student work samples will be blind reviewed by discipline experts to further evaluate the usefulness of the resources to support consistency in teacher judgement.

Across the project, various quantitative and qualitative methodologies have been used to identify consistency of teacher judgements and identify exemplar samples. Forms of statistical and social moderation have occurred across the different activities. The proposed outcome is a form of online moderation that takes forward teachers' explanations of their judgement decisions through the cognitive commentary to enable teachers across diverse locations to moderate their own judgement decisions.

Project 2: The Graduate Teacher Performance Assessment (GTPA®)²

Since 2019, Australian initial teacher education (ITE) providers have been required to implement a teaching performance assessment to final-year preservice teachers, with successful completion required for graduation and licensure. The GTPA is one such validated assessment, led by the Australian Catholic University, and used by 19 universities across Australia. Supporting consistency of judgement across a large and diverse group of universities involved the design of customised digital infrastructure and app design (Wyatt-Smith et al., 2022).

1 ARC LP180100046. We acknowledge the funding received for this project from the Australian Research Council, Australian Catholic University, University of Western Australia and Partner Organisations: the Queensland Department of Education, Western Australian School Curriculum and Standards Authority, and Catholic Education Western Australia Limited. The research team includes Associate Professor Lenore Adie, Professor Claire Wyatt-Smith, and Professor Michele Haynes (Australian Catholic University), Associate Professor Stephen Humphry and Professor David Andrich (University of Western Australia), Professor Chris DeLuca (Queen's University, Canada), and Mr Terry Gallagher (Queensland Department of Education).

2 The Graduate Teacher Performance Assessment (GTPA®) was created by the Institute for Learning Sciences and Teacher Education (ILSTE), Australian Catholic University and has been implemented in a Collective of Higher Education Institutions in Australia (graduatetpa.com). ILSTE has led the validation of the instrument, standard-setting and cross-institutional moderation with the engagement of teacher educators, policy personnel and a multidisciplinary research team.

At the centre of these processes is online cross-institutional moderation (CIM) that is conducted annually (CIM-Online™)³. CIM involves teacher educators from each university within the GTPA Collective individually scoring online de-identified samples from each university within the group (Wyatt-Smith & Adie, 2021). To ensure reliability, supporting processes and artefacts are provided, including:

- performance-level descriptors that support teacher educators to gain a sense of overall level, and criteria specifications that focus the gaze on required features
- calibration processes where judges grade previously validated exemplars prior to undertaking CIM to gain a sense of the expected standard at three levels of performance (meets, above and below the standard).
- exemplars in the form of validated preservice teacher samples from a previous moderation process with accompanying cognitive commentaries of judgement decisions that illustrate application of the standard used in calibration and for review when making judgements.
- anchor samples used to track any change in the applied standard over time.

The data from CIM are analysed statistically to investigate the comparability of judgements in terms of applying the established standard and endorsing submitted samples, as well as the performance characteristics of preservice teachers across program characteristics. These analyses are provided back to universities through a confidential report to use in ongoing programming decisions and for consideration in their teaching (formative purpose). The data are also used for accreditation purposes by the universities when reporting to state regulatory authorities (summative purpose).

For the first time, cumulative data from CIM-Online have provided information about the quality of teacher education graduates. Our data collected from 2017 have shown that the endorsement of samples increases over time though this is dependent on factors such as staff changes and university funding models. The collated data have contributed to rigorous professional conversations among teacher educators of the expected qualities for graduating preservice teachers, as well as investigation of context-specific responses within each university.

Conclusion

So, what have we learnt? By combining statistical and social processes of moderation with digital technologies and customised supporting resources, the consistency and reliability of judgement decisions can be improved over time. Significantly, through these processes, teachers have moved beyond moderation as an end process following summative assessment to one that views moderation as an ongoing process throughout teaching and learning, informed by the use of data for formative, improvement purposes. The main purpose of moderation may just be the opportunities it affords to interrogate and improve teaching and learning.

³ The online model of cross-institutional moderation (CIM-Online™) has been supported by digital architects in the Institute for Learning Sciences and Teacher Education, Australian Catholic University.

References

- Bloxham, S., Hughes, C., & Adie, L. (2016). What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices. *Assessment and Evaluation in Higher Education*, 41(4), 638–653. <https://doi.org/10.1080/02602938.2015.1039932>
- Crisp, V. (2017). The judgement processes involved in the moderation of teacher-assessed projects. *Oxford Review of Education*, 43(1), 19–37. <https://doi.org/10.1080/03054985.2016.1232245>
- Kippin, S., & Cairney, P. (2022). The COVID-19 exams fiasco across the UK: Four nations and two windows of opportunity. *British Politics*, 17(1), 1–23. <https://doi.org/10.1057/s41293-021-00162-y>
- Porter, J. (2020, 17 August). UK ditches exam results generated by biased algorithm after student protests. *The Verge*. <https://www.theverge.com/2020/8/17/21372045/uk-a-level-results-algorithm-biased-coronavirus-covid-19-pandemic-university-applications>
- Wyatt-Smith, C., & Adie, L. (2021). Introducing a new model for online cross-institutional moderation. In C. Wyatt-Smith, L. Adie, & J. Nuttal. (Eds). *Teaching performance assessments as a cultural disruptor in initial teacher education: Standards, evidence and collaboration* (pp. 39–58). Springer.
- Wyatt-Smith, C., Adie, L., Haynes, M., & Day, C. (2022). *Professionalizing teacher education: Performance assessment, standards, moderation, and evidence*. Routledge.
- Wyatt-Smith, C., & Gunn, S. (2009). Towards theorising assessment as critical inquiry. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory and practice* (pp. 83–102). Springer.