Effect of online practice exams on student performance

Muxin Zhang, Andrew Engel, and Tim Stelzer
Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL, USA/61801

Jason W. Morphew
School of Engineering Education, Purdue University, 701 West Stadium Drive, West Lafayette, IN, USA/47907

Online practice exams have been important tools for students to prepare for their actual exams in our large introductory-level physics courses at the University of Illinois. Using data collected in the past few years, we found that, although students valued these online tools, their participation in the practice exams did not correlate with their actual exam performance. We also found that, with our old practice exam delivery format, students' performance in the practice exams might not reflect their actual abilities and could give them an "illusion of understanding".

2019 PERC Proceedings edited by Cao, Wolf, and Bennett; Peer-reviewed, doi.org/10.1119/perc.2019.pr.Zhang

I. INTRODUCTION

Studies have shown that practice exams are highly effective learning tools [1] and can significantly improve students' performance on their actual exams [2]. Compared to other exam reviewing strategies, practice exams are a form of testing that encourages active retrieval and helps students recognize gaps in their knowledge [3]. They can also provide students with the feedback to self-assess and develop formative studying strategies [4]. However, the effectiveness of practice exams could depend on many factors, such as, their similarity to the actual exams, students' prior abilities, and the format of feedback [1,5].

At the University of Illinois, we started implementing online practice exams with solution videos for our large introductory level physics courses approximately five years ago. Before every exam, students have access to old exam problems that have been given to the same course in previous years. Students can submit their answers online, receive immediate feedback, and watch solution videos. Clinical studies have shown that these solution videos significantly increased student learning [5]. Our end-of-semester survey also shows that students value these online practice exams more than any other component of the course including lecture, discussion, homework, and lab. In Fall 2018, 72.4% of the students reported that the practice exams were "essential" or "very important" in helping them understand the material.

Despite the positive literature and survey results, these practice exams may not be meeting the needs of some students. Anecdotally, low-performing students are often frustrated by their actual exam performance even though they have worked through multiple practice exams. A common complaint among students in our introductory level courses seems to be: "I did pretty well on the practice exams, so I thought I was well-prepared. But I did worse than I expected in the actual exam!"

Some literature shows that the effectiveness of practice exams depends a lot on the format it's delivered in and how students use it. Bol and Hacker found that in some situations practice exams can be less effective than traditional review and lead to inaccurate judgements of ability [6]. Balch found that, for two groups of students who were given access to the same practice test, the group that did the test before viewing solutions performed better than the group that only viewed solutions [7], however this may only apply to problems close to students' current ability level [8].

Anecdotes and literature prompted us to look closer into how students are using our online practice exams and examine possible issues that could have caused negative reactions from some students.

II. RESEARCH QUESTIONS

In this paper, we attempt to answer the following research questions: 1) How effective were the online practice exams? Is students' performance improving while they are using the current online practice exam tool? 2) Do online practice exams help students prepare for their actual exams? 3) By adjusting the delivery format of the practice exams, can we encourage more productive study behaviors and increase the effectiveness of practice exams?

III. PROCEDURE

Physics 211 is the first mechanics course in the introductory physics sequence for engineering students at University of Illinois. There are three "hour exams" spread out over the entire semester. Starting from one week before each hour exam, students could log into the course website (FlipIt Physics), where they access their regular online homework every week, and view up to 4 practice exams with solutions. Students also had access to paper copies of the same practice exam questions, so the online practice exams were optional. Paper copies of practice exams had answer keys in the back, but they didn't provide students full solutions. Each practice exam contained about 24 multiple-choice questions. However, in the spring of 2019 a new format for the practice exams was introduced. In this in-situ study, we will be using data from both Spring 2017 and Spring 2019 semesters, which had 1,121 and 1,016 students respectively.

In Spring 2017, as students approached each question, they could see a "submit" and a "help" button under the statement of each question. They could submit an answer, and receive immediate feedback on whether their answer was right or wrong. They could change their answer up to ten times and receive a feedback each time. Meanwhile, they could click the "help" button any time during the practice and view a solution video. Students could choose to watch the solution video before they submitted any answer, and could change their answer after they watched the video.

In Spring 2019, we adjusted the format that these practice exams were delivered such that, when students were working on the problems, they had an experience closer to what they would have in a realistic exam. Each practice exam was divided into "clusters". Students could view one cluster at a time, which contained about 3 questions. Students could submit answers to these questions, but they wouldn't get immediate feedback on whether their answers were right or wrong. Only after they had submitted answers to all the questions in that cluster, could they click "Submit Cluster", and get feedback on correctness and access the solution videos. Unlike in Spring 2017, students could not see any feedback or solutions before they submitted the entire cluster, and they could no longer change their answers after they had submitted the entire cluster and had access to the solutions.

We collected students' submissions data online for both semesters to look at how students used the practice exam tool. We also collected their actual hour exam 1 and hour exam 2 scores to look at their performance. These hour exams covered the same content as the practice exams and were written to have similar difficulties.

IV. RESULTS

A. No correlation between practice exam participation and actual exam performance

Figure 1 shows that the number of practice exam problems that students attempted and their actual exam performance are very weakly correlated (r = 0.14). Here "attempted a question" means that the student has either submitted an answer or watched the solution video for that question. This means that, in this context, higher quantity of practice exams didn't necessarily lead to improvement in exam performance. We found this result for both Spring 2017 and Spring 2019 semesters despite the format difference.

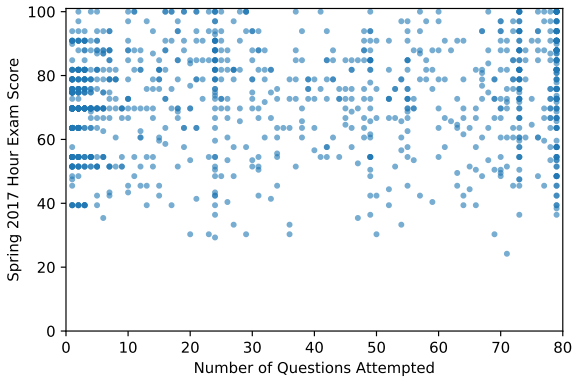
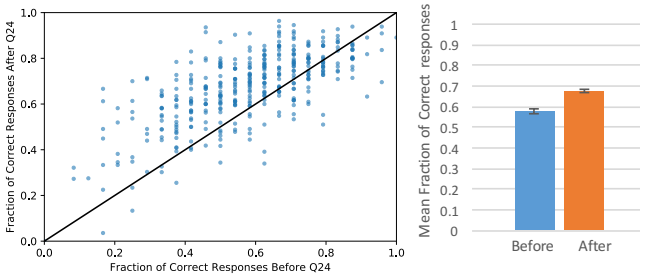


FIG. 1. Physics 211 Spring 2017 Hour Exam 1 score of each student was plotted against the number of practice exam questions that the student attempted. This plot only counts each practice question once; for example, a student who attempts a practice question and then re-attempts the same question again the next day will only have the questions counted once towards this total. The lighter dots in this graph are single students, while the darker dots have more than one students overlapping each other.

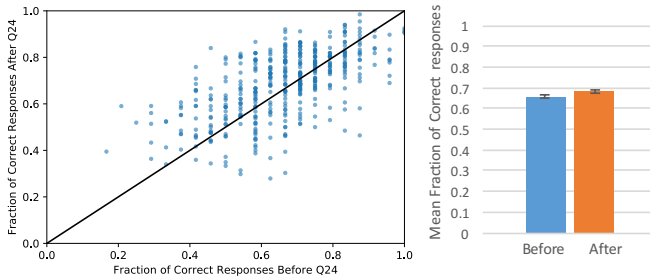
This lack of correlation doesn't necessarily mean that the practice was not effective. Students' prior ability might affect how much they participate in the practice exams and how they perform in the actual exam. Stronger students are more likely to engage in using testing as a study strategy [1]. However, weaker students are incentivized to work on more practice exam problems because they get more questions incorrect and can view more solutions.

B. No significant improvement during practice

Another way to measure the effectiveness of practice exams is to see if there is improvement within each student during the practice. In Figure 2, we compared students' performance on their first 24 practice exam questions, which correspond to the first full practice exam they did, and their performance on the rest of the practice exam questions to see if there is any longitudinal improvement.



(a) Spring 2017



(b) Spring 2019

FIG. 2. For each student, their fraction of correct responses after (but not including) Question 24 is plotted against their fraction of correct responses before (and including) Question 24. These plots only included students who answered more than 50 practice questions in total. (a) For Spring 2017, the average performance of a student before Q24 is 0.576, and after is 0.673. (b) For Spring 2019, the average performance before Q24 is 0.66, and the average performance after Q24 is 0.68.

For Spring 2017, a "correct response" is defined as, if the student answered correctly on their first submission, and did not look at help before that first submission, then the answer is counted as correct. This is an important distinction because students can submit as many answers as they please and receive feedback. However, we counted only their first submission. For Spring 2019, the format is slightly different and students could no longer access feedback or solutions after they have submitted their answers, so we simply collected their last submitted answer before they clicked the "submit cluster" button, and calculated the fraction of correct response from that.

Figure 2 shows that, for Spring 2017, there was a significant increase of fraction of correct responses from the first 24 practice questions to the rest of the practice questions by 9.6 percentage points, $t(382) = 13.8$, $p < .001$. However, we found that, in Spring 2017, most students started with a practice exam that was 10% harder than the rest of the

practice exams, where the difficulties of practice exams were measured using the mean scores students received when they took them in previous years. Due to this difficulty difference, we cannot conclude that there was a significant improvement as students worked on more practice exams.

For Spring 2019, there was a 2.5 percentage points improvement, $t(433) = 4.02$, $p < .001$, which is statistically significant but very small. A possible reason for the small learning gain may be that students take these practice exams in a later stage of their exam preparation process and they use the practice exams to measure their learning rather than as a learning tool itself.

C. Most of the practice was done less than two days before the actual exam

To look closer at how students used the online practice exams, we collected the time stamps of students' submissions online. As shown in Figure 3, almost all practice exam participation happened within three days before the actual exam. In fact, about 50% of the submissions were made within 24 hours before the exam, and about 75% of all submissions were made within 48 hours before the exam. This behavior might explain why practice exams were not effective for some students. A short amount of time for practice might not allow significant improvement in performance or better understanding of the material.

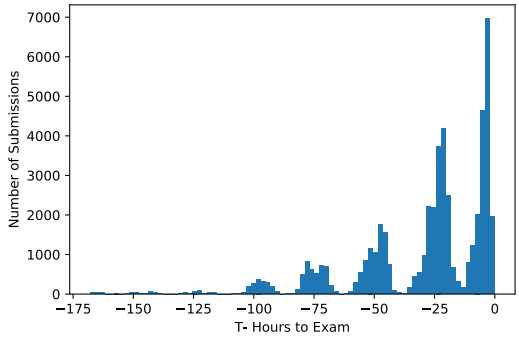


FIG. 3. Histogram of when students submitted answers to the online practice exams for Spring 2019. Each bin width is 2 hours.

D. Students' practice exam performance might not reflect their actual abilities in the old format

To answer our second research question, we want to see if students' practice exam performance predicts their actual exam performance. However, there are some subtleties in how we calculate the fraction of correct responses and how we define students' practice exam performance.

Since in Spring 2017 the format of the practice exams made it possible for students to change their answers after they got feedback and solutions, we found that most students corrected their answers such that eventually they got almost all questions correct (Figure 4). The "fraction of correct responses" calculated here is what the website records at the end of practice and, we believe, a close estimation of students' perception of their ability, which clearly does not predict their actual exam performance.

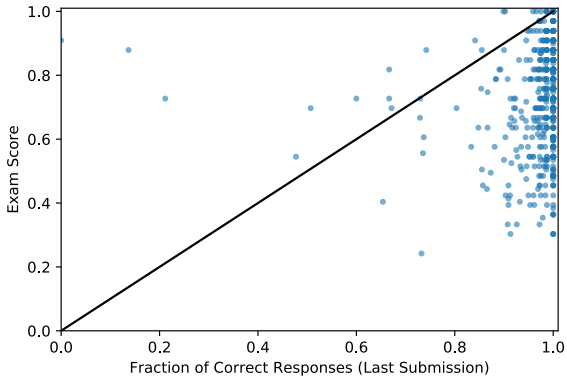
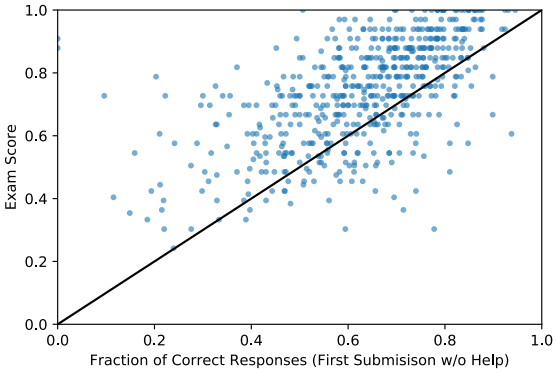
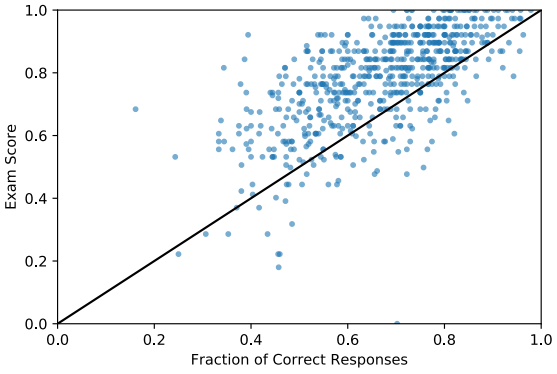


FIG. 4. For Spring 2017 Hour Exam 1, we plotted students' actual exams scores against their fraction of correct responses in practice exams using their last submitted answers. We included only students who did more than 24 practice questions and that included 566 out of 1,121 students.

If we count only students' first submitted answers without viewing any solutions, as shown in Figure 5(a), we found that their practice exam performance is closer to their actual exam performance ($r = 0.60$). However, with this format, students are likely to not use their first submitted answers to judge their ability.



(a)



(b)

FIG. 5. For both plots, we included only students who did more than 24 practice questions, which filtered out roughly 40% of the total population. (a) For Spring 2017 Hour Exam 1, we plotted students' actual exam scores against their practice exam performance using their first submitted answers before viewing any solutions. Of students plotted, 19.79% of students performed worse on the actual exam than on the practice exam (under the diagonal line). (b) For Spring 2019 Hour Exam 1, we plotted students' actual exam scores against their practice exam performance using their last submitted answers. 15.7% of students performed worse on the actual exam than on the practice exams.

After we adjusted the format, in Spring 2019, we used students' last submitted answers to calculate their practice exam performance (Figure 5(b)). The "fraction of correct responses" calculated here is what the website records at the end of practice and, we believe, the best estimation of students' perception of their ability, which does predict their actual exam performance ($r = 0.65$).

V. CONCLUSIONS

First, our data indicated that there was no correlation between the quantity of practice exams done and actual exam performance. This may be partly due to the fact that there were paper copies of the practice exams also available to the students. In addition, there may have been an interaction between prior ability and the number of problems attempted. However, we also did not find significant improvement within students during the practice. This could be due to our observation that most students did the practice exams a short time before their actual exams.

Second, we found that the old practice exam format in Spring 2017 might have given students an "illusion of understanding," where students believed that they were prepared for the exams, when in fact they were not able to solve some of the problems without the support of the solution videos. The fraction of correct responses calculated from student final submitted answers did not reflect their actual ability and might not help students predict their actual exam performance. In contrast, with the new format in Spring 2019, students' final answers better reflect their actual exam performance.

Third, comparing effectiveness, we did not find any difference between the Spring 2017 format and the Spring 2019 format. However, we found that the new format might help students better understand their ability. Since we did not ask students to explicitly predict their exam score, we cannot make any concrete conclusion about this improvement.

Overall, we have two visions for our online practice exams. First, we want them to help students learn the material and improve their performance. Second, given that prior work has demonstrated that students are often poor predictors of their level of preparedness for exams [8, 9], we want to help students develop more accurate metacognitive monitoring of their current ability.

This study helped us to understand the current effectiveness of practice exams and how their delivery format can make a difference. In future work, we want to look closer at how students interact with the online practice exams and understand how practice exams help students develop judgements of their ability. In particular, we wonder if the new format can help students predict their actual exam performance more accurately.

ACKNOWLEDGMENTS

This project was supported by NSF grant DUE 16-08002.

[1] Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. Psychological Science in the Public Interest: A Journal of the American Psychological Society, 14(1), 4–58. https://doi.org/10.1177/1529100612453266

[2] Kulik, J. A., Kulik, C.-L. C., & Bangert, R. L. (1984). Effects of Practice on Aptitude and Achievement Test Scores. American Educational Research Journal, 21(2), 435–447. https://doi.org/10.2307/1162453

[3] Roediger, H. L., III, Putnam, A. L., & Smith, M. A. (2011). Ten Benefits of Testing and Their Applications to Educational Practice. PSYCHOLOGY OF LEARNING AND MOTIVATION: COGNITION IN EDUCATION, 55, 1–36. https://doi.org/10.1016/B978-0-12-387691-1.00001-6

[4] Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74. https://doi.org/10.1080/0969595980050102

[5] Fakcharoenphol, W., Potter, E., & Stelzer, T. (2011). What students learn when studying physics practice exam problems. Physical Review Special Topics - Physics Education Research, 7(1). https://doi.org/10.1103/PhysRevSTPER.7.010107

[6] Linda Bol, & Douglas J. Hacker. (2001). A Comparison of the Effects of Practice Tests and Traditional Review on Performance and Calibration. The Journal of Experimental Education, 69(2), 133. https://www-jstor-org.proxy2.library.illinois.edu/stable/20152656

[7] Balch, W. R. (1998). Practice versus Review Exams and Final Exam Performance. Teaching of Psychology, 25(3), 181–185. https://doi.org/10.1207/s15328023top2503_3

[8] J. W. Morphew, "Metacognitive Calibration in Introductory Physics Courses: Predictors and Interventions," doctoral dissertation, University of Illinois at Urbana-Champaign (2019)

[9] Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology, 77(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121