AI in Assessment

Case study 15: Fast and Fair Marking with Gradescope

Keywords
AI-assisted marking, Gradescope, handwritten assessments, digital marking, reliability, efficiency, feedback, fairness, MCQ, SAQ

Scenario Summary
Marking a stack of handwritten mathematics scripts is slow, inconsistent, and easy to get subtly wrong – a mis-added total here, a slightly different standard applied there. In this case study, a course convenor turned to Gradescope, a platform that can mark both multiple-choice and handwritten short-answer questions, to mark coursework and an exam. The AI-assisted grading features didn't just speed things up, they made marking more reliable, gave the convenor new ways to reach struggling students, and - perhaps most surprisingly – changed how students engaged with their feedback. What Gradescope offered wasn't automation replacing judgement, but an optimal balance between software capable of spotting patterns and human oversight through checking.

Context
This first year mathematics course for engineering students sits within an extended degree programme, designed for students repeating the course with double the contact time, smaller classes, and more frequent feedback. One convenor typically runs the whole class – lecturing and tutoring – with assessment split evenly between exams and coursework, the latter built from tutorials, workshop style activities, and tests done individually and in groups.

The trigger for adopting Gradescope was about both marking efficiency and feedback going to waste. Students rarely collected their marked paper scripts to learn from their feedback, they also tended to selectively review memos. The convenor wanted feedback to do its actual job and support students with their learning: "the point of feedback is to improve. It's not to sort of stamp you as, "You are mediocre," or "You are a failure." It's to say, "You haven't got this concept yet," you know, figure out what went wrong."

Purpose
Adopting Gradescope promised benefits on both sides of the red pen:
(1) More efficient marking
(2) Reduced marking error
(3) Improved inter-rater and intra-rater reliability
(4) Greater marking flexibility
(5) Greater fairness in marking
(6) Improved assessment security
(7) Better student engagement with feedback

CC-BY 4.0

Process

Adopting Gradescope meant moving away from paper-based hand-marked assignments, to scanning students' written submissions – done by students themselves or administrative support staff, depending on the assessment - and uploading them to Gradescope for marking, where they were then matched against a template. From there the software's real value kicked in: grouping similar answers so the marker isn't grading the same mistake, or the same correct answer, one script at a time. The convenor described the workflow directly: "You just click on the screen all the ones that have that common mistake and make a group...Then you click on all the ones that have the right answer and make a group...You click on all the blank ones and make a group and then you group as much as you can." Once grouped, a rubric item attached to the whole batch at once – multiple scripts marked in the time it used to take to mark one. Marking could run either direction too, building up from zero of trimming down from full marks, whichever suited the question.

Feedback got more precise as a result. The convenor could choose whether comments were visible to everyone or only to the student they applied to – meaning a student's feedback could be entirely their own: "You can select what is specific to them. So when they see their marked script, it's that much quicker and more relevant to read through the feedback because it's specific to you and the mistakes you made." A built-in regrade request feature gave students a direct, structured way to flag potential marker errorsonce results were released.

Lessons Learnt

Marking through Gradescope turned out to double as a teaching tool. Watching answers accumulate in real time during tutorials or in class let the convenor spot patterns as they emerged, allowing them to quickly offer individual or group feedback to students or adjust their teaching. The convenor could also see whether students were actually opening their feedback - closing the loop that paper-based marking never could. Marks could even be used deliberately as a nudge, awarding a point simply for encouraging students to review their scripts or to draw attention to something particular in an answer.

The gains went beyond feedback. Totals stopped getting mis-added. The grind of marking large stacks of scripts eased. Consistency held up better over the length of a marking session, and when the convenor did misjudge something, it was easy to fix since Gradescope let memos be amended mid-marking. It was also easy for students to flag which they valued as a way to identify errors. In high-stakes exams especially, the regrade option functioned as "safety nets" for everyone involved. And it tightened security – fewer scripts lost in transit, less room for tampering.

Regrade requests, in practice, were rare - and usually the result of a genuine misunderstanding rather than a fishing exercise. Still, the convenor found it worth setting a boundary: after three unsuccessful regrade requests, a student would lose the option to request further reviews. The effect wasn't to shut students out, but to make them look harder before they came knocking- while still protecting them when a real error had occurred: "it's so much fairer to students when there was a mark in error. Perhaps the marker didn't notice this or the student did this totally different style of answer."

Adequate resourcing was central to Gradescope's usefulness. A fast, good quality scanner is essential infrastructure, and administrative staff had to pick up new skills such as scanning scripts - though this largely just swapped one set of manual tasks (alphabetising and loading marks by hand) for another (scanning, roster corrections). Where students uploaded their own scripts, they needed to learn the skill too.

When the convenor checked in with colleagues across the department, the preference for Gradescope over hand-marking was near-universal - and the single biggest factor in getting comfortable with the tool was simply watching someone else demonstrate it first. They encouraged colleagues to provide similar demonstrations.

CC-BY 4.0

Next Steps

The convenor's verdict was unambiguous: highly satisfied, strongly recommended. The one wishlist item was for Gradescope to get smarter over time - learning a courses' rubric patterns and a marker's preferences well enough to start suggesting rubric items itself.

Relevance Policy Principles

By leveraging AI features in Gradescope, the convenor made marking meaningfully more feasible at scale (Principle 6) and gave students a genuinely usable channel for feedback (Principle 2). The fist also strengthened reliability (Principle 4), fairness (Principle 5), and assessment integrity (Principle 7) all at once.

Case Study Contact

Anita Campbell (UCT, ADP: EBE): anita.campbell@uct.ac.za

CC-BY