Alternative Grading at Scale: Insights from Implementing Weekly Checkpoint Quizzes in a Large Introductory CS Course

Sadia Sharmin
Computer Science
University of Toronto
Toronto, Ontario, Canada
s.sharmin@utoronto.ca

Paul He
Computer Science
University of Toronto
Toronto, Ontario, Canada
paulhe@cs.toronto.edu

Abstract

In our pilot study, we implemented bi-weekly low-stakes checkpoint quizzes, inspired by alternative grading paradigms such as mastery learning, within a large (500-student) introductory programming and proofs course. The quizzes offered students frequent feedback and each quiz could be attempted a total of three times without penalty.

We collected survey data from 491 students through Likert-scale questions about confidence, stress, and learning, as well as open-ended questions for qualitative feedback. Based on quantitative data analysis looking at mean ratings, and qualitative analysis involving a grounded coding process to identify emerging themes, we found that student response was overwhelmingly positive, with the majority of students reporting improved perceived self-assessment, reduced stress, and greater engagement.

Our poster contributes to the growing conversation around flexible and equitable assessments and scalable alternative grading practices.

CCS Concepts

• **Social and professional topics → Computer science education**.

Keywords

mastery, CS1, alternative grading, qualitative, surveys

ACM Reference Format:
Sadia Sharmin and Paul He. 2025. Alternative Grading at Scale: Insights from Implementing Weekly Checkpoint Quizzes in a Large Introductory CS Course. In *25th Koli Calling International Conference on Computing Education Research (Koli Calling '25), November 11–16, 2025, Koli, Finland.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3769994.3770057

1 Introduction

First-year CS students often learn at different paces, leading some students—often those from minoritized backgrounds or without prior programming experience—to struggle with a lower sense of belonging [3, 7]. Others may develop a false sense of proficiency often due to the trending usage of AI tools for homework, resulting in inaccurate self-evaluation and subsequent inadequate preparation for major supervised assessments [4].

To address these challenges, drawing from various existing implementations of alternative grading, consulting guidelines and recommendations for such assessments, and keeping the challenges of scale and effectiveness in mind, we developed a "mastery learning"-informed approach that aims to balance flexibility with structure.

In the Fall 2024 offering of our large (~500 student) undergraduate introductory programming and proofs class, we introduced bi-weekly low-stakes checkpoint quizzes. For each checkpoint quiz, students received full credit for any score above 77% and were allowed up to two retakes, as needed. (Note: The threshold of 77% was used as this is currently our department's threshold requirement for program entry). The first two attempts were written, while the third was a verbal interview with an instructor or Teaching Assistant, serving as an opportunity for one-on-one support. To prevent procrastination and scheduling conflicts, retake deadlines were strictly spaced one week apart.

Our course had 12 weeks of class, and 10 checkpoint quizzes in total (the first and last week did not have a quiz). Each week included three two-hour lectures on Mondays, Tuesdays, and Thursdays, and weekly two-hour tutorial sessions on Fridays facilitated by Teaching Assistants during which the weekly checkpoint quizzes took place. Each quiz focused on one unit of the course (the same unit discussed during the previous 1-2 weeks of lecture); the quizzes were fairly short (about 3-4 questions) and meant to be completable within 30 minutes (although students were provided the full two hours, to reduce stress and to support students working at different paces).

2 Results

Across the 10 quizzes, an average of 34% of students completed a second attempt, and an average of 10% of students completed a third attempt (interview with instructional staff). To explore the impact of checkpoint quizzes on student experience, we surveyed all students once before their first midterm and again before their final exam. The survey combined Likert-scale questions about confidence, stress, engagement, and learning benefits with two open-ended questions for qualitative feedback. Our data collection and analysis procedures were approved by our institution's ethics review board.

The mean ratings on the quantitative survey questions (presented below) show that students believed the quizzes had the following positive impacts (higher score = stronger agreement):

- Improved self-assessment:

Koli Calling '25, Koli, Finland

ACM ISBN 979-8-4007-1599-0/25/11
https://doi.org/10.1145/3769994.3770057

  - How useful do you find the checkpoint quizzes in helping you assess your knowledge of course material? - 4.3/5 (midterm survey), 4.4/5 (final survey)
  - How accurately can you assess your understanding of the material after completing a checkpoint quiz? - 4.1/5 (both surveys)
- Reduced stress/anxiety:
  - How much does knowing you have the option to retake checkpoint quizzes reduce your stress? - 4.3/5 (midterm survey), 4.5/5 (final survey)
  - How helpful do you find the retake opportunities are for managing your stress about mastering course material overall? - 4.4/5 (both surveys)
- Improved learning:
  - How helpful are the checkpoint quizzes in identifying which concepts need more of your attention? - 4.3/(midterm survey), 4.4/5 (final survey)
  - How effective are checkpoint quizzes in helping you master the course content over time? - 4.1/5 (midterm survey), 4.2/5 (final survey)

Our two open-ended questions were: (1) "What suggestions would you make to improve the checkpoint quizzes or how they are used, for future offerings of this course?", and (2) "Do you have any other thoughts you want to share?". We conducted a qualitative analysis of the open-ended responses guided by Miles et al.'s [5] grounded theory procedures. The following recurring themes were found to be most prevalent across both surveys:

(1) **Praise for Checkpoint Quizzes:** The open-ended question responses were largely positive; 108 of 174 (62%) responses in the pre-midterm survey and 98 of 176 (56%) in the final survey expressed praise and appreciation for the quizzes. Even in response to the question asking for improvements, many students described the quizzes as "good", "great", "amazing", or "perfect".
(2) **More difficult questions:** The second most prevalent theme across both surveys was a desire for more difficult questions, to better align the quiz difficulty to that of the midterm/final exams. To preserve the "low-stress" nature of the current quizzes however, some students suggested adding an *optional* challenge question.
(3) **More comprehensive coverage:** Several students wanted each week's quiz to incorporate material from previous units, rather than each quiz focusing on one chapter. Students also suggested longer quizzes with more questions and a larger spread of points would help reduce the weight of each individual mistake and provide a more balanced assessment of their understanding.
(4) **Feedback and Solutions:** Several students requested that solutions be posted for each quiz shortly after grading. In our course, we chose not to post solutions for the checkpoint quizzes and instead encouraged students to work through the solutions on their own or seek personalized guidance and feedback from instructors.

Other less common, yet notable, suggestions we wanted to take note of for future iterations of the quizzes were: making the unfamiliar format of the third attempt interviews feel less daunting, computer-based quizzes rather than handwritten ones, more TA interaction during tutorials rather than simply conducting quizzes, and a different weekly schedule that allowed more time to review content before quizzes (e.g. specifically, they proposed scheduling the first quiz attempt on the Monday following the week's lessons, rather than on Friday of the same week).

3 Implications

Based on both the quantitative and qualitative analysis, it is clear that the general sentiment regarding the checkpoint quizzes was largely positive. Several students expressed enthusiastic support for their implementation, referring to them using words such as "perfect" and "amazing", and wishing for them to be used in their future courses. Students cited the quizzes as a valuable tool for self-assessment, which aligns well with the goal of ensuring students are adequately prepared for midterms and finals. The feedback highlights that the quizzes were an effective way to gauge their readiness and identify areas that needed further attention. Feedback from neurodivergent students and those with mental health issues (e.g. anxiety) particularly highlighted the ability to accommodate diverse learning needs with flexible assessment structures. Several students mentioned that the course design (specifically the weekly feedback, and the ability to retake the quizzes) made them feel like the course instructors genuinely cared about their success. This sentiment aligns well with our key goal being to improve student well-being and feelings of inclusivity.

On the other hand, the feedback also highlighted areas for improvement related to quiz difficulty, grading, interview structure, and quiz modality. Future iterations of the checkpoint quizzes could benefit from incorporating optional challenge questions to better align quiz difficulty with midterm and final exams, while maintaining the quizzes' low-stress nature. Additionally, revising grading criteria to reduce the impact of minor mistakes is being considered. Lastly, to further enhance the effectiveness of the quizzes overall, we have planned to administer programming quizzes in staffed computer labs for future offerings of the course (rather than the hand-written format used in our pilot study). This would allow for automated testing, providing students with immediate feedback, while also reducing administrative workload for instructional staff. We intend to examine the impact of these computerized checkpoint quizzes in future studies.

Overall, this study explores the pedagogical benefits and potential future improvements for a promising "mastery learning"-informed implementation suitable for large courses. Checkpoint quizzes, with their low-stakes nature and multiple attempts, provide students with meaningful opportunities to engage with course material and receive frequent feedback in a supportive environment while preparing for larger assessments. Learning from previous alternative grading attempts within CS courses and guided by recommendations about such approaches [1, 2, 6, 8], our implementation aimed to balance flexibility with structure, giving students opportunities to revisit and refine their understanding while maintaining steady progress through the course material. We hope these findings encourage other instructors to explore integrating mastery-learning principles within large-scale undergraduate courses for more inclusive and flexible assessments.

References

[1] Claudio Álvarez, Nickolas Falkner, Päivi Kinnunen, Jaromir Savelka, and Lisa Zhang. 2025. Show Me the Mastery Learning! Obstacles to Adoption and Opportunities for New Solutions. In *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 1*. 639–645.

[2] James Garner, Paul Denny, and Andrew Luxton-Reilly. 2019. Mastery learning in computer science education. In *Proceedings of the Twenty-First Australasian Computing Education Conference*. 37–46.

[3] Michail N Giannakos, Ilias O Pappas, Letizia Jaccheri, and Demetrios G Sampson. 2017. Understanding student retention in computer science education: The role of environment, gains, barriers and usefulness. *Education and Information Technologies* 22 (2017), 2365–2382.

[4] Rabab Marouf, Iouri Kotorov, Yuliya Krasylnykova, and Hamna Aslam. 2024. Evidence of Learning in the Presence of AI Tools in Computer Science Higher Education. In *KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications*. Springer, 331–341.

[5] Matthew B Miles and A Michael Huberman. 1994. *Qualitative data analysis: An expanded sourcebook*. sage.

[6] Linda B Nilson, David Clark, and Robert Talbert. 2023. *Grading for growth: A guide to alternative grading practices that promote authentic learning and student engagement in higher education*. Routledge.

[7] Eric Roberts. 2000. Strategies for encouraging individual achievement in introductory computer science courses. In *Proceedings of the thirty-first SIGCSE technical symposium on computer science education*. 295–299.

[8] Claudia Szabo, Miranda C Parker, Michelle Friend, Johan Jeuring, Tobias Kohn, Lauri Malmi, and Judithe Sheard. 2025. Models of Mastery Learning for Computing Education. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1*. 1092–1098.