

## Assessment, Evaluation, and Grading: A Reflection on My Radicalization

Priscilla Bremser  
*Middlebury College*

Follow this and additional works at: <https://scholarship.claremont.edu/jhm>



Part of the [Arts and Humanities Commons](#), and the [Mathematics Commons](#)

---

### Recommended Citation

Bremser, P. "Assessment, Evaluation, and Grading: A Reflection on My Radicalization," *Journal of Humanistic Mathematics*, Volume 11 Issue 2 (July 2021), pages 291-297. DOI: 10.5642/jhummath.202102.15 . Available at: <https://scholarship.claremont.edu/jhm/vol11/iss2/15>

©2021 by the authors. This work is licensed under a Creative Commons License.

JHM is an open access bi-annual journal sponsored by the Claremont Center for the Mathematical Sciences and published by the Claremont Colleges Library | ISSN 2159-8118 | <http://scholarship.claremont.edu/jhm/>

The editorial staff of JHM works hard to make sure the scholarship disseminated in JHM is accurate and upholds professional ethical guidelines. However the views and opinions expressed in each published manuscript belong exclusively to the individual contributor(s). The publisher and the editors do not endorse or accept responsibility for them. See <https://scholarship.claremont.edu/jhm/policies.html> for more information.

# Assessment, Evaluation, and Grading: A Reflection on My Radicalization

Priscilla Bremser

*Department of Mathematics, Middlebury College, Vermont, USA*  
bremser@middlebury.edu

---

## Synopsis

The COVID-19 pandemic forced instructors to adapt their assessment practices. For this author, that adaptation led to a reconsideration of evaluation systems, given the ways in which such systems can interfere with learning and perpetuate inequities. The author describes resulting changes in her grading of student work and evaluation of faculty colleagues.

**Keywords:** assessment, grading, ungrading, student evaluations of teaching.

---

The last time I stood in a classroom, a student asked, “Do you think we’ll be able to come back to campus this semester?” It was March of 2020, and at the time all I could say was “I have no idea.” That familiar space near the blackboard suddenly became murky, the floor less solid. All we knew was that students had just a couple of days to get home to Massachusetts or California or Morocco or Bhutan, and faculty had just a couple of weeks to move our courses and advising online. I welcomed the extra week of spring break to re-engineer my teaching, but it meant I’d have to remove some course material from the plan. How would I decide what other changes to make?

Years before, a literary friend had introduced me to John Keats’ idea of “negative capability,” or as she put it, “the ability to be content floundering in the muck.” We may not have been content, but we were certainly floundering. In those subsequent days of mastering Zoom breakout rooms and downloading patterns for fabric face masks, my students were studying in vastly dif-

ferent spaces with varying internet access. I could not imagine implementing closed-book tests under those conditions, so I replaced the two remaining exams in Calculus I with portfolio assignments.

My syllabus already included a portfolio, meaning a set of seven or eight substantial problems, from which each student would choose five and write complete solutions. For the first pandemic portfolio, I assigned this instead:

Choose any three problems that improved your understanding of the course material from the first five weeks of the course. They can come from class work, homework, quizzes, or the first exam. For each problem,

- Write out the problem as it was assigned.
- Write out a full solution to the problem. Use full sentences, and justify any steps that need clarification, in your judgement.
- End with a “Discussion” section, explaining how this problem helped your understanding.

The remaining portfolio assignments included a similar section, along with some new problems. Emerging from the muck was a conviction that prioritizing student learning means designing assessments that focus on what I want students to be able to do well and asking them to engage in some metacognition.

Meanwhile, college faculty meetings featured an extended debate about adopting a temporary grading system. The administration proposed that the faculty give students a choice between a grade on the A-F scale or a Pass/Fail for each course. A colleague and I offered a different proposal: mandatory Credit/No Credit for the spring of 2020. Our rationale included our beliefs that students now had unequal access to A's, that to include any form of standard grading falsely suggested that the semester was like any other, and that students (and faculty) were under unusual levels of stress and needed some relief. In the end, we were outvoted. At least the final version gave students a choice between an A-F grade or a Credit/No Credit, in which No Credit meant that nothing about that course would appear on the transcript.

The debate itself, however, put weaknesses in our grading systems into high relief. It's been clear for a long time that, on my campus and beyond, we on the giving end don't agree on what grades mean. Are we using a norm-referenced system, a criterion-referenced one, or some hybrid? What does a "B" signify? Does extreme grade inflation render grades useless? More to the point, what are we claiming to measure? What became clearer to me during the debate was how problematic the enterprise is on the receiving end. Who is our audience when we assign grades: the students, our colleagues, potential employers, or graduate program admissions officers? And how do they interpret what they see? Do we have any idea, really, what effect our reliance on grading has on our students, in the short term and later on?

What cracked that last question open for me, in part, were some emails I received from students between faculty meetings. I had expected resistance to a binary grading system from pre-medical students and their advisors, but I wasn't prepared for the vehemence. One of my Calculus students told me it would be a "punishment" to remove the possibility of a letter grade, because students applying to medical school would have to take the course again for "real" grades. Another student described a difficult home situation and then wrote, "I would still like to show that despite it all, I was able to try my hardest and obtain the grade I deserved for this semester." Said a third, "I need to show medical schools that I can handle working in times of stress."

The extent to which these students see grades as an assessment of their character, rather than their performance on some specific tasks at singular moments in time, is distressing. Calculus isn't just a necessary hurdle on the track to a medical career; it's a proxy for the ability to handle the pressures of the emergency room, in their minds.

The internalization of one's GPA as a measure of personal worth wasn't limited to our students, judging from some of the faculty comments in opposition to our proposal. Those colleagues were apparently not moved by others who spoke of the myth of meritocracy [6], though those comments helped me to take a broader perspective.

After I submitted my letter grades, I signed up for a "summer camp" offered by our digital learning crew for those of us who had decided to teach entirely online in the fall. The course could have been called "Instructional Design for

Online Instruction During a Pandemic,” and while I had read the chapter on Design Practices in the *MAA Instructional Practices Guide* [1], I appreciated the chance to write up my learning goals and teaching philosophy anew for an audience of mostly non-mathematicians. Our instructors recommended more frequent and more varied assessments, not just as a way to remove the stress of a final exam worth 40% of the final grade, but because well-designed quizzes, for example, support learning (see, for example, [3]).

A clearer picture was coming out of the muck now. Added to my previous understanding of summative assessment and formative assessment, I was thinking more about the learning that can happen *during* assessment. I had seen this in real time during an observation of Deborah Ball interviewing a student during a “Critical Issues in Mathematics Education” conference at the Mathematical Sciences Research Institute in 2013. The interview was meant to give a detailed picture of the student’s understanding, but I could see the student gaining some understanding as the conversation proceeded. This is an illustration of the strength of active learning environments. We learn mathematics by being in conversation with others. I wanted students to write up clear solutions as well as reflect on their learning not because I wanted to grade those assignments (though they are much more interesting than a standard midterm exam), but because I know that those activities support long-term retention of the concepts in the course.

I taught two sections of Linear Algebra and one section of Abstract Algebra online in the fall of 2020. I got better at dropping into breakout rooms, but I still missed listening in on small groups in an actual classroom. It didn’t feel right to give out grades for participation, given what I knew about individual challenges for some of my students, so I chose generosity for that element. But really, what did I know about similar challenges in pre-pandemic times? Only what students, or their deans, shared with me; just a glimpse.

During that term, I was asked to submit a letter on behalf of a colleague under review. While Course Response Forms — our version of student evaluations — are generally only visible to the individual, the department chair, and the review committee, this individual had shared theirs with the senior department members. Having primed myself to reconsider my participation in assessment systems, I realized that I could not in good conscience rely

on this problematic instrument. (See, for example, [4, 5] for biases found in student evaluations of instructor performance.) I now include a version this statement in all such letters:

I have not looked at X's course response forms, as I have come to believe, based on abundant evidence, that student evaluations should not be used in hiring or reappointment/promotion decisions. (This is both because they reflect and may even amplify inherent biases in society, and because students are poor judges of their own learning. I'll spare you the citations here.) I am glad that X has found them useful in their own development as an instructor, which is all they are good for.

At the end of the fall semester, I joined a faculty reading group to discuss *UNgrading* [2], a collection of essays by a variety of instructors. In the book, I found support for my emerging sense that the impulse to measure and rank has negative impacts on learning. It's not just that our students have a warped view of what their grades mean; on top of that, the focus on grades interferes with their intellectual growth. From colleagues in the reading group who have already done some ungrading, including a few in STEM fields, I heard strategies for shifting the emphasis away from sorting students and toward learning. One is to respond to student work with comments, but no grades, and then have students grade themselves at the end of the term. This isn't simple; these instructors reported that students' grades for themselves were close to what the instructors would have given, but sometimes women seemed to be harder on themselves than men. Still, I am going to try a version of this method in the fall.

Am I a radical now? Not according to Alfie Kohn, who wrote the Introduction to *UNgrading*. In Kohn's framework, while I'm familiar with "research showing that grading has three predictable effects — less interest in learning, a preference for easier tasks, and shallower thinking" [2, page *xiv*] — I'm only partway along the path to doing all I can in my own classroom. That, in turn, is only the beginning, according to Kohn, toward his preferred goal of eliminating grading entirely at a systems level. As for my small act of defiance against high-stakes uses of student evaluations of teaching, there, too, I am acting mostly at an individual level. It's a start.

What is it about living and working through a pandemic — floundering through this particular muck — that has moved me in these directions? I suspect it has to do with being compelled by new limitations to consider what’s essential. Also, when so much we’ve taken for granted is no longer available, we may be more inclined to question *all* of our assumptions. On a deeper level, being a witness, however distant, to so much loss compels me to consider what’s compassionate. I am now more likely to resist established systems that, in my mind, interfere with people’s intellectual and professional growth. I want to support my students and my colleagues without relying on deeply flawed evaluation systems just because we’ve been in the habit of using them. I would have called that radical a few years ago, but now it seems like the only viable path.

## References

- [1] Martha L. Abell, Linda Braddy, Doug Ensley, Lewis Ludwig, and Hortensia Soto, *MAA Instructional Practices Guide: Guide to Evidence-Based Instructional Practices in Undergraduate Mathematics*, Mathematical Association of America, 2018. Available online at <https://www.maa.org/programs-and-communities/curriculumresources/instructional-practices-guide>, last accessed on July 23, 2021.
- [2] Susan D. Blum, editor, *UNgrading: Why Rating Students Undermines Learning (and What to Do Instead)*, West Virginia Press, Morgantown WV, 2020.
- [3] Peter C. Brown, Henry L. Roediger III, and Mark A. McDaniel, *Make it Stick: The Science of Successful Learning*, The Belknap Press of Harvard University Press, Cambridge MA and London, 2014.
- [4] Louis Deslauriers, Logan S. McCarty, Kelly Miller, Kristina Callaghan, and Greg Kestin, “Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom,” *Proceedings of the National Academy of Sciences of the United States of America*, Volume 116 Number 39 (2019), pages 19251–19257.

- [5] Michelle Falkoff, “Why We Must Stop Relying on Student Ratings of Teaching,” *The Chronicle of Higher Education*, April 25, 2018; available at <https://www.chronicle.com/article/why-we-must-stop-relying-on-student-ratings-of-teaching/>, accessed on June 15, 2021.
- [6] Michael J. Sandel, “Toppling the myth of meritocracy,” *The Harvard Gazette*, January 25, 2021; available at <https://news.harvard.edu/gazette/story/2021/01/the-myth-of-meritocracy-according-to-michael-sandel/>, accessed on June 15, 2021.

---

**Author bio:** Priscilla Bremser is a professor of Mathematics at Middlebury College, a small liberal arts college in Vermont. Trained in number theory, she now focuses more on mathematics education. In addition to teaching undergraduates, Bremser has worked for over a decade as an instructor in a master’s degree program for practicing K-12 math teachers. That has been a rich and rewarding experience, and informed her understanding of mathematics education on many levels.

---