

An Online Assessment Strategy to Improve Student Engagement, Performance, and Retention: Certification-based Retesting

Rebecca Pierce¹, Matthew Brenneman², Martin Pierce³, Eric Wilken⁴

¹Department of Mathematical Sciences, Ball State University, Muncie, IN, USA

²Department of Mathematics, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA

³Department of Finance and Insurance, Ball State University, Muncie, IN, USA

⁴Department of Mathematical Sciences, Ball State University, Muncie, IN, USA

Correspondence: Rebecca Pierce, Department of Mathematical Sciences, Ball State University, Muncie, IN, USA.

Received: June 2, 2021

Accepted: June 28, 2021

Online Published: July 11, 2021

doi:10.11114/jets.v9i7.5252

URL: <https://doi.org/10.11114/jets.v9i7.5252>

Abstract

Many courses still utilize a traditional one chance testing model to assess student understanding. If the purpose of assessment is to reflect the mastery a student has in a course, then there is benefit for students to have multiple opportunities to show mastery. This paper outlines the results of a course policy of full grade replacement retesting that required students to first pass a “recertification” quiz. The goal of this policy was to adopt a pedagogical style that more readily reflected the opportunity of continued learning that many workers experience in the professional world while simultaneously aiming to engage students in an online course during the COVID-19 pandemic. A hypothesis test was conducted to determine if this retake policy helped to improve student grades during the course. The results indicate there was a statistically significant difference between the mean score on the first exam and the retest where, on average, students who utilized the exam retake increased their score. Time was found to have a positive relation with retest scores, but even after accounting for time, retesting was found to have a practical and significant effect on student performance. Retesting policies consistently show positive impacts on grades (e.g. Roszkowski & Spreat, 2016; Herman et al., 2019) and should be more widely considered when developing and updating course policies.

Keywords: assessment, retaking, testing, retesting, online teaching, finance

1. Introduction

Due to the COVID-19 pandemic, most courses at all levels of education were forced to be offered online in 2020. Teachers responded with a wide variety of innovative pedagogical methods to engage students in an educational format unfamiliar to many of them (for example, see Jandrić, 2020). One specific change to which teachers adjusted was the use of online examinations. This change in modality for exams provided an opportunity to reconsider the role traditional testing policies have in the modern classroom.

This paper describes the retesting policy and results of one professor’s new assessment strategy for a required finance course in a college of business at a Midwest university. The intent was to improve student engagement, performance, and retention of course material. Giving students a second opportunity to take an exam incentivizes continued engagement in the material while providing a structure for feedback within a course. Retesting may also help prepare students for adult life where one can often learn from their failures and move past them without being weighed down forever (Wormeli, 2011).

Traditional testing uses exams as an assessment for knowledge and provides students one chance to show they have mastered the material. An alternative testing strategy, mastery testing, aims to periodically evaluate if a student grasps certain objectives in the course through frequent exams. If a student does not demonstrate mastery, they do not continue on to new material and instead review until they can pass (Bangert-Drowns et al., 1991). However, under the time-constraints of a semester system, mastery testing is often not possible (Juhler et al., 1998; Herman et al., 2019). Optional retesting of exams is a compromise between traditional and mastery testing which garners students some of the benefits of knowledge retention while still trying to fit into the semester format (Herman et al., 2019). There is some concern retesting is unfair to students who succeed on the first attempt or scores may increase due to practice effects or familiarity, but this does not seem to impact the validity of exams as a rule (Geving et al., 2005; Roszkowski & Spreat, 2016).

As testing became easier to conduct at the start of the 20th century, the impact of frequent testing began to be investigated (Bangert-Drowns et al., 1991). A meta-analysis of 35 studies, conducted between 1929 and 1989, identified 37% of the studies as showing statistically significant positive effects on student learning associated with frequent testing and 3% of studies as showing statistically significant negative effects (Bangert-Drowns et al., 1991). More recent studies indicate testing itself provides an opportunity for students to improve their memory, in addition to the learning facilitated by the feedback provided on an exam (Pyc & Rawson, 2010; Agarwal et al., 2012). Exams may provide an environment where mediators (a word, phrase, or concept) are created that are more easily retrieved and decoded (Pyc & Rawson, 2010). In other words, exams or quizzes provide practice retrieving previously learned information which, in turn, allows students to better retrieve it again in the future (Agarwal et al., 2012). However, Downs (2015) found the opposite effect where the benefit of testing without feedback was not statistically significant, and the overall impact when feedback was small.

The effect optional retesting has on learning is still mixed throughout the literature. Most studies find allowing students to retake exams will increase their exam scores with diminishing returns (Cates, 1982; Rohm et al., 1986; Friedman, 1987; Kennedy, 1994; Juhler et al., 1998; Abraham, 2000; Geving et al., 2005; Kunz et al., 2011; Roszkowski & Spreat, 2016; Morphew et al., 2019). However, few studies found improved performance on a final exam at the end of the semester (Friedman, 1987; Abraham, 2000; Herman et al., 2019; Morphew et al., 2019). These increases in exam scores did not always happen for all students, rather anywhere between 30% to over 60% of students which appeared to be moderated by variables such as time between test and retest, the amount a student studied, and the number of retests available to the student (Elbrink, 1973; Geving et al., 2005). Studies capturing student attitudes found students preferred optional retesting to one shot testing and it can reduce testing anxiety (Friedman, 1987; Juhler et al., 1998). There is not a singular retesting policy used throughout the literature, rather the specific policies tend to be what fits best with a professor's teaching style or course content.

When developing a retesting policy, a number of factors must be considered such as the logistics of administering retests, the number of retests allowed, who gets to retake the exam, and how the retest will replace previous exam scores. These differences in policy have an impact on how many and which students choose to retake exams (Herman et al., 2019). Scheduling retake exams can create logistic problems in terms of additional work for the instructor outside of normal class time or it requires additional lecture/lab days be used as retake days (Kunz et al., 2011). The use of an online test proctoring system can be used to reduce some of these logistical difficulties while still helping to preserve academic integrity. The number of retakes available to a student overall and per exam can impact the overall retention of material. When a student is given unlimited attempts, they tend to view only the last one as the "real" test (Elbrink, 1973, Herman et al., 2020). Still, there may be benefit to providing additional attempts past one retake for an exam (Kennedy, 1994). In terms of who gets to take the exam, it can either be open to all students or those who score below a particular score on the first exam (Herman et al., 2019, Herman et al., 2020). Lastly, one can decide whether the score from the retested exam fully replaces the previous one or only partially (Herman et al., 2019, Herman et al., 2020).

Deciding on what policy to use for score replacement opens up the pedagogical question of "what is the purpose of a test score?" If the purpose is simply to reflect the mastery of content, then if a student can demonstrate mastery at any point they should receive credit (Wormeli, 2006; Herman et al., 2019). Often grades also try to capture if a student is on the predefined timeline a teacher wants (Wormeli, 2006). If this accountability is not the goal, then full replacement of a test grade should be done, but this full grade replacement can lead to students not trying on the first exam as they know they have another chance (Friedman, 1987; Herman et al., 2019; Herman et al., 2020). One way around this is to use a form of partial grade replacement where the two or more test scores are averaged in some way to incentivize students to try on the first chance (Herman et al., 2019; Herman et al., 2020). Grade replacement can also include caveats in the policy such as insurance, where a grade cannot go down, a grade cap, where after the first attempt there is a cap on one's max score moving forward, or imposing an additional assignment required to be completed prior to taking the retest. These strategies aim at making it worthwhile for students not to use retests as an opportunity to procrastinate (Cates, 1982; Herman et al., 2019).

The policy utilized in this paper embraces full replacement with requalification. Requalification means that a student is required to pass a quiz prior to their one retake for each exam. The policy embraces the idea that assessment indicates mastery, and a student should have the ability to fully change their grade after they had the opportunity to study more and receive feedback. The requalification quiz aims to provide a disincentive to students who did good, but not excellent, on the first attempt from taking the exam by requiring a student to put in time and effort between the first and second attempts. The requalification quiz both requires a student to put in study time to pass and also is a way to study itself (Pyc & Rawson, 2010; Agarwal et al., 2012). Requiring the requalification quiz also increases the time between test and retest which is associated with improving one's score (Geving et al., 2005).

The research problem presented in this paper is to investigate the results from a course policy of full grade replacement retesting that required students to first pass a "recertification" quiz. The first research question was whether requalification improved the student's average exam score. This was tested via the alternative hypothesis: $\mu_{\text{after-before}} \neq 0$

(where “before” and “after” refer to the scores before and after completion of a requalification quiz, respectively). The second research question was estimating the effect of requalification on the change in the average test scores, while also considering the time required to complete the initial test and the retest after requalification. This was tested via the alternative hypothesis, $\beta_1 > 0$ for the regression model:

$$D_{\text{TestScore}} = \beta_0 + \beta_1 * D_{\text{Time}} + \varepsilon$$

where $D_{\text{TestScore}}$ represents the difference in test scores, D_{Time} represents the difference in times, and ε represents the error term.

2. Method

The re-testing policy was used in 4 sections of the class “Principles of Finance” (FIN 300) at a midwestern university. A total of 143 students were enrolled in the 4 sections. The course policy consisted of 3 phases: an initial exam, a second “requalification quiz”, and a second exam. The specifics of each phase will be discussed in detail. In the remainder of paper, the words *exam* and *test* are used interchangeably.

Phase 1: Initial exam All students were given four exams throughout the course and each exam covered multiple sections of the course textbook. All of the exams were timed (2-hour limit), accessed online (via Canvas), and monitored for cheating (using Respondus Lockdown Browser + Monitor). Each exam was created from its own set of question banks, and questions for each exam were randomly chosen for each student from the question bank for that exam. Although each exam was scored as 100 points, the number of questions on each exam were not the same, nor was the size of the question banks for each exam the same (see Table 1, below). On each exam, roughly 8% of the questions were matching and fill in the blank. The remainder of the questions were multiple-choice, with about a quarter of the multiple-choice questions being computational in nature. At the end of the exam, students were able to view their total score, in addition to viewing the results for each question, where they could see whether or not they got the problem right and if they did not, what the correct answer was. However, because the students were still using the lockdown browser when they viewed the exam, they could not record this information.

Table 1. Properties of Exams Used in Phases 1, 2, and 3

Exam	Number of questions on phase 1 and 3 exams (Regular Exams)	Question bank size for phase 1 and 3 exams	Number of questions on phase 2 exam (Requalification Quizzes)
1	63	282	45 30
2	44	363	60 40
3	42	125	30 20
4	53	177	45 30

Phase 2: Requalification Quiz Students who were not satisfied with their first test score had the opportunity to take “requalification quizzes”. The requalification quizzes were similar to the first exam in content. To be granted the opportunity to retake the exam a second time, the student had to score above 50% on the requalification quiz. One potential problem with the requalification quizzes was the possibility students might simply get the same problems they had on the first test. To reduce this possibility, questions for the requalification quizzes were drawn from two question banks. The original question bank used for the first exam was split into 2 sub-banks. The first sub-bank, which was about 40% of the original question bank, had their answers modified. All the multiple-choice questions involving computations were modified (so students could not just remember which number went with the problem), while other questions were modified so the correct choice was replaced with the option “none of the above”. The remaining 60% of the questions in the other sub-bank were not modified in any way. All of the matching and fill in the blank questions were in this sub-bank. The requalification exams were not the same length as the initial exam (see Table 1). During the last 2 weeks of the semester, the number of questions drawn from the question pool changed from 15 questions per chapter to 10 questions per chapter (to prevent students from being overwhelmed with work). This is shown in the last column of Table 1; two numbers are given with the first number being the number of questions on the requalification quiz before the last 2 weeks of class, while the second number reflects the lower number of questions on the requalification quiz for the last 2 weeks of the course. The requalification quizzes were timed, but the time period was much longer (2 days) to allow the students to stop the test and review the relevant material in the text. Students could take the requalification quizzes as many times as they wished, until they either reached the goal of obtaining a score of 50% or higher or they did not reach the threshold score of 50% required to retake the exam. Those students who did not reach the threshold goal did not get to take the second exam and data on them was therefore not recorded for the purposes of this study. Most students took the requalification quiz one or two times, with a few students taking it 3 times or more.

Phase 3: Retest Those students achieving a score of 50% or higher on the requalification quiz were permitted to take the exam again. The second exam consisted of a new random selection of questions from the test bank. Assuming there were no overlapping questions between the first exam and the requalification quiz, the probability that the retest consists of all new questions is near zero for any given test. Already seen questions should not be a serious issue as there is some evidence that test scores will not substantially increase due to a student already having seen a question (Geving et al., 2005). The same time limit as the exam in phase 1 was used (2 hours) as was the use of the Respondus Lockdown Browser + Monitor.

For each exam, data were collected on the students who underwent the requalification quiz successfully. From Canvas, both the scores and the time of completion for their first exam scores and their second exam scores were obtained.

Table 2 shows the total number and the number of students (out of a total of 143 students from all 4 sections) who successfully completed the requalification process and took the second exam. To ensure the scores obtained were valid, the time it took to complete each exam was inspected. If the completion time for a student's test was unusually low (< 10 minutes), the video recording from the Respondus Browser was reviewed to check if the student actually tried to complete the exam. On the other hand, since the Respondus Browser closes the exam after the specified two-hour time limit has been exceeded, the video recording of students' having low scores but who used the full 2 hours was also reviewed to check if the exam was started but not actually attempted for whatever reason. In total, one student's score was flagged for a short completion time and deleted since the student said in the video (from Respondus) they had started the exam and realized immediately they didn't have time to finish it. Another student was flagged for having taken the entire allotted exam time but earning a very low grade. Reexamination of the Respondus Browser video, and log, (which records the detailed activity of each student) revealed the student worked just 17 minutes, so their first exam time was altered accordingly.

Table 2. Number of students who requalified and retested

Exam	Number of students who took the second exam
1	24
2	12
3	21
4	12
Total	69

Although the data is not included, it is available upon request from the authors.

3. Results

The scores from phase 1 (the original test) and phase 3 (the retest) for each student from all the exams was compiled. Table 3 shows the summary statistics for the difference in the 69 pairs of exam scores.

Table 3. Summary statistics for the differences in test scores

n	Mean	Std. dev.	Median	Range	Min	Max	IQR
69	21.93	18.84	20.02	97.07	-13.57	83.5	20.93

Plotting the data revealed the differences between the second and first exam scores had a slightly skewed distribution (see Figure 1). From a two-tailed paired t-test on the score differences, we find there is sufficient evidence to support the claim, the (population) average of the score differences is not zero. The 95% confidence interval for the differences in the average test scores, is found to be (17.8%, 26.9%). The results from the hypothesis test is shown in Table 4.

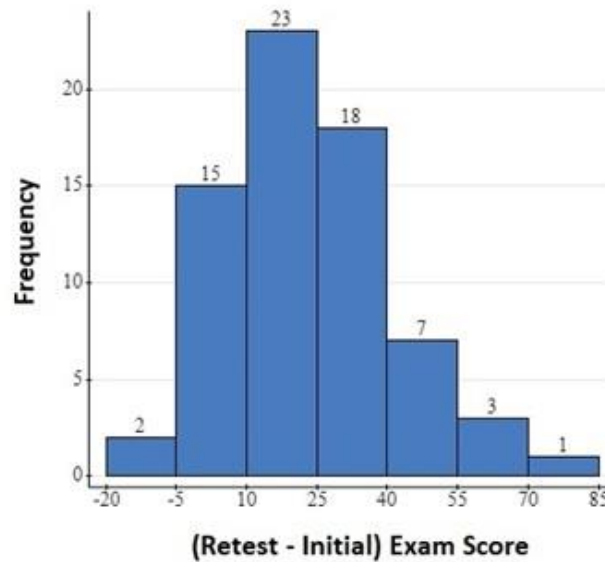


Figure 1. Distribution of Test Score Differences (n=69)

Table 4. Hypothesis test output for the mean of the difference between retake and first test score

Variable	Mean	Std. Error	DF	Test statistic	p-value
Retake – First Test Score	22.36	2.27	68	9.85	<0.0001

To account for the effect of the time taken to complete the exam on the difference in the exam scores, simple linear regression was performed using the following statistical model:

$$Y = \beta_0 + \beta_1 * X + \varepsilon$$

where $Y = (\text{Retest} - \text{Initial})$ exam score and $X = (\text{Retest} - \text{Initial})$ time to complete exam and $\varepsilon \sim N(0, \sigma)$ iid. We note β_0 accounts for the effect of the retest on the average student score, since it gives us the change in the average score due to retesting when there is no difference in the time it took to compete both the initial exam and the retest exam. The parameter β_1 accounts for the effect of the time to complete the exam on the difference in the average exam scores. The scatterplot and the best fit line for the data is shown below in Figure 2.

As expected, the plot shows a moderately strong positive relation between the change in the exam score and the difference in the time it took to compete the exam.

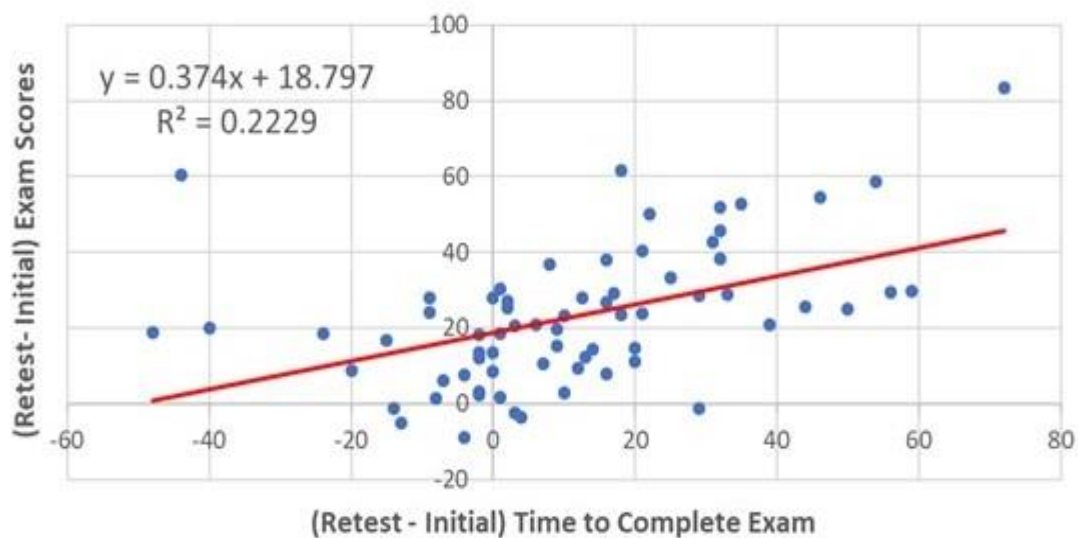


Figure 2. Relationship between Difference in TestScore and Difference in Time

A check for model validity was performed. To test for outliers in the independent variable, the leverage was computed and 4 data values were found to have values above the conservative threshold of $3p/n$, where p is the sum of the leverages taken over all the data points. One of these data points was also found to be influential, having a Cook's D over 0.5. The final model was run after these 4 data points were deleted. Deletion of these 4 data values did not have a substantive effect on any of the results we obtained.

A plot of the histogram of the studentized residuals (see Figure 3) shows they have a distribution consistent with having come from a normal distribution. Performing the Shapiro-Wilks test, a goodness fit test for normality, yields a p-value of 0.469, supporting the claim there is not enough evidence to conclude the residuals do not come from a normal distribution.

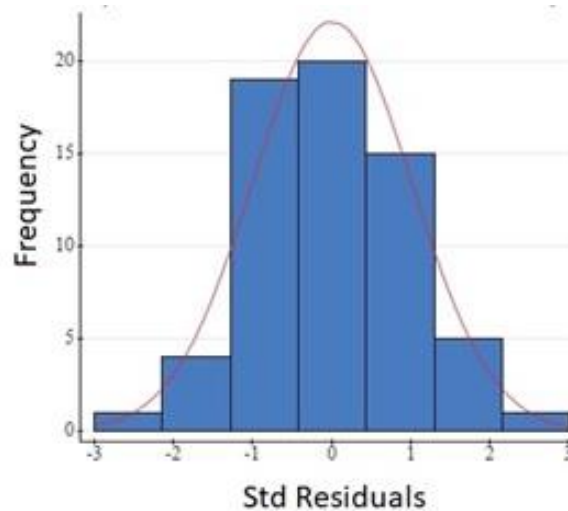


Figure 3. Standardized Residuals (with best fit normal distribution)

A plot of the residuals versus the independent variable, shown in Figure 4, was used to check the remaining model assumptions for the residuals. Specifically, the residuals are independent and identically distributed with a mean of zero and constant standard deviation. Independence of the residuals is evident in the plot since the points appear to randomly bounce around the horizontal axis and do not seem to show any sign of correlations. The identical distribution assumption is supported by the fact the spread of the residuals seems relatively uniform.

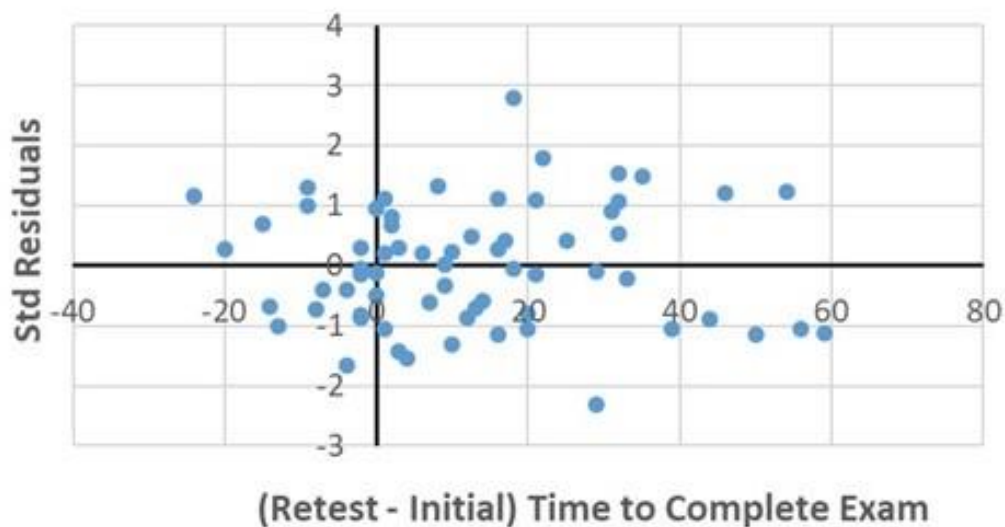


Figure 4. Residuals plot versus the independent variable

The significance of the model was determined from the p-value for the slope coefficient, which was 3.8×10^{-7} , indicating the model is statistically significant. Furthermore, the y-intercept was statistically significant (p-value = 2.6×10^{-10}). The coefficient of determination, R^2 , was 33.8%.

4. Discussion

Our paper presents two analyses of the effect of the retest policy on student exam scores. First, the results from the hypothesis test showed the mean of the difference between the retake and first test scores was statistically significant. The result of the confidence interval for the differences in the average test scores, (17.8%, 26.9%), gives an estimate of the effect considering only retesting as a factor. Second, the effect of time was considered through the regression analysis. The value of the slope, 0.505, tells us for each additional minute taken to complete the exam, the estimated increase in the average score is about one half a point. While time was found to have a positive relationship with the change of the students' scores, the coefficient of determination indicates it only accounts for about a third of the variation in the data. The value of the y-intercept tells us the average increase in the test score upon retesting (holding the time it took to complete both the initial and retest constant) is 15.1 points. This is consistent with the estimated difference in the average scores of 22.36 points, although the effect of the retest is smaller, having accounted for the effect of the time it took to take the exam.

However, caution should be made in interpreting the results from the regression analysis and the paired t-tests. For example, a student realized after taking the original test, they needed to take more time. So, although the increased time might have helped to increase their score, it was the fact they were able to retake the exam which made this possible. But without putting too fine a point on our results, it seems reasonable to conclude the effect of the retest in this study accounted for a one and a half to about 2 letter grade improvement. These results provide another replication of past studies in general showing that retesting improves exam scores (Cates, 1982; Rohm et al., 1986; Friedman, 1987; Kennedy, 1994; Juhler et al., 1998; Abraham, 2000; Geving et al., 2005; Kunz et al., 2011; Roszkowski & Spreat, 2016; Morphew et al., 2019).

This improvement highlights how a student's learning of one topic can continue alongside learning another. It also emphasizes how a student's journey to mastery on a particular topic might now follow the course schedule developed prior to the start of the semester. Retesting provided these students an opportunity to continue to learn the material, especially during a time where personal challenges, not related to school, were likely at a high level due to the COVID-19 pandemic. The course policy outlined in this paper may not produce the same results for some instructors or courses and there are adaptations that can be made to it which would likely improve its effectiveness.

One potential change to this policy is a move away from a singular test bank for the two exam attempts. While a singular test bank means that concepts are represented fair amongst the initial exam and the retake, it opens up the ability for students to have seen a question and its answer prior to taking the retest. For the first exam outlined in Table 1, the median number of questions that will match on a second random selection from the question bank computed from the hypergeometric distribution is 13. There is an approximate 43% chance that 13, or fewer questions (out of 63 questions) will be the same on a retake for a randomly generated test. In other words, there is a substantial chance that a notable portion of the retake are questions a student has already seen. This means a student potentially could study to memorize answers instead of learning in order to succeed on the retest. Though, there is evidence that the use of the repeated questions may not lead to increased exam score alone (Geving et al., 2005). One way to ensure that students do not see the same questions multiple times is to create separate test banks for the initial exam and the retake, alter the test bank questions (similar to what was done for the requalification quiz) or increase the overall number of exam questions in the test bank. The higher the number of questions the lower the probability will be of overlap. Further discussion about the necessary size of question banks has been addressed in Murdock & Brenneman (2020).

Another change that can occur to the testing policy is the threshold percentage needed to requalify to take a retake. The 50% threshold was selected to make students put in the effort to study for the second chance exam, but not be so high of a level that it deterred students who would benefit substantially from the retake. Students did not find out their score on the requalification quiz until after they turned it in and got unlimited opportunities to reach the 50% threshold on the assignment. Some students took a guess and check approach until they eventually reached the threshold level while others spent hours answering a portion of the questions correctly and then ignoring the remaining questions. The goal of the requalification quiz was to get students to continue to look at the material between retakes which occurs even when students guess and check as they are still receiving valuable feedback. Changing the threshold barrier, or the number of opportunities a student has to requalify for an exam, would also substantially change how students respond to such a policy.

Some limitations of this study are the findings are not generalizable to all students and we cannot make any conclusions regarding causality. The data collected for this study comes from a single course in a single semester during the COVID

pandemic. The composition of students in courses can vary largely which further emphasizes the ways this study may not generalize to all classrooms. It is also worth noting the condition in which the data were collected were not “normal” in the sense a pandemic will, hopefully, not be a concern in the future. It would be worthwhile to collect a larger sample of students in courses with a similar course policy on retesting across multiple disciplines to understand the overall effectiveness of the policy.

Variations in course policies may be fruitful endeavors to investigate moving forward alongside instructor variation. Retesting policies have a substantial impact on student behavior including exam scores (Herman et al., 2019). Though, it would be impactful to better understand how these policies interact with the personality and teaching modality of different instructors. Teaching does not occur in a bubble and a particular retesting policy may work great for one instructor but may not work for another and understanding why this occurs may lead to better future prescription of policy.

Overall, this study reinforces support for retesting policies and provides more evidence for instructors to consider adding them to their courses. Adding a retesting policy should not just be a shift in wording in a syllabus, rather should involve reflecting on one’s own teaching philosophy. In general, the addition of a retesting policy is beneficial, but understanding how it fits into an instructor’s academic ideology is important to make a sustainable shift in the way that assessments are viewed and administered.

References

- Abraham, P. (2000). A Microscopic Look at Assessment: Dropping a Lowest Test Score versus Allowing a Retake Test. *Proceedings of the 1999 Regional Campus Mathematics Conference*.
- Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. *Educational Psychology Review, 24*(3), 437-448. <https://doi.org/10.1007/s10648-012-9210-2>
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C. (1991). Effects of Frequent Classroom Testing. *The Journal of Educational Research, 85*(2), 89-99. <https://doi.org/10.1080/00220671.1991.10702818>
- Cates, W. M. (1982). The Efficacy of Retesting in Relation to Improved Test Performance of College Undergraduates. *The Journal of Educational Research, 75*(4), 230-236. <https://doi.org/10.1080/00220671.1982.10885386>
- Downs, S. D. (2015). Testing in the college classroom: Do testing and feedback influence grades throughout an entire semester? *Scholarship of Teaching and Learning in Psychology, 1*(2), 172-181. <https://doi.org/10.1037/stl0000025>
- Elbrink, L. C. (1973). A Flexible and Forgiving Testing Program A Report of the Current Status of the Testing Program of the CRIMEL Project. *School Science and Mathematics, 73*(8), 686-690. <https://doi.org/10.1111/j.1949-8594.1973.tb09141.x>
- Friedman, H. (1987). Repeat Examinations in Introductory Statistics Courses. *Teaching of Psychology, 14*(1), 20-23. https://doi.org/10.1207/s15328023top1401_4
- Geving, A. M., Webb, S., & Davis, B. W. (2005). Opportunities for Repeat Testing: Practice Doesn't Always Make Perfect. *Applied H.R.M. Research, 10*(2), 47-56. <https://doi.org/10.1037/e518612013-432>
- Herman, G. L., Cai, Z., Bretl, T., Zilles, C., & West, M. (2020). Comparison of Grade Replacement and Weighted Averages for Second-Chance Exams. *Proceedings of the 2020 ACM Conference on International Computing Education Research*. <https://doi.org/10.1145/3372782.3406260>
- Herman, G., Varghese, K., & Zilles, C. (2019). Second-chance Testing Course Policies and Student Behavior. *Proceedings of 2019 IEEE Frontiers in Education Conference (FIE)*. <https://doi.org/10.1109/fie43999.2019.9028490>
- Jandrić, P., Hayes, D., Truelove, I., Levinson, P., Mayo, P., Ryberg, T., ... Hayes, S. (2020). Postdigital dialogue. *Postdigital Science and Education, 2*, 163-189. <https://doi.org/10.1007/s42438-020-00169-6>
- Juhler, S. M., Rech, J. F., From, S. G., & Brogan, M. M. (1998). The Effect of Optional Retesting on College Students' Achievement in an Individualized Algebra Course. *The Journal of Experimental Education, 66*(2), 125-137. <https://doi.org/10.1080/00220979809601399>
- Kennedy, R. (1994). A Comparison of Limited vs. Unlimited Retakes of a Multiple-Choice Test. *Proceedings of the 1994 Mid-South Educational Research Association Annual Meeting*.
- Kunz, R., Bubacz, M., & Mahaney, J. (2011). Retests: A Rescue Plan for the Sophomore Slump. *Proceedings of the 22011 ASEE Southeast Section Conference*.

- Morphew, J. W., Silva, M., Herman, G., & West, M. (2019). Frequent mastery testing with second-chance exams leads to enhanced student learning in undergraduate engineering. *Applied Cognitive Psychology*, 34(1), 168-181. <https://doi.org/10.1002/acp.3605>
- Murdock, M., & Brenneman, M. (2020). Designing tests from question pools with efficiency, reliability, and integrity. *Journal of Contemporary Chiropractic*, 3(1), 92-100.
- Pyc, M. A., & Rawson, K. A. (2010). Why Testing Improves Memory: Mediator Effectiveness Hypothesis. *Science*, 330(6002), 335-335. <https://doi.org/10.1126/science.1191465>
- Rohm, R. A., Sparzo, F. J., & Bennett, C. M. (1986). College Student Performance under Repeated Testing and Cumulative Testing Conditions: Report on Five Studies. *The Journal of Educational Research*, 80(2), 99-104. <https://doi.org/10.1080/00220671.1986.10885731>
- Roszkowski, M. J., & Spreat, S. (2016). Retaking the SAT may boost scores but this doesn't hurt validity. *Journal of the National College Testing Association*, 2(1), 1-16.
- Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American Secondary Education*, 14-27.
- Wormeli, R. (2011). Redos and retakes done right. *Educational Leadership*, 69(3), 22-26.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.