Determining the Best Policies for Second-Chance Tests for STEM Students

Mr. Chinedu Alexander Emeka, University of Illinois at Urbana-Champaign

Chinedu Emeka is a PhD Candidate in Computer Science at the University of Illinois at Urbana-Champaign. His research interests include Computer Science Education and improving assessments for CS and other STEM students. Mr. Emeka also has a passion for teaching CS, and he has received two awards for his teaching.

David H. Smith IV, University of Illinois at Urbana-Champaign

David Smith is a PhD candidate at the University of Illinois at Urbana-Champaign in the area of Computers and Education. He has experience teaching multiple computer science courses and has played a central role in creating curricula that is used for teaching and testing hundreds of introductory CS students at the University of Illinois. Prior to joining University of Illinois he completed his B.S. in Computer Science at Western Washington University.

Prof. Craig Zilles, University of Illinois at Urbana-Champaign

Craig Zilles is a Professor in the Computer Science department at the University of Illinois at Urbana-Champaign. His research focuses on computer science education and assessment with an emphasis on effective teaching at scale.

Prof. Matthew West, University of Illinois at Urbana-Champaign

Matthew West is a Professor in the Department of Mechanical Science and Engineering at the University of Illinois at Urbana-Champaign.

Dr. Geoffrey L. Herman, University of Illinois at Urbana - Champaign

Dr. Geoffrey L. Herman is the Severns Teaching Associate Professor with the Department of Computer Science at the University of Illinois at Urbana-Champaign.

Prof. Timothy Bretl, University of Illinois at Urbana-Champaign

Timothy Bretl is a Severns Faculty Scholar at the University of Illinois at Urbana-Champaign, where he is both Professor and Associate Head for Undergraduate Programs in the Department of Aerospace Engineering. He holds an affiliate appointment in the Coordinated Science Laboratory, where he leads a research group that works on a diverse set of projects in robotics and education (http://bretl.csl.illinois.edu/). He has received every award for undergraduate teaching that is granted by his department, college, and campus.

Comparing Second-Chance Testing Grading Policies for Effective Mastery Learning in STEM Courses

Abstract

In this full research paper, we examine various grading policies for second-chance testing. Second-chance testing refers to giving students the opportunity to take a second version of a test for some form of grade replacement. Second-chance testing as a pedagogical strategy bears some similarities to mastery learning, but second-chance testing is less expensive to implement. Previous work has shown that second-chance testing is associated with improved performance, but there is still a lack of clarity regarding the optimal grading policies for this testing strategy. We interviewed seven instructors who use second-chance testing in their courses to collect data on why they chose specific policies. We then conducted structured interviews with some students (N = 11) to capture more nuance about students' decision making processes under the different grading policies.

Afterwards, we conducted a quasi-experimental study to compare two second-chance testing grading policies and determine how they influenced students across multiple dimensions. We varied the grading policies used in two similar sophomore-level engineering courses. We collected assessment data and administered a survey that queried students (N = 513) about their behavior and reactions to both grading policies. Surprisingly, we found that the students' preference between these two policies were almost perfectly split. We conclude that there are likely many policies that perform well by being simple and encouraging serious attempts on both tests.

Introduction

Tests can be used not just to assess student learning, but to promote learning as well. Second-chance testing is a pedagogical approach that can be deployed with summative assessment to improve learning outcomes and students' experiences in courses. Second-chance testing refers to offering an assessment optionally a second time for some form of grade replacement. The two assessments are typically referred to as the first-chance exam and the second-chance exam (or retake exam). The first- and second-chance exams generally have the same structure, content and difficulty, although questions are not typically identical for both assessments [1, 2].

Prior work (discussed below) has shown that second-chance testing benefits learning, but there is a lack of clarity regarding the best grading policies for second-chance testing. For many instructors new to second-chance testing, a "full replacement" policy that takes the maximum of the student's first- and second-chance exam scores seems good, because, after all, if a student

masters the material to achieve a score on an exam, it is only fair to give them that score.

In practice, however, instructors that we interviewed found that such a full replacement grading policy led students to procrastinate and disincentivized studying for the first-chance exam. Many students would not study for the first exam and a few would not even take the first exam, because, in their minds, it does not really count. An important aspect to the learning benefit of second-chance testing is that students study for the first exam, receive feedback on what they do not understand, and then remediate that material so they can demonstrate the knowledge on the second-chance exam. As such, the instructors that we interviewed have employed a range of policies that have one feature in common: the student's score on the first exam contributes to their final score, at least a little.

In this work, we explore the relative merits of a pair of grading policies. The two grading policies were as follows:

- "One-third two-thirds" policy, with insurance: Under this policy, a third of the grade came from the first-chance exam and two-thirds came from the maximum of the first-chance and second-chance score. This means that students could never be worse off after taking a second-chance test. The fact that a grade could not go down is referred to as "insurance" or "grade protection." This was the policy used in course A.
- "90-10" policy: Under this policy, 90% of the grade came from the higher of the two assessment scores and 10% of the grade came from the lower of the two assessment scores. This was the policy used in course B. This policy offers more potential for students to improve their grades from taking the second-chance exam, but their grades could go down.

We considered two research questions related to the grading policies:

RQ1. Do students prefer a particular grading policy? Why?

RQ2. How does the choice of grading policy affect student studying, anxiety, and willingness to take second-chance exams?

We explore these questions through a quasi-experimental study where we administered second-chance testing in two similar courses, but varied the second-chance grading policies used. Afterwards, we surveyed students about their preferences and how the policy affected their studying habits, anxiety, learning, and experiences in courses. The survey was supplemented by a small number of faculty and student interviews.

Related Work

Second-chance testing as a testing strategy is supported by research on the testing effect and test-potentiated learning [1, 3, 4, 5, 6, 7]. According to the testing effect, the act of taking tests (i.e. engaging in retrieval practice) leads to better retention than simply restudying the same material [6]. The theory of test-potentiated learning suggests that taking a test primes a person to better retain knowledge if it is followed by a subsequent study session [5]. Tests can therefore facilitate learning, in addition to gauging students' knowledge.

There is evidence that second-chance testing is beneficial to student learning [1, 8]. Studies report

increases in student grades and generally observe a reduction in failure rates from the addition of second-chance testing [1, 9, 10, 11, 12]. Second-chance testing also reduces test anxiety [13].

Despite the potential benefits of second-chance testing, caution is needed when determining the policies for courses. A large variety of grading configurations exist. For instance, a simple average can be used (i.e. 50% comes from the first-chance score and another 50% comes from the second-chance score, if the second-chance exam is taken). Another grading policy involves a cap on the points that can be regained (e.g. students can only get a maximum of 90% even if they score a 100% on the second-chance). Previous work has explored some specific grading policies to determine how the policies used may influence performance [1, 2]. The choice of policy may also influence who attempts the retakes [2]. Herman et al. found that marginally competent students were more likely to retake exams with a policy that does not cap the students' possible scores if the second-chance exam is taken.

Previous work suggests that the policies can have an impact on behavior and experiences in courses, but has mostly focused on an analysis of students' grades. Herman et al. [2] attempted to use several proxies to measure some criteria of interest, such as time spent on practice exams as an indicator of students' overall studying habits, but it is unclear if the measures used are able to accurately gauge students' behavior and experiences, particularly the sentiments of students in relation to the specific policies used. The previous work has significant gaps, including how policies influence other dimensions of interest, such as students' anxiety, study habits and experiences in courses. Our work fills those gaps. Our work also discusses the differing goals of various faculty members who use varied second-chance testing policies, and whether the faculties' objectives align with student behavior.

Method

Data for this paper was collected from three sources: faculty interviews, student interviews, and student surveys. All data was collected from the same large university in the midwestern US. All aspects were approved by our Institutional Review Board.

We conducted Zoom interviews with seven faculty members who had offered second-chance testing previously. Collectively, the instructors had used second-chance testing in ten distinct courses across a range of STEM departments: Chemistry, Computer Science, Aerospace Engineering, and Mechanical Engineering. Faculty were not remunerated for their time. These interviews included questions about the grading policies used in their respective courses, why they chose those policies, and their primary aims and considerations when administering second-chance testing. The faculty had taught the courses for several semesters. The faculty interviews motivated this research.

We conducted Zoom interviews with eleven students who had taken a course that used second-chance testing. The courses were drawn from the same diverse set of departments listed above (i.e. Chemistry, Computer Science, Aerospace Engineering and Mechanical Engineering). The interviews were largely used to guide the survey design, but will also be referenced briefly in the results.

The student surveys are our primary source of data and were deployed in the Spring 2022 semester in conjunction with a quasi-experimental study. At the institution, there exist two sophomore-level required courses (course A and course B) in the same engineering department that are often taken concurrently. The organization of the two courses and their course policies are very similar, as they are overseen by the same group of faculty. Both courses employ second-chance testing. In Spring 2022, these courses employed the two different grading policies, in an attempt to decide which policy would be more beneficial for students.

Each course had 6 quizzes, and each quiz offered a second chance. The quizzes were worth a total of 40% of the final grade in both courses. The first and second chance quizzes were spaced one week apart in the two courses. The courses had a similar numbers of students. Course A had 401 students and course B had 491 students.

We administered surveys to students in both courses to ask about their experiences with second-chance tests in the courses. The surveys were comprised of multiple choice questions, Likert items, and free response questions on students' sentiments regarding both policies, such as their overall preferences, study habits and the importance of policy features ("insurance" versus the ability to recoup more points). We also collected quiz performance data with students' consent. Out of the 401 students in course A, 265 students filled out the survey, a response rate of 66.1%. In course B, 247 out of 491 students filled out the survey, a response rate of 50.3% .

Analysis: For analysis, we focused primarily on the students enrolled in both course A and B who filled out both surveys. There were 98 such students; 71 (72.4%) of them were male, 21 (21.42%) were female, and 6 (6.12%) did not specify gender. Using the subset of students who took both courses and filled out both surveys allowed us to further mitigate the impact of any possible confounder between the two courses. Students in both courses are better able to make comparisons between the two policies. We analyzed both sets of data, but only report on results from one survey because the results were qualitatively the same.

The free-response questions, which allowed students to expound on their sentiments, were analyzed qualitatively. We used grounded theory for our analysis [14, 15]. Two researchers coded the students' responses to a set of four questions.

- Which of the two grading policies do you prefer? And why?
- How would the choice of these two policies affect your studying? If the policy would have no effect, please indicate so.
- Does one of these policies make you more anxious about taking tests than the other? Why?
- Is having two courses that have retake quizzes in the same semester overwhelming?

The two researchers coded the students' responses independently, but met at regular intervals to reconcile the code book and the set of codes assigned to each student response. There were 72 codes in total, and each student response received one or more codes. To ensure inter-rater reliability, a third researcher independently coded one third of the students' responses, which were picked randomly. Inter-rater reliabilty was found to be satisfactory (Krippendorff's alpha α = 0.84). From the codes, we were able to identify themes in the data. A similar approach was taken to analyze the transcripts from the faculty and student interviews, but there the bulk of the

data was coded by a single member of the research team after an initial calibration.

Results

Faculty Interviews

Instructors indicated a number of motivations for including second-chance testing in their courses. First, they wanted to encourage remediation; they wanted to motivate students to address their misunderstandings instead of just moving on to the next topics in the course. Second, instructors wanted to encourage metacognition. Most of the courses that were employing second-chance testing were mainly freshmen- and sophomore-level courses where students may not have fully developed study strategies appropriate for college. Therefore, they wanted students to think more about how they approached studying and learning course material and they felt that second-chance testing facilitated this. Third, a majority of the instructors indicated that they offered second-chance testing to reduce students' stress. Interestingly, two faculty members had a contrasting view, fearing that second-chance testing may inadvertently induce student stress, resulting from the increased testing frequency.

The instructors used a variety of grading policies in their courses. As previously noted, no instructor used a "full replacement" policy that would have allowed students to recoup all of the points on a second-chance exam. The instructors indicated that early attempts using full grade replacement led to student procrastination. One instructor reported that he had initially allowed for full replacement, but then a non-negligible number of students did not show up for the initial exam, defeating the purpose of offering second-chance tests.

The policies included the aforementioned "90-10" and "one-third two-thirds" policies and a simple average ("50-50"), among others. The instructors provided a range of explanations for their choice of grading policies. The instructor that used a simple average ran second-chance exams that used the same problems but with different numbers than the first-chance exam; less "come back potential" on the second exam was in line with the lower effort of mastering the specific questions that one had already seen. Another instructor said they were trying to encourage lower performers to retake the tests, but discourage high scorers from retaking by using a complex grading scheme that limited the points high first-chance scorers could earn. Instructors that included insurance did so to encourage retakes. Those that did not include insurance wanted to discourage "frivolous retakes" where students did not study in between the two exams.

Overall Preference

Most students indicated that the choice of policy had either a moderate or a significant impact on their decision to do the retake, as opposed to a minor or no impact (81 vs. 17 students). From the structured interviews with 11 students, it appears that students who earn a B grade on a first-chance exam are most likely to be infuenced by the grading policy. Students indicated that they would be more likely to retake the quizzes regardless of the grading policy used if they had earned a C grade or lower. Conversely, the students stated they would be unlikely to complete the retakes regardless of the grading policy used if they had scored in the A range.

Although the survey results indicate that the choice of policy did have an impact on students' decision to take second-chance tests, neither policy was preferred by the majority of students. A

total of 44 students indicated they preferred or strongly preferred the 90-10 policy and 43 students stated that they preferred or strongly preferred the "one-third two-thirds" policy with insurance, as shown in Figure 1. This indicates that there are distinct groups of students. There was no correlation between students' score goals, i.e. the minimum desirable score, and their policy preferences. However, interestingly, students with the highest score goals did not have a policy preference (i.e. the students who answered "Neutral" when asked which policy they preferred had the highest expectations for their performance in both courses).

Looking at students' explanations of their policy preferences, the most frequently expressed preferences (27 students, 27.6%) was for the guarantee that their scores would not be decreased, even if they could only receive relatively modest score boosts from second-chance testing. These students were risk-averse and leaned towards protecting the scores they had already earned.

> I don't like risk.
>
> [I prefer the "one-third two-thirds policy" with insurance] because typically students that have to attempt a retake quiz do not understand the material well, so having the possibility of dropping your grade further is not beneficial to these students.

A second group of students (21 students or 21.4%) preferred the ability to gain back more points, at the risk of their scores decreasing further. We coded these students as the "Prefers High Risk High Reward" group. Some of the students in this category stated that if they were going to take a second-chance test, then their score on the first-chance quiz was likely low and there was no need to "insure" or "protect" that score. The insurance feature was described as moot. Other students in this group stated that they would study harder for the second-chance quiz, which could reduce the likelihood of their scores going down.

> I rarely, if ever, do worse on the retakes as I tend to study more and make sure the content is solidified. So I would prefer to improve my grade more.

Studying

To measure studying habits, we looked at a Likert item that asked, 'Have the differences between these two retake grading policies affected how you approach studying in the two courses?' Results are shown in Figure 2. The responses ranged from 'No Impact' to 'Significant Impact.' We then looked at the free-response question that asked students to elaborate on their studying habits.

In response to the Likert item, most students stated that the second-chance grading policy did not have an impact on their studying habits: 60 students (61.2%) indicated the policy had a minor impact or no impact on studying, as opposed to 38 students (38.8%) who indicated that the specific policy used influenced studying habits.

The students who reported that the policy had no impact on their studying generally stated that they tried to study hard the first time around irrespective of policy, in order to maximize their grades or to avoid having to complete a second-chance quiz because of the workload involved.

> [The policy] has no effect because I study to try to avoid doing a retake in the first place.

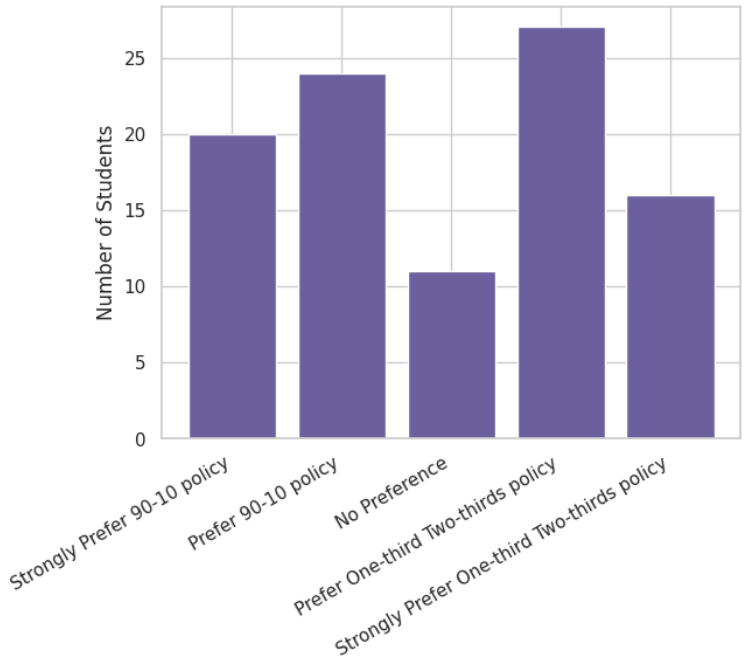

Figure 1: Overall Policy Preference

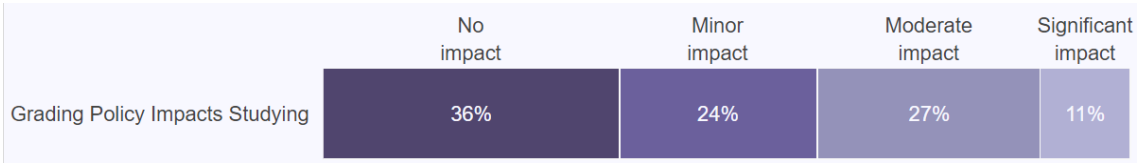

Figure 2: Does Grading Policy Impact Studying Behavior?

> [The policy] would not affect my studying. I want to do as well as possible on my first attempts no matter what.

However, for those whose studying was influenced by the policy used, they preferred the 90-10 policy because there was a higher potential to recover from a poor grade on an assessment. From the qualitative analysis, 17 out of 98 students (17.98%) indicated that they would study harder under the "90-10" policy, as opposed to 4 out of 98 students (4%) who stated they would study more under the "one-third two-thirds policy" with insurance. The students indicated that they could get a large point boost from second-chance testing under the "90-10" policy, which motivated them to study harder for the retakes (second-chance tests).

> In general, I'll study harder for [course's B] retakes because of the potential payoff

Additionally, the 90-10 policy did not have insurance, unlike the two-thirds one-third policy, which some students indicated led them to study more seriously so as to minimize the possibility that their scores would decrease on second-chance assessments.

> I would say I study harder for [retakes in course B] because I know the retakes could lead to a lower score.

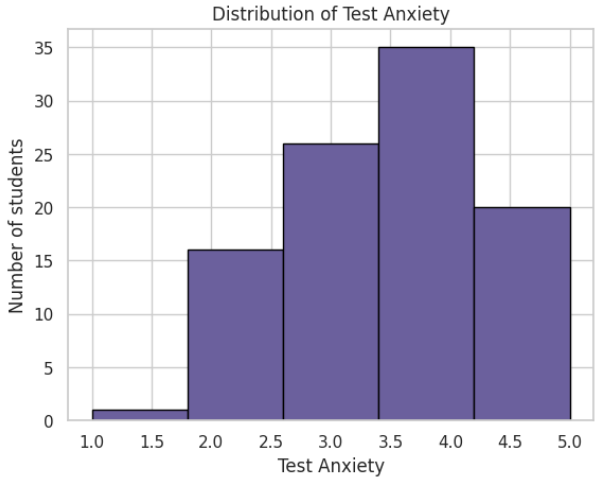

Figure 3: Students' Test Anxiety

Therefore, the 90-10 policy may lead to increased study over the two-thirds one-third policy.

Anxiety

We sought to determine which policy was better at mitigating test anxiety. Test anxiety refers to anxiety in evaluative situations, and is often problematic for students. Test anxiety can lead to diminished academic performance [16]. We measured overall test anxiety of the students in both courses by using 3 Likert items which were combined to form a Likert scale for test anxiety. The Likert items asked students to rate their nervousness during tests and whether that made them forget facts, whether they felt very tense during tests, and whether thoughts about poor performance interfered with their concentration while testing. Each Likert item had a 5 point scale that ranged from "Strongly Disagree" (1) to "Strongly Agree" (5) and included a "Neutral" option (3). The questions on test anxiety were motivated by a short-form of Spielberger's Test Anxiety Inventory (TAI) [17], which may be used when constraints prohibit administering a full instrument on test anxiety. We computed the average anxiety for each student. The distribution of anxiety is shown in Figure 3. The average anxiety score was 3.6 with a standard deviation of 1.0.

A large fraction, 55 students (56.1%) indicated they had high levels of test anxiety, i.e. they marked "Agree" and "Strongly Agree" on the Likert items for anxiety. To ascertain how the policies influenced anxiety, we examined students' answers to a free-response queston, 'Does one of these policies make you less anxious about taking tests?'

The results revealed that both policies reduced test anxiety, although this effect was achieved through different ways.

A plurality of students indicated that the two-thirds one-third policy with insurance led to less anxiety, specifically because of the insurance component. 33 students (33.7%) said that insurance or protection reduced anxiety.

> The [policy with insurance] makes me less anxious because I know taking the retake

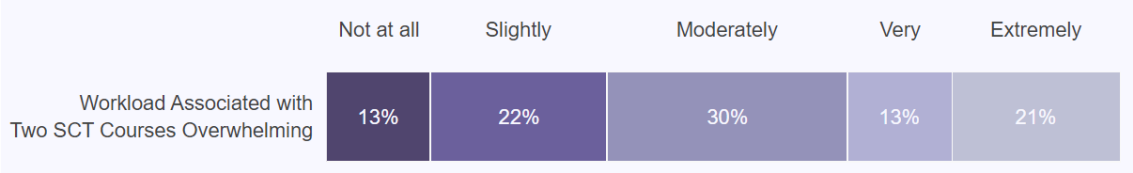

Figure 4: Workload Associated With Second-Chance Testing

> can not hurt my grade.
>
> [The policy with insurance makes me less anxious] as I know there won't be a penalty for doing worse.

Although more students reported a reduction in test anxiety under the "one-third two-thirds policy" with insurance, some students (19 or 19.4%) stated that the higher recovery potential under the "90-10" policy was better at reducing anxiety, because the students could re-coup most of the lost points.

> [The "90-10 policy"] is much safer because you can still get 90 even if i get a 0 on the first.

A third, smaller group of students indicated both policies were equally effective, or alternatively neither worked at reducing anxiety.

> They both make me less anxious knowing there's an opportunity to improve.
>
> [The two policies] both make me just as anxious. With the [90-10 policy], if you do worse then you can get a lower combined score but with [the two-thirds, one-third policy], you can't change your score that much.

The "one-third two-thirds policy" with insurance was associated with less worry about assessments and second-chance testing specifically, which may encourage more students to utilize the retake opportunity under that policy.

Workload Associated with Retakes

Second-chance testing provides opportunities for remediation, but there is some concern among instructors about the additional workload associated with second-chance testing. Therefore, we surveyed students on whether the workload associated with second-chance testing was manageable, particularly when students were simulateneously taking two or more courses that offered second-chance testing.

We included a Likert item that asked about whether the workload was overwhelming. Second-chance testing was expected to lead to extra work for students who decided to complete the retakes because of the additional time spent on testing and possibly remediation, so we phrased the question so as to capture only an increase in work that went beyond a level that may be ordinarily expected. Results from that item are shown in Figure 4. The results show that sentiment is almost evenly split among students: 35 students (35.7%) indicated that the workload was not all overwhelming or just slightly overwhelming, while 34 students (34.6%) indicated that the workload was very or extremely overwhelming.

We included a follow-up free-response question, asking students to expound further on if and why they thought second-chance testing was overwhelming when used in more than one course at a time.

A majority of students (50 students, 51%) stated that the presence of second-chance testing in two courses led to multiple quizzes a week or bi-weekly quizzes, which could potentially make it difficult to manage both courses.

> Because with the retake, I have 2 quizzes every week plus trying to study for the new material, it is a lot and more room to fall behind.

> Having retakes in two classes could lead you to having 2 quizzes every week of the semester. Because you have the option of a retake, some people may find themselves obligated to do the retake just to improve their score a little bit. The fact that it's offered could push people to do it. Having one class with retakes isn't bad, but I could see how having two classes could be very overwhelming.

From the results, we observe that students are concerned about the testing frequency. Nevertheless, the number of students who reported the retakes were manageable even for two courses was larger than those who said the retakes were not manageable (17 students, 17.3% vs. 10 students, 10.2%).

> Having a retake option will always be less stressful than having no retake option, so I prefer that.

Still, a small group of students opted to avoid the retakes because of the extra time needed to prepare.

> Really trying not to have to take two quizzes per week with retakes so often I am accepting a score I otherwise would have tried to retake

The qualitative results show that although second-chance testing is optional, it may still be considered burdensome by students. One possible reason why a group of students would say that the workload associated with second-chance testing is overwhelming is because for both course A and B, second-chance testing was combined with frequent testing. Students in the survey data were taking both courses concurrently. Because of the large number of quizzes, students may have felt fatigued from testing regardless of the availabilty of second-chance testing and may have confounded the two testing approaches (frequent testing and second-chance testing). In spite of these workload concerns, a large majority of students (83 students, 83%) indicated that they would like to have second-chance testing available in more of their courses; they responded with "Agree" or "Strongly Agree" to a Likert item that stated, "I wish more of my courses would offer retake quizzes or exams."

Limitations and Future Work

Some of the results may have been confounded by the differences in difficulty of the second-chance tests when compared to the first-chance tests in both courses. For course B in particular, students reported that the second-chance tests were harder than the first-chance tests. This may have dissuaded students from taking the second-chance tests in course B, irrespective of

course policy. We have revised the second-chance tests in course B in the current semester, and plan on re-running our experiment.

A majority of students in both courses participated in the survey. Our response rates for the two courses were 66.1% and 50.3% for course A and B, respectively. However, there may still be selection bias if the students who did not respond had different reactions to second-chance testing than those who responded. In future work, we hope to survey a larger cross-section of students across more STEM departments to determine if current findings hold and bolster transferability.

Discussion and Conclusion

Second-chance testing is a testing strategy that can lead to improved performance, albeit with additional effort on the part of the students. While students largely like having second-chance testing, they are split about what aspect of a grading policy is most valuable to them: "insurance" which prevents their score from going down by taking a second-chance exam or a larger opportunity to improve their score by taking the second-chance exam. Insurance seems to be favored by test anxious students and the larger come back potential may provide additional encouragement to study.

An obvious question to ask would be "why not provide both insurance and a high come back potential?" This is a great question, but we fear that doing both will bring us too close to a full replacement grading policy and its negative consequences. A few students report studying less for the first-chance exam in the 90-10 grading policy and others suggest that they would study less for second-chance exams that provided insurance. If we offer both, it could be that we encourage both kinds of sub-optimal student behaviors, as was observed for full replacement.

Based on the findings from this paper, our current belief is that the grading policy should provide some "friction" relative to a full replacement policy, but it may not be important exactly what that friction is. By friction, we mean features that encourage broadly studying for the first exam and focused study of misunderstood concepts for the second exam. Both the 90-10 and one-third two-thirds policies in large part do this, and there are likely other policies that would result in the same behavior.

Acknowledgements

We would like to thank Rishi Gulati for his help with reviewing our code book and for ensuring that the coding scheme was reliable. This material is based upon work supported by the National Science Foundation under Grant No. DUE 19-15257. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] G. Herman, K. Varghese, and C. Zilles, "Second-chance testing course policies and student behavior," in *2019 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2019, pp. 1–7.

[2] G. L. Herman, Z. Cai, T. Bretl, C. Zilles, and M. West, "Comparison of grade replacement and weighted averages for second-chance exams," in *Proceedings of the 2020 ACM Conference on International Computing Education Research*, 2020, pp. 56–66.

[3] C. Yang, W. Zhao, L. Luo, B. Sun, R. Potts, and D. R. Shanks, "Testing potential mechanisms underlying test-potentiated new learning." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2021.

[4] K. T. Wissman and K. A. Rawson, "Test-potentiated learning: Three independent replications, a disconfirmed hypothesis, and an unexpected boundary condition," *Memory*, vol. 26, no. 4, pp. 406–414, 2018.

[5] K. M. Arnold and K. B. McDermott, "Test-potentiated learning: Distinguishing between direct and indirect effects of tests." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 39, no. 3, p. 940, 2013.

[6] M. A. McDaniel, J. L. Anderson, M. H. Derbish, and N. Morrisette, "Testing the testing effect in the classroom," *European journal of cognitive psychology*, vol. 19, no. 4-5, pp. 494–513, 2007.

[7] H. L. Roediger III and A. C. Butler, "The critical role of retrieval practice in long-term retention," *Trends in cognitive sciences*, vol. 15, no. 1, pp. 20–27, 2011.

[8] C. Emeka, T. Bretl, G. Herman, M. West, and C. Zilles, "Students' perceptions and behavior related to second-chance testing," in *2021 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2021, pp. 1–8.

[9] S. M. Juhler, J. F. Rech, S. G. From, and M. M. Brogan, "The effect of optional retesting on college students' achievement in an individualized algebra course," *The Journal of experimental education*, vol. 66, no. 2, pp. 125–137, 1998.

[10] O. E. Fernandez, "Second chance grading: An equitable, meaningful, and easy-to-implement grading system that synergizes the research on testing for learning, mastery grading, and growth mindsets," *PRIMUS*, vol. 31, no. 8, pp. 855–868, 2021.

[11] J. W. Morphew, M. Silva, G. Herman, and M. West, "Frequent mastery testing with second-chance exams leads to enhanced student learning in undergraduate engineering," *Applied Cognitive Psychology*, vol. 34, no. 1, pp. 168–181, 2020.

[12] C. D. Schmitz, G. L. Herman, and T. Bretl, "The effects of second-chance testing on learning outcomes in a first-year stem course in engineering," in *2020 ASEE Virtual Annual Conference Content Access*, 2020.

[13] C. Emeka, T. Bretl, G. Herman, M. West, and C. Zilles, "Comparing second-chance testing grading policies for effective mastery learning in stem courses," in *2023 ASEE Annual Conference& Exposition*, 2023, pp. 1–11.

[14] K. Charmaz and L. Belgrave, "Grounded theory. the blackwell encyclopedia of sociology," 2007.

[15] A. Strauss and J. Corbin, *Basics of qualitative research techniques*. SAGE Publications, 1998.

[16] J. C. Cassady and R. E. Johnson, "Cognitive test anxiety and academic performance," *Contemporary educational psychology*, vol. 27, no. 2, pp. 270–295, 2002.

[17] J. Taylor and F. P. Deane, "Development of a short form of the test anxiety inventory (tai)," *The Journal of General Psychology*, vol. 129, no. 2, pp. 127–136, 2002.