

Reconstructing Publicness in Educational Assessment through Reflective Consensus Moderation

Insights from Group Moderation Practice in Japanese Senior High Schools

Takahiro ENDO

University of Fukui, Japan

This study examines how evaluation criteria can be justified among actors with differing educational views. Drawing on Japanese group moderation involving teachers and students, it proposes “reflective consensus moderation” as a framework for the deliberative development of criteria, reconceptualizing assessment as a democratic practice of public justification rather than a technical exercise in grading.

Keywords: public reason, reflective equilibrium, overlapping consensus, normative validity, plurality

This study examines how evaluation criteria can be justified among actors with differing educational views. It explores how publicness in educational assessment can be reconstructed through “reflective consensus moderation,” drawing on group moderation practices in Japanese senior high schools.

In recent years, group moderation has been increasingly implemented in Japanese high schools to enhance the reliability of performance assessments. Typically, it involves multiple teachers evaluating the same student work and comparing their judgments. More recently, its scope has expanded to include students as evaluators and to extend beyond single schools to inter-school contexts. A notable example is a student-led inter-school group moderation initiative organized by students at Fukui Prefectural Takefu Senior High School (Endo, 2024).

While group moderation was originally introduced to ensure inter-rater reliability, emerging practices suggest a shift in its function. Rather than simply aligning scores, it increasingly serves as a space for dialogue among evaluators. In this process, not only assessment results but also the underlying criteria, educational aims, and conceptions of inquiry are critically examined.

Such practices reveal differences in evaluators’ views of scientific inquiry that are often implicit in curriculum design and assessment. Group moderation makes these differences visible as variations in how student performance is interpreted, thereby bringing them into collective discussion. In this sense, it plays an important role in curriculum management in an increasingly diverse educational context.

However, educational assessment cannot be reduced to the coexistence of multiple perspectives; it also requires the capacity to distinguish more and less justified judgments. This plurality is

not merely empirical but constitutive of human coexistence (Arendt, 1958/2018). This raises a key question: How is it possible for there to exist evaluation criteria that can be publicly justified among educational actors who remain profoundly divided by reasonable educational views?

To address this question, this study draws on the idea of “public reason” (Rawls, 1993/2015). While Arendt (1958/2018) frames plurality as an irreducible condition of human coexistence, Rawls (1993/2015) offers a framework for constructing publicly justifiable criteria under conditions of reasonable disagreement among educational actors. In Rawls’s political philosophy, this condition corresponds to what he calls the fact of reasonable pluralism, in which free and equal citizens inevitably hold diverse and often conflicting yet reasonable comprehensive doctrines. In this study, this idea is transposed to educational assessment, where evaluation criteria are reconceptualized as public reasons that must be justifiable across different educational views (Endo, 2024).

This study reconceptualizes group moderation in educational assessment as a process of generating and examining publicly justifiable criteria under conditions of persistent differences in educational values, thereby extending Rawls’s idea of public reason into assessment practice.

From this perspective, group moderation is not merely a technique for achieving consistency in grading, but a practice through which evaluation criteria are constructed, negotiated, and critically examined as public reasons. This study defines this process as “reflective consensus moderation.”

While prior research on consensus moderation has focused primarily on achieving agreement in grading (Sadler, 2013), reflective consensus moderation shifts attention to the “normative validity” (Biesta, 2020) of evaluation criteria and the reflective dialogue through which these criteria are continuously reconsidered.

From this perspective, a key issue is how the legitimacy of negotiated criteria can be ensured. Even when evaluators reach consensus, the resulting criteria may still lack educational validity or fail to align with educational aims. Therefore, moderation must go beyond intuitive comparison of judgments and instead make explicit the criteria underlying those judgments, enabling their examination across cases.

This process can be interpreted in relation to the method of “reflective equilibrium” (Rawls, 1971/1999), in which principles and judgments are mutually adjusted to achieve coherence (Endo, 2022). Ultimately, ensuring the legitimacy of criteria requires an “overlapping consensus” (Rawls, 1993/2015) grounded in publicly justifiable reasons that can be shared across diverse and potentially conflicting educational values.

At Takefu Senior High School, teacher-led group moderation is conducted in parallel with student-led group moderation, forming a cycle in which criteria emerging from both processes are compared to revise existing rubrics. This case illustrates how reflective consensus moderation can function as a practical site for aligning evaluative judgments with publicly justifiable criteria through processes akin to reflective equilibrium and overlapping consensus.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP21KK0040, JP22K02283, JP24K21473, and JP25K00862.

This manuscript was prepared with the assistance of AI language models (including ChatGPT and Google Gemini) to improve linguistic readability and refine the logical structure.

References

- Arendt, H. (2018). *The human condition* (2nd ed.). University of Chicago Press. (Original work published 1958)
- Biesta, G. (2020). *Educational research: An unorthodox introduction*. Bloomsbury Academic.
- Endo, T. (2022). Moderation as reflective equilibrium: Methodological perspectives on learning assessment in the period for inquiry-based cross-disciplinary study. *Studies in and on Teacher Education*, 15, 419-427. United Graduate School of Professional Development of Teachers, University of Fukui. [In Japanese] <http://hdl.handle.net/10098/00029381>
- Endo, T. (2024). Criteria as public reasons: Focusing on the development of group moderation in high schools in Fukui Prefecture. *Studies in and on Teacher Education*, 17, 381-392. United Graduate School of Professional Development of Teachers, University of Fukui. [In Japanese] <http://hdl.handle.net/10098/0002000518>
- Rawls, J. (1999). *A theory of justice* (Rev. ed.). Harvard University Press. (Original work published 1971)
- Rawls, J. (2015). *Political liberalism* (Expanded ed.). Columbia University Press. (Original work published 1993)
- Sadler, D. R. (2013). Assuring academic achievement standards: From moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, 20(1), 5–19. <https://doi.org/10.1080/0969594X.2012.714742>