Investigating the Efficacy of Mastery-Based Tests in Fostering Effective Self-Regulated Learning Behaviors in CS1 Courses

Joyce Gill*
gilljoyc@grinnell.edu
Grinnell College
Grinnell, Iowa, USA

Michael W. Asher
masher@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Paulo F. Carvalho
pcarvalh@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Abstract

Given the cumulative nature of computer science, success in introductory computing (CS1) courses requires students to not only learn the material but also develop effective self-regulated learning (SRL) habits. While theories of SRL emphasize planning, performance, and self-reflection as essential phases of effective learning, there is limited evidence on how to help learners put these phases into practice. In this context, Mastery-Based Tests (MBT), which allow students to retake assessments after receiving feedback, have shown promise for improving learning outcomes. However, prior work in computer science is largely observational and does not directly test MBT's impact on SRL behaviors. This paper presents a pilot study (N = 6) exploring this relationship in CS1. Using a between-subjects design, we observed that learners who first completed an MBT achieved higher post-test scores, demonstrated lower metacognitive calibration errors, and self-reported more productive SRL behaviors. These patterns suggest that MBTs warrant further investigation as a viable scaffold for fostering self-regulation in CS1.

CCS Concepts

• **Applied computing** → **Education**; • **Social and professional topics** → **Student assessment**; **CS1**.

Keywords

Mastery-Based Tests; Self-Regulated Learning; CS1

ACM Reference Format:
Joyce Gill, Michael W. Asher, and Paulo F. Carvalho. 2026. Investigating the Efficacy of Mastery-Based Tests in Fostering Effective Self-Regulated Learning Behaviors in CS1 Courses. In *Proceedings of the 57th ACM Technical Symposium on Computer Science Education V.2 (SIGCSE TS 2026), February 18–21, 2026, St. Louis, MO, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3770761.3777163

1 Problem and Motivation

Introductory computer science (CS1) courses are notorious for high failure rates, averaging 30% worldwide [3]. A primary cause is the cumulative nature of the curriculum, a phenomenon known as Learning Edge Momentum (LEM), where misunderstanding an early concept can trigger a cascade of failure as later topics build upon each other [1]. As a result, students must not only learn complex material, but also develop strong self-regulated learning (SRL) skills to accurately identify and remediate their own knowledge gaps. Unfortunately, traditional instruction, with fixed timelines and broad summative assessments, makes it difficult for students to pinpoint the specific source of their struggles [1, 6].

Mastery-Based Tests (MBT) address this problem by providing targeted feedback and multiple attempts for students to achieve mastery [2, 6]. While MBTs have shown promising results in CS1 [1], existing work is largely observational, and the link between MBT and the underlying student behaviors remains underexplored.

This paper addresses this gap with a controlled experiment designed to isolate the causal impact of MBT on student SRL behaviors in CS1. Our work is guided by the following research questions:

- **RQ1** How do mastery-based tests shape students' self-regulated learning behaviors in CS1 courses?
- **RQ2** Do these changes in self-regulated learning behavior lead to improved learning outcomes?

2 Background and Related Work

2.1 Self-Regulated Learning

Prior work characterizes effective SRL as a three-phase cycle: (1) *Planning*, where students set study goals; (2) *Performance*, where students engage in deliberate practice; and (3) *Reflection*, where students evaluate outcomes to adjust future study [2, 7–9].

2.2 Mastery-Based Tests

MBTs are an assessment approach that evaluates a student's understanding of specific learning objectives, rather than measuring overall performance on a mixed set of problems [5, 7]. To do this, students receive targeted feedback on specific learning objectives and multiple opportunities to retake tests. MBTs create a structured loop of performance, reflection (via feedback), and re-planning that could scaffold the development of these crucial SRL skills. Figure 1 further demonstrates this interaction [2].

2.3 Related Work

Prior work has examined MBT outcomes in classrooms, but evidence on how MBT affects SRL behaviors in the context of computer science is sparse and predominantly observational, making it difficult to unify results and establish strong causal claims [1, 2, 4, 6, 7].

A close example is a quasi-experiment by Capovilla *et al.* [4], which showed that students who took a mastery-learning based CS0 course performed significantly better on the practical part of a subsequent CS1 course. However, the authors cautioned about its

*Work conducted over a summer during an NSF REU at Carnegie Mellon University's Human-Computer Interaction Institute (CMU HCII)

SIGCSE TS 2026, St. Louis, MO, USA

ACM ISBN 979-8-4007-2255-4/2026/02
https://doi.org/10.1145/3770761.3777163

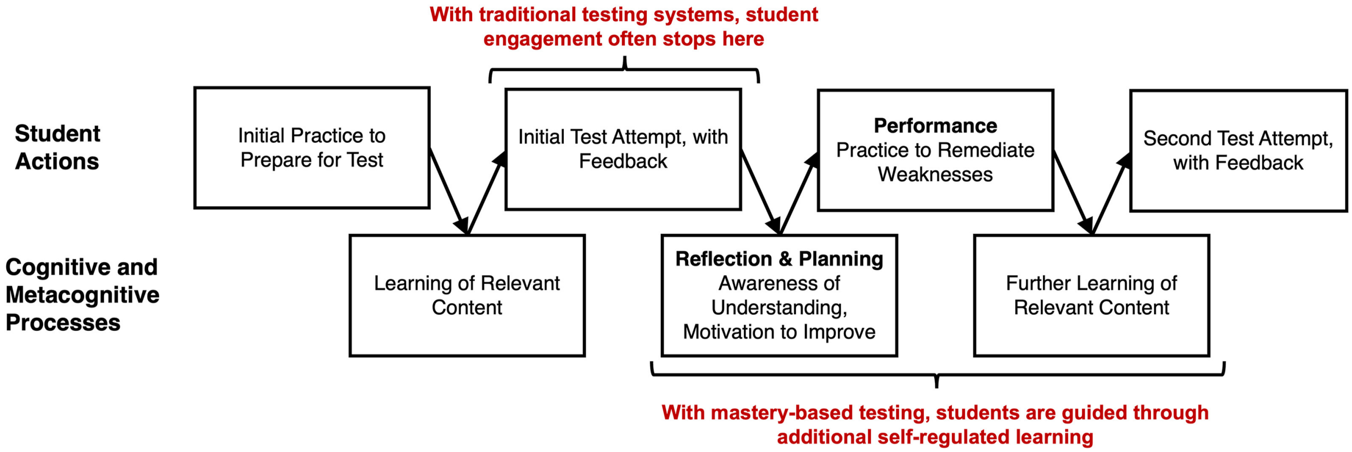

Figure 1: This diagram illustrates the relationship between student actions and their subsequent cognitive and metacognitive processes. The structured, diagnostic feedback provided by MBTs is designed to scaffold the core SRL cycle, specifically the metacognitive processes of reflection, planning, and practice. Adapted from [2]

generalizability due to a probable self-selection bias, as participation in the CS0 course was voluntary.

On the other hand, Asher *et al.* [2] investigated MBTs in a general chemistry course (N = 234) and found behavioral evidence that MBTs can effectively motivate SRL. However, the study relied on correlational methods and was conducted in a chemistry setting, limiting both causal claims and generalizability to computer science.

Thus, a clear need remains for a controlled experiment to isolate the causal effects of MBTs on SRL behaviors specifically within CS1.

3 Approach and Uniqueness

To test whether MBTs can foster productive SRL behaviors, we built an introductory Python course on Carnegie Mellon University's OLI Torus platform and designed a controlled experiment.

3.1 Learning Environment

The course had four fundamental learning objectives: (1) Identify and use different variable types; (2) Display text and values using print() statements; (3) Create and assign values to variables; and (4) Modify variables using simple expressions. To explicitly support SRL, we integrated three learning scaffolds:

3.1.1 Objective-forward pages. Each page of our learning module displays the relevant learning objectives at the top, coupled with a bucket-style mastery progress tracker, giving students a clear framework for setting study goals.

3.1.2 Objective-organized practice. Practice questions were grouped by objective, allowing students to engage in deliberate practice by targeting areas of weakness.

3.1.3 Objective-linked feedback. After each question, students received personalized feedback in the following structure: (1) correctness; (2) the correct solution; (3) the associated learning objective; and (4) a direct link to the corresponding course section. This immediate, objective-linked feedback directed students to review specific concepts.

3.2 Experimental Design and Procedure

We conducted a between-subjects pilot study and recruited participants from Prolific. Participants were screened to ensure they had no prior programming experience and were compensated for their time.

All participants were first given 15 minutes to learn the material through our course. We then randomly assigned participants to either the control or experimental (MBT) condition. Next, both groups received a total of 18 minutes to practice. The control group utilized the full duration for unstructured self-study. The experimental group, first completed an 8-minute MBT, followed by 10 minutes of self-study. Finally, both groups concluded with a 2-minute post-cognitive questionnaire and an 8-minute post-test. However, due to participant attrition during the study protocol, our final dataset consisted of 6 participants (N = 4 experimental, N = 2 control).

3.3 Measures

We used data from the post-cognitive questionnaire and post-test to examine three main outcomes:

3.3.1 Self-reported study strategies. The post-cognitive questionnaire asked participants how they decided what to focus on during practice (e.g., reviewing problems they struggled with, practicing problems in the given order). We used these responses to characterize self-reported SRL strategies.

3.3.2 Post-test scores. Learning outcomes were measured using a post-test with ten questions that aligned with the four course objectives. Each item was scored as correct or incorrect, and we report total percentage scores.

3.3.3 Metacognitive calibration error. Immediately before the post-test, participants completed a post-cognitive questionnaire in which they predicted their score on a scale of 0-100%. We operationalized metacognitive calibration error as the absolute difference between each participant's predicted and actual post-test scores, with smaller differences indicating better calibration.

4 Results and Contributions

Our pilot study yielded promising results, suggesting that the MBT may foster effective SRL behaviors. Qualitatively, participants in the experimental group explicitly reported using the MBT to guide their practice. For example, one participant stated they *"used the MBT to figure out what to practice,"* while another noted that *"the test showed [them] exactly what [they] didn't understand."*

These qualitative insights aligned with our quantitative observations, as demonstrated in Figure 2. Three out of four participants in the experimental (MBT) group achieved a perfect score on the post-test. They also showed lower metacognitive calibration error, with the same three participants perfectly predicting their scores. The fourth participant slightly overestimated their performance, predicting 70% but achieving 60%. In the control group, one participant earned a perfect score and predicted it correctly, while the other scored 90% while estimating 80%.

Although both groups performed well overall, a larger proportion of MBT participants achieved higher post-test scores and lower metacognitive calibration errors. Given the small sample size, however, these patterns should be viewed as preliminary and descriptive, rather than conclusive.

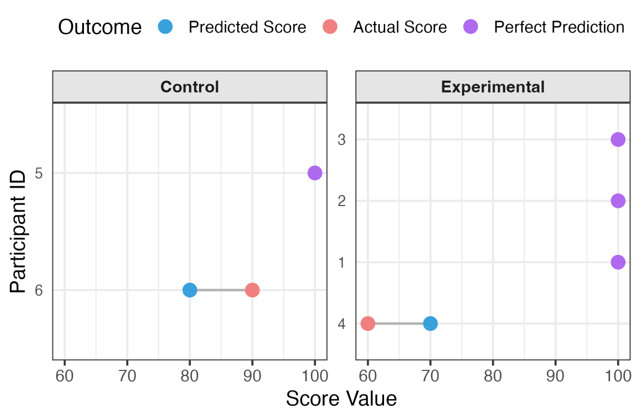

Figure 2: Predicted vs. actual post-test scores for each participant in the control and experimental conditions. Each dumbbell connects a participant's predicted score (blue point) to their actual score (pink point), with perfect prediction being highlighted by a single, overlapping point (purple point).

4.1 Limitations and Future Work

As a small pilot study (N = 6), our work was primarily intended to validate the experimental design and instruments rather than to draw definitive conclusions. In addition, participant attrition led to unequal group sizes (N = 4 experimental, N = 2 control), further limiting our ability to compare conditions. Future work should scale this design to a larger and more balanced CS1 sample to enable more robust statistical analyses of the causal impact of MBTs on SRL.

Our current characterization of SRL behaviors is also limited by relying on overall test scores and brief self-reported items. Future research should incorporate fine-grained log data (e.g., which practice problems students choose, how often they revisit objectives, how they follow feedback links) to more precisely capture how MBTs shape SRL behaviors.

Finally, we examined only a single implementation of an MBT with a fixed structure. Future work should systematically vary key design parameters–such as the number and timing of retakes and the alignment between objectives and items–to investigate which MBT configurations most effectively foster self-regulation in CS1.

4.2 Contributions

To our knowledge, this is the first study to experimentally test how MBT directly influences SRL behaviors in CS1. Our design intentionally isolates the immediate impact of MBTs on students' study choices, moving beyond the correlational findings of prior work. This work provides a validated framework for future experiments on leveraging assessments that foster effective SRL habits in CS1. Ultimately, this line of research can inform practical guidelines for designing more efficient and supportive CS1 courses.

Acknowledgments

This material is supported by the National Science Foundation under Awards No. 2418656 and No. 2349558. Thank you to everyone involved with the CMU HCII REU Program. Special thank you to Professors Samuel and William Rebelsky of Grinnell College.

References

[1] Claudio Alvarez, Maira Marques Samary, and Alyssa Friend Wise. 2024. Modularization for mastery learning in CS1: a 4-year action research study. *Journal of computing in higher education* 36, 2 (2024), 546–589.

[2] Michael W. Asher, Joshua D. Hartman, Mark Blaser, Jack F. Eichler, and Paulo F. Carvalho. 2025. The promise of mastery-based testing for promoting student engagement, self-regulated learning, and performance in gateway STEM courses. *Computers Education* 237 (2025), 105387. doi:10.1016/j.compedu.2025.105387

[3] Jens Bennedsen and Michael E. Caspersen. 2019. Failure rates in introductory programming: 12 years later. *ACM Inroads* 10, 2 (April 2019), 30–36. doi:10.1145/3324888

[4] Dino Capovilla, Marc Berges, Andreas Mühling, and Peter Hubwieser. 2015. Handling Heterogeneity in Programming Courses for Freshmen. In *2015 International Conference on Learning and Teaching in Computing and Engineering*. 197–203. doi:10.1109/LaTiCE.2015.18

[5] 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* (Sydney, NSW, Australia) *(ACE '19)*. Association for Computing Machinery, New York, NY, USA, 37–46. doi:10.1145/3286960.3286965

[6] Andrew Luxton-Reilly, Brett A. Becker, Yingjun Cao, Roger McDermott, Claudio Mirolo, Andreas Mühling, Andrew Petersen, Kate Sanders, Simon, and Jacqueline Whalley. 2017. Developing Assessments to Determine Mastery of Programming Fundamentals. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education* (Bologna, Italy) *(ITiCSE '17)*. Association for Computing Machinery, New York, NY, USA, 388. doi:10.1145/3059009.3081327

[7] Claudia Ott, Brendan McCane, and Nick Meek. 2021. Mastery Learning in CS1 - An Invitation to Procrastinate?: Reflecting on Six Years of Mastery Learning. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1* (Virtual Event, Germany) *(ITiCSE '21)*. Association for Computing Machinery, New York, NY, USA, 18–24. doi:10.1145/3430665.3456321

[8] Ernesto Panadero. 2017. A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology* 8 (2017), 422.

[9] Barry J. Zimmerman. 2002. Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice* 41, 2 (2002), 64–70. arXiv:https://doi.org/10.1207/s15430421tip4102$_2$ doi:10.1207/s15430421tip4102_2