Can Technology Make Grading Fairer and More Efficient?

Strategies for escaping grading hell.

Steven Mintz

December 9, 2019

Many of us would agree with the following statement: we'd teach for free, but must be paid to grade.

Grading can be hell, as we agonize over the difference between an A or A-minus or a B or B-plus or respond to student complaints that our grades are unfair and inconsistent.

So what can be done to escape grading hell?

Machine-graded tests provide one solution. Grade inflation offers another.

One instructor even experimented with outsourcing grading to Bangalore -- a strategy that, of course, raises FERPA concerns.

But if you genuinely want to provide an accurate assessment of the quality of students' work and detailed, substantive feedback, grading becomes the most grueling, time-consuming aspect of teaching.

And the pressure is likely to worsen as caps on courses increase. The increases may be modest, with seminars adding a student or two, or they may be larger. But in either case, instructors will be expected to respond to more and more student work.

So let's look at five possible responses.

1. Improving workflow

Canvas's SpeedGrader and Gradescope are online grading and analytics tools designed to accelerate the process of assigning grades and providing

student feedback and discovering what concepts, content or methods large numbers of students do not understand or have not mastered.

These tools streamline the grading process by allowing instructors to grade question by question rather than by student, automatically post grades to the LMS, easily draw upon a common set of comments and view the distribution of grades for each question. Both tools allow instructors to hide student names to support anonymous grading, use rubrics to make grading more precise and fair, alter the value of each rubric item globally, and deliver text and even video or audio comments to students.

SpeedGrader and Gradescope make grading more efficient -- but still require the grader to look at each individual student response. But there are other approaches that do reduce the burden on graders, although each raises difficult questions involving validity, reliability, and consistency.

2. Automated peer assessment

All the major LMSes can automate peer assessment. These systems can distribute questions to students to view and gather their responses.

Peer assessment can be a powerful learning tool. When students evaluate a classmate's work and provide feedback, they not only better understand the criteria that instructors use in assessment, but they learn how to give constructive feedback. But whether we can rely on peers to accurately and fairly assess student work is a source of great concern.

A particularly powerful tool for automated peer assessment is the University of Michigan's M-Write, which that institution uses to incorporate writing into large STEM classes.

Faculty at Michigan identify key concepts that students need to master, develop prompts that require students to write about the concepts, use automated text analysis to identify students who need help, deliver automated feedback based on the types of errors and automatically distribute student responses, along with detailed rubrics, for peer assessment.

3. Autograding of formulas

The rise of MOOCs helped spur the development of tools designed to grade and provide feedback to students at the scale of tens of thousands. While this usually involved multiple-choice or fill-in-the-blank questions, MOOCs in such fields as computer science and engineering use mathematical language processing to machine grade formulas. One tool used widely by computer science departments to automatically grade formulas is Web-CAT. Autograding of formulas can help instructors easily identify concepts that substantial numbers of students don't understand or methods that students do not know how to apply correctly.

4. Autograding of essays

The holy grail, for many instructors, would be a tool that autograded essays. Although such tools exist, and are sometimes used by the big testing firms, such as Pearson or ETS, to grade standardized tests, such tools are not ready for widespread adoption. As *Forbes* magazine put it, "Automated Essay Scoring Remains an Empty Dream." To ensure fairness and to make sure that students don't game the system, autograding of essays is generally combined with the use of human graders.

Essay autograders use natural language processing and semantic and syntactic analysis to assess grammar, vocabulary (including key words) and word choice, sentence structure, the number of subordinate clauses, reading level, and writing mechanics. Usefulness is largely limited to fluency, diction, grammar and structure. But these tools can tell if a student is writing off-topic and help students improve their writing skills.

I should add: autograding of content can work if the responses are not open-ended and require students to respond in predictable ways with certain formulaic language. Automatically grading a student's argument, organization and use of evidence in open-ended essays, alas, remains beyond the capacity of existing tools.

5. Personalized, adaptive learning

If our ultimate goal is not simply to rank students and evaluate performance, but to bring all students to competency, another approach -- involving personalized, adaptive learning -- makes sense. Personalized, adaptive courseware gives students frequent diagnostic questions and uses data mining to track student mastery of key concepts, content and problem-solving skills. Such courseware can flag problems and automatically refer students to resources -- to e-tutors -- that can provide remedial help.

Personalized, adaptive courseware has sometimes been dismissed as "digital snake oil," but if Knewton, which held out the prospect of a "mind-reading robo tutor in the sky," failed to live up to its promise, other examples are much more promising. These include the University of Texas at Austin's Dana Center's Mathematics Pathways, Carnegie Mellon's Open Learning Initiative's courseware, and OpenStax tutor.

As someone who has created personalized, adaptive courseware for use in my very large introductory history classes at the University of Texas at Austin and who makes extensive use of a home-grown classroom response system (UT Instapoll) that uses students' cellphones, and which can automatically display distribution of student responses to prompts and questions, I can personally attest to the ways that such approaches and tools can improve student performance while significantly easing the burden of grading.

For more information about computer-assisted grading, see: Computer Aided Personalized Education and Automated Scalable Assessment: Present and Future.

Steven Mintz is senior adviser to the president of Hunter College for Student Success and Strategic Initiatives.

Steven Mintz