

A Collaborative Rubric Development and Scoring Program

Bernadette Mullins Justin Hoffmeier Jared Bunn
Florida Polytechnic University Florida Polytechnic University Florida Polytechnic University

Satyajith Boyana Dipali Swain
Florida Polytechnic University Florida Polytechnic University

Keywords: coordinated courses, rubric scoring, inter-rater reliability

Research Aims and Relationship to the Research Literature

The literature speaks to the benefits of course coordination (Rasmussen et al., 2014) and the importance of consistency in grading across multi-section courses. Achieving grading consistency can be challenging (Reddy & Andrade, 2010) but can be enhanced by using task-specific rubrics and rater training (Jonsson & Svingby, 2007; Crisp, 2017; Schoepp, et al., 2018). It is well-documented in the literature that rubric training can be a time-consuming process and that faculty time constraints can be a barrier to implementation of assessment practices in general and rubric scoring in particular (Charlton & Newsham-West, 2024; Holmes & Oakleaf, 2013; Williams, 2024). However, the research literature does not seem to adequately address how, after an initial time investment, rubric use can become routine. We seek to investigate the following research question: after instructors invest time in rubric development and training, how does the time needed for rubric implementation change?

Preliminary Data Collection and Results

Our investigation of this research question has two phases. Phase one is the initial time investment in rubric development and training. Phase two is rubric utilization. For the first phase, we collected student responses ($n = 147$) on assessment items involving exponential growth, cooling, and trigonometry. Faculty members (two associate professors and two assistant professors in their first year of teaching post Ph.D.) teaching a coordinated precalculus course participated in a collaborative rubric development and training program: after selecting a task aligned with learning outcomes (e.g., problem-solving, communication, modeling exponential growth), we collaboratively developed a draft task-specific rubric specifying partial credit based on the demonstrated level of understanding, then refined it by anticipating potential student solutions, errors, and omissions. We scored a small sample of blinded student responses as a whole group to discuss how to apply the rubric and make necessary revisions. Next, we independently scored a sample of student work and discussed results to identify discrepancies and come to a consensus rating. After the training session, we independently scored the remaining student solutions, then reviewed discrepancies to identify lessons learned. In round one, the percent agreement across all raters was 50.3% and the percent adjacent agreement was 73.1%. In round four, the percent agreement was 84.7% and the percent adjacent agreement was 91.5%. Implementation of the second phase is forthcoming. Using already collected student work on different assessment items, each faculty participant will write a task-specific scoring rubric without additional training beyond the initial time investment in phase one. For each rubric, all faculty participants will independently score the student work and percent agreement will be computed. Faculty perceptions of changes in time required for rubric development and implementation will be captured through a survey and a structured interview process with each faculty participant.

References

- Charlton, N., & Newsham-West, R. (2024). Enablers and barriers to program-level assessment planning. *Higher Education Research & Development*, 43(5), 1074–1088.
- Crisp, E. A. (2017). Calibration: Are you seeing what I'm seeing? *Intersection*, Winter 1(3), 7-13.
- Holmes, C., & Oakleaf, M. (2013). The official (and unofficial) rules for norming rubrics successfully. *The Journal of Academic Librarianship*, 39, 599-602.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.
- Rasmussen, C., Ellis, J., Zazkis, D., & Bressoud, D. (2014). Features of successful calculus programs at five doctoral degree granting institutions. In S. Oesterle, P. Liljedahl, C. Nicol, & D. Allan (Eds.), *Proceedings of the 38th Conference of the International Group for Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education* (Vol. 5, pp. 33–40). Vancouver, Canada: PME.
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448.
- Schoepp, Kevin, Danaher, Maurice, & Ater Kranov, Ashley. (2018). An Effective Rubric Norming Process. *Practical Assessment, Research & Evaluation*, 23(11).
- Williams, Melodie A. R., "Barriers To Faculty Adoption Of Assessment: A Case Study In Engineering" (2024). Theses and Dissertations. 6581. <https://commons.und.edu/theses/6581>