University of Reading

Benefits of additional online practice opportunities in higher education

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Schwerter, J., Dimpfl, T., Bleher, J. and Murayama, K. (2022) Benefits of additional online practice opportunities in higher education. Internet and Higher Education, 53. 100834. ISSN 1096-7516 doi: https://doi.org/10.1016/j.iheduc.2021.100834 Available at https://centaur.reading.ac.uk/102243/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.iheduc.2021.100834

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Benefits of additional online practice opportunities in higher education

Jakob Schwerter[a,e], Thomas Dimpfl[b,c], Johannes Bleher[c], and Kou Murayama[d,e]

[a] Technical University Dortmund, Center for Research on Education and School Development, Vogelpothsweg 78 44227 Dortmund, Germany, *jakob.schwerter@tu-dortmund.de*
[b] University of Hohenheim, Department of Business Mathematics and Data Science, Fruwirthstr. 49, 70599 Stuttgart, Germany, thomas.dimpfl@uni-hohenheim.de
[c] University of Tübingen, Department of Econometrics, Statistics, and Empirical Economics, Sigwartstr. 18, 72076 Tübingen, Germany, *johannes.bleher@uni-tuebingen.de*
[d] University of Tübingen, Hector Research Institute of Education Sciences and Psychology, Europastr. 6, 72072 Tübingen, Germany, *k.murayama@uni-tuebingen.de*
[e] LEAD Graduate School & Research Network, Tübingen, Germany

Author Note

Jakob Schwerter is an affiliate at the LEAD Graduate School & Research Network [GSC1028], funded by the Excellence Initiative of the German federal and state governments.

Correspondence concerning this article should be addressed to Jakob Schwerter, Technical University of Dortmund, Center for Research on Education and School Development, Campus Nord (CDI Gebäude), Vogelpothsweg 78, 44227 Dortmund, Germany. Email: *jakob.schwerter@tu-dortmund.de*

This work was partly funded by Jacobs Foundation Advanced Research Fellowship to Kou Murayama.

Abstract

Are exam grades predetermined by students' prior performance and personal characteristics, or can underperforming students catch up? We evaluate whether additional e-learning practice opportunities improve learning outcomes for a group of undergraduate business students enrolled in a university math course (N = 281). During the semester, students were offered two types of voluntary additional e-learning practice opportunities (some earned extra credit, others did not). These practice opportunities incorporated the study techniques of self-testing and spacing, as well as knowledge of correct responses feedback. After controlling for a large number of personal characteristics, we find that voluntary practice has a statistically significant effect on exam performance, which indicates that practicing leads to better grades. Our results show that students currently performing at any level can improve their learning outcomes through additional practice. Furthermore, the overall effect is most significant for weak students who would otherwise be expected to score low on the exam.

Keywords: mathematics, learning outcomes, study techniques, knowledge of correct response feedback, external rewards

Highlights

- We analyzed the effect of additional online practice opportunities on exam points.
- Additional practice included self-testing, spacing, and direct feedback.
- University students in a math course were observed during a semester and not in a lab.
- Practice participation and performance led to better outcomes on the final exam.
- Results are robust to a rich set of control variables.

1. Introduction

E-learning practice opportunities can mitigate challenges of teaching higher education courses (Ardac & Sezen, 2002; Dusi & Huisman, 2021). When instructors design their courses, they can offer only a limited set of study materials that promote a selected set of study techniques; including too many materials will overwhelm students. Instructors face particularly narrow limitations when teaching heterogenous first-year introductory courses with a large number of students. E-learning allows instructors to offer practice materials that indirectly promote a variety of study techniques. These techniques have different effects on students' learning gains, but students (and instructors) might not know which techniques are most effective (Dunlosky et al., 2013; Karpicke, 2016). Dunlosky et al. (2013) review articles on evidence-based study techniques and points out the minimal learning gains of rereading, highlighting, and summarizing. However, these are the study techniques that most students use most frequently (Karpicke, 2016). Teaching students how to use other study techniques takes time and resources; furthermore, this instruction might not reach all students in courses where attendance is not mandatory. Therefore, it is essential to know whether e-learning practice can be used to augment a course so that students can apply more effective study techniques without explicit instruction in those techniques. Furthermore, while intrinsically motivated students will take on additional tasks (Walker et al., 2006), the additional e-learning materials should be designed to help all students in a course, particularly underperforming students. Therefore, the general objective of the present study is to analyze whether and which students benefit from additional e-learning opportunities that are based on effective study techniques.

A rich body of literature highlights the study techniques of self-testing (also called active recall) and spacing (also called spaced learning), which outperform rereading, highlighting, and summarizing (Dunlosky et al., 2013; Hartwig & Dunlosky, 2012; Rawson & Dunlosky, 2012; Rodriguez et al., 2021; Roediger & Pyc, 2012). Moreover, additional e-learning practice opportunities could be designed to provide students with immediate feedback about the correct response. Especially in classes that are too large for instructors to provide individual feedback, automated feedback is helpful (Butler, Karpicke, & Roediger III, 2008; Finn et al., 2018). However, previous studies have mixed findings on whether additional practice opportunities incorporated within the design of a course can improve students' course achievement (B. W. Brown & Liedholm, 2002; Panus et al., 2014). Therefore, more research on blended learning is needed to see if there are effective ways to provide students with practice opportunities and whether previous findings on study techniques and feedback can be translated into real-life higher education settings.

In this paper, we analyze the effect of a course augmented with e-learning practice opportunities. We observed a first-year college mathematics course with 280 undergraduate students. Students enrolled in the course were either majoring or minoring in business administration or economics at a large public German university. Therefore, we used a real-life educational setting, which is important to establish the external validity of laboratory results (Morrison & Anglin, 2005; Ross & Morrison, 1989). In addition, the study goes beyond existing results by identifying which students benefitted the most from e-learning practice opportunities. In particular, we seek to determine whether e-learning practice only helps students who are already doing well or whether additional digital practice opportunities can help underperforming students catch up.

2. Literature Review

The study relates to three strands of existing literature: study techniques (self-testing and spacing), knowledge of correct response feedback, and blended learning. In addition, to select an appropriate set of control variables, we surveyed the literature in these areas, which identifies essential predictors/drivers of students' learning and achievement.

2.1 Practice Techniques That Enhance Student Learning

The literature on study techniques identifies self-testing and spacing as some of the most effective learning techniques (Dunlosky et al., 2013; Hartwig & Dunlosky, 2012; Rawson & Dunlosky, 2012; Rodriguez et al., 2021; Roediger & Pyc, 2012). However, despite their effectiveness, students rarely use these techniques, either because they underestimate the possible gains or because they are unaware of the effectiveness of these approaches (Karpicke et al., 2009; Kornell & Son, 2009).

Self-testing is when students actively try to recall knowledge; it is a form of metacognitive monitoring (Dunlosky et al., 2013; Rodriguez et al., 2021). Dunlosky et al. (2013) demonstrate that this study technique has high utility for low-level learning. It is less time-intensive than other techniques and can be implemented with minimal explanation. Its positive effect has been demonstrated with various test formats, materials, learner ages, outcome measures, and retention intervals (Butler, 2010; Karpicke & Blunt, 2011; Spitzer, 1939). Rodriguez et al. (2021) extended the results to high-level learning in a university STEM course. During self-testing, learners assess how much of the learned material they can actively recall while solving exercises. Learners then become aware of their deficits and focus on improving these areas; this is called the *potentiation effect* (Izawa, 1971). Furthermore, students who see that their answer is wrong may benefit from the *error generation effect* (Kornell et al., 2009) or, if errors are made with high confidence of correctness, from the *hyper-correction effect* (Butler, Karpicke, & Roediger, 2008).

Spacing involves repeated learning of a specific topic over an extended period of time (in contrast to cramming). Learning gains through spacing are explained by the *forgetting curve* (Ebbinghaus, 1885; Murre & Dros, 2015). Over time, humans forget topics at an exponential rate. Interrupting this curve through spaced learning leads to longer retention of the learned material. Therefore, students can learn more and better when they spread learning sessions over a long period of time, as repeated efforts reinforce memory traces (Bjork, 1975; Rodriguez et al., 2021). In addition, longer intervals might encourage students to learn more carefully and to understand the material better than if they only cram for a few days (Melton, 1970; Rodriguez et al., 2021). Laboratory studies (e.g., Karpicke et al., 2009; Karpicke & Bauernschmidt, 2011) and intervention studies (e.g., Rodriguez et al., 2021; Stanger-Hall et al., 2011) demonstrate the potential of self-testing and spacing. The latter studies also show that explaining the benefits of spacing to students leads to more consistent learning.

Combining both of these approaches into one study promises to be particularly valuable (Cull, 2000; Rodriguez et al., 2021; Roediger & Karpicke, 2006). Spaced out self-testing should lead to more efficient encoding of the information to be retrieved, stored, and/or recalled (Jonides, 2004). This positive influence of combining self-testing and spacing on (low-level) learning has been found in numerous studies (Baker et al., 2020; Hartwig & Dunlosky, 2012; Panus et al., 2014; Park et al., 2018; Rodriguez, Kataoka, et al., 2018; Rodriguez, Rivas, et al., 2018) — even when the results are controlled for prior achievement (Rodriguez et al., 2021; Stanger-Hall et al., 2011).

An issue from a research perspective is that most previous studies on self-testing and spacing have measured declarative and procedural knowledge. One exception to this, Rodriguez et al. (2021), shows that these results also translate into high-level learning in an advanced STEM course. The additional practice opportunities analyzed in the present study are motivated and shaped by the previous findings discussed above. The course design aims to increase the time students spend on learning the material during the semester (instead of cramming just before the exam) as well as the effectiveness of that time. This should eventually set students on a more consistent learning path throughout the semester.

2.2 Feedback

Immediate knowledge of correct response feedback in e-learning environments is another key feature of our study. The effectiveness of feedback is well established and highlights a significant advantage of e-learning environments (Azevedo, 2005; Bangert-Drowns et al., 1991; Finn et al., 2018; Hattie & Timperley, 2007; Jaehnig & Miller, 2007; Kluger & DeNisi, 1996; Kulhavy, 1977; Kulik & Kulik, 1988; Wang et al., 2019). A beneficial feedback setting is characterized by four features: (1) the student is able and (2) willing to make use of the feedback (Shute, 2008); (3) the feedback is given immediately (Shute, 2008); and (4) the feedback is about the task and not the person (Hattie & Gan, 2011). E-learning environments facilitate points 1, 3, and 4. Direct feedback and the opportunity to make errors are important aspects of learning in general (P. Kirschner et al., 2006; Sloboda et al., 1996; Wisniewski et al., 2020) and of our additional practice opportunities. Immediate feedback about the correct response in e-learning settings has been shown to foster students' learning (Attali, 2015; Attali & van der Kleij, 2017).

Particularly in high-level learning in STEM subjects, students can make mistakes. They might, for example, find a result for a calculation without knowing that it is wrong. Therefore, knowledge of the correct response fosters guided learning (P. A. Kirschner et al., 2006). Guided instruction during active knowledge construction in computer-based learning environments promotes process and content knowledge (Ardac & Sezen, 2002). Thus, automated feedback after self-testing could lead to an additional boost in learning.

2.3 Additional Practice in a Blended Learning Environment

Though there is no clear definition for blended learning (M. G. Brown, 2016), it is characterized by the incorporation of digital features that accompany or partly replace a face-to-face teaching format. Thus, our additional online practice opportunities convert the original face-to-face teaching format to a blended teaching format. The potential of blended (and online) teaching is often praised. However, numerous studies find that students who are taught in a blended format do not outperform students who are taught face-to-face (Alpert et al., 2016; B. W. Brown & Liedholm, 2002; Joyce et al., 2015); adult learners may be an exception to this (Deschacht & Goeman, 2015). A possible reason for this failure of students in blended environments to outperform those in face-to-face environments is the wide range of variations in blended learning approaches. For example, Alpert et al. (2016) only replaced one lecture per week with videos without providing additional practice opportunities. Similarly, in Joyce, Crockett, Jaeger, Altindag, and O’Connell (2015), only that face-to-face and blended format taught students were required to take online quizzes. In addition, Joyce et al. (2015) document that students in the lower performance percentiles did worse in a blended environment. In that study, good students performed well in both settings, while poorer students tended to suffer in the blended format.

Brown and Liedholm (2002) is one of few studies to use the blended learning feature of additional problem sets. However, they do not specify what type of feedback was given to students or determine the impact of different types of feedback on student achievement. Still, this result could suggest that instructors could incorporate additional practice opportunities in different formats in existing face-to-face courses. However, if the newly introduced blended learning features promote self-testing and spacing or include additional feedback, we would expect students learning through this blended approach to outperform students who are only taught face-to-face. Finally, for low-level learning, only Panus, Stewart, Hagemeier, Thigpen, and Brooks (2014) find a positive correlation between more frequent self-testing during a semester and improved exam performance. However, these results could be biased by prior achievement, performance during self-testing, personal traits, or demographic characteristics. Therefore, the analysis of e-learning practice in the present study contributes to the debate over the best design for blended learning formats.

2.4 Student-Level Factors Associated with Learning Performance

The selection of control variables in the present study is based on the literature on student-level factors associated with learning outcomes. A rich set of variables is essential to understanding which students perform better than others. Important variables included in the present study related to motivation are achievement goals (mastery–approach, mastery–avoidance, performance–approach, and performance–avoidance) (Elliot & McGregor, 2001; Hulleman et al., 2010; Pintrich, 2000, 2003), self-concept, task values (intrinsic motivation, attainment value, and utility value), and cost (Eccles et al., 1983; Feather, 1992; Pintrich, 2003; Wigfield & Eccles, 2000). In addition, we surveyed personality traits as measured by the Big Five (extraversion, agreeableness, openness to experience, conscientiousness, and neuroticism)

(Digman, 1990; Goldberg, 1993), procrastination behavior (measured by present bias preferences—risk, discount factor, and present bias) (see Becker, Deckers, Dohmen, Falk, & Kosse, 2012; Frederick & Loewenstein, 2002; O'Donoghue & Rabin, 1999), prior achievement, and demographic information.

In general, achievement goals can be separated into two types: performance goals (i.e., ego orientation, relative ability, and self-enhancement) and mastery goals (i.e., task orientation, learning, and mastery challenge). They can be divided into approach and avoidance goals (Elliot & McGregor, 2001; Hulleman et al., 2010; Pintrich, 2000, 2003). Hence, mastery–approach achievement goals measure whether students want to learn and improve their knowledge on a particular topic, while mastery–avoidance is about not getting worse. The distinction is similar for performance goals, although performance goals concern achievement relative to others rather than individual achievement. Elliot, McGregor, and Gable (1999) and Harackiewicz et al. (2002) find that mastery–approach goals positively impact the educational outcomes of undergraduate students, while Plante, O'Keefe, and Théorêt (2013) deem this effect to be unstable. Yperen, Blaga, and Postmes (2014) document mixed results for performance–approach goals among university students. Avoidance approaches are negatively related to successful learning (Baranik et al., 2010; Hulleman et al., 2010; Linnenbrink-Garcia et al., 2008; Payne et al., 2007).

There are two types of expectancy-value variables. The expectancy part of the theory captures "beliefs about one's competence and self-efficacy" (Plante et al., 2013 p. 67), while the value part "refers to the reasons for engaging in a specific task" (Plante et al., 2013 p. 67). Specific values, according to expectancy-value theory, include self-concept, dispositional interest (or intrinsic value), attainment value, utility value, and cost. Self-concept measures how students see themselves in relation to a specific subject. Dispositional interest measures students' pleasure in learning, while attainment value is about the personal importance students place on mastering a topic. Utility value bears some similarity to achievement goals, as it reflects students' views on how well the learned topics relate to their future goals. Less prominent, but no less important, the cost variable indicates the psychological stress students may face when engaging with a topic. Numerous studies, including Bailey and Phillips (2016); Finney and Schraw (2003); Lane, Hall, and Lane (2004); Macher, Papousek, Ruggeri, and Paechter (2015); Marsh and Martin (2011); McKenzie and Schweitzer (2001); Schunk (1981); and Wigfield and Eccles, (2000), find that expectancy, value, and cost measures have clear relationships with educational outcomes. While expectancy-value theory and achievement goals are indirect measures of student motivation, Plante et al. (2013) show that they complement each other. The explanatory power of these variables increases when both concepts are included. Since students self-selected to take part in the additional practice in our study, measuring multiple dimensions of motivation is of great importance. This is also highlighted by Dunn and Kennedy (2019), who show that intrinsically motivated learners make more effort to work meticulously through e-learning exercises. However, extrinsically motivated learners are more likely to use e-learning more frequently. Thus, including motivation is crucial, as it explains learning outcomes as well as practice behavior.

The Big Five personality traits include five categories (Digman, 1990): (1) Extraversion in individuals is divided into outgoing/energetic vs. solitary/reserved; (2) agreeableness is divided into friendly/compassionate vs. critical/rational; (3) openness, into inventive/curious vs. consistent/cautious; (4) conscientiousness, into efficient/organized vs. extravagant/careless; and (5) neuroticism, into sensitive/nervous vs. resilient/confident. Rimfeld, Kovas, Dale, and Plomin (2016) document the high predictive power of conscientiousness on academic effort and achievement, and Komarraju, Karau, and Schmeck (2009) and Komarraju, Karau, Schmeck, and Avdic (2011) show that the Big Five, combined, explain about 14% of GPA variance among college students.

Present-bias preferences measure individuals' risk aversion, discounting rate, and present-bias behavior (Frederick & Loewenstein, 2002; Horn & Kiss, 2018). Risk preference captures how much individuals value certain outcomes relative to risky or uncertain alternatives. Here, "risky" refers to situations in which the probabilities of various events are known, while "uncertain" means that these probabilities are not known. The discount rate measures (im)patience and compares the likelihood of immediate and future returns. The smaller an individual's discount rate, the more patient that individual is. The size of a person's discount rates (which compare how much an individual values present returns to how much they value future returns) indicates whether that individual has a present bias. That is, if the discount rate is larger in the present than in the future, then the individual is more impatient in the short term than in the long term and suffers from present bias. Bisin and Hyndman (2020) find that present-biased students are more prone to procrastination and spend less time studying during the semester. Horn and Kiss (2018) find that present-bias preferences help explain university exam grades and that risk-averse students outperform risk-tolerant students. Becker et al. (2012) show that time preferences complement personality traits and that both are useful predictors of educational outcomes.

3. Aims of the Study

Most literature on blended learning compares students who are taught in different formats, but few previous studies have explicitly studied the effects of specific features in a blended format. Though the objective of Panus et al. (2014) is similar to that of our study, their analysis focuses on multiple-choice quizzes. Our study aimed to test the effectiveness of two different e-learning practice opportunities on high-level learning in a higher education mathematics course. In addition, we explicitly distinguish the effects of engagement in additional practice from those of performance during e-learning practice on students' exam outcomes. Studies on self-testing and spacing have already shown that additional practice using these techniques should improve students' exam outcomes. The literature, however, focuses on observing or promoting study techniques like self-testing and spacing, while we designed our e-learning practice to explicitly incorporate self-testing and spacing. We also included knowledge of correct response feedback.

To the best of the authors' knowledge, there are no studies demonstrating the influence of additional practice in an e-learning environment on exam points in a higher education setting. To approach a causal interpretation of the practice effect more closely, we have also included a set of control variables to learn whether the effect is solely driven by students' characteristics, such as motivation, conscientiousness, or risk-averseness. Furthermore, we wanted to test whether it is important to measure both engagement (i.e., the number of attempts) and practice performance to obtain a clearer picture of the impact of practice on exam scores. Finally, we used quantile regression to determine whether low- or high-performing students benefited more from practice engagement. Therefore, in this study, we have analyzed an innovative approach to e-learning practice and demonstrated how well this approach supplements face-to-face lecture instruction in higher education. The analysis is guided by the following research questions (RQs):

- RQ1: Do students benefit from additional e-learning practice opportunities? In other words, do both practice engagement (RQ1a) and practice performance (RQ1b) positively impact exam outcomes, or does only performance affect outcomes, suggesting that only high-performing students benefit from practice?

- RQ2: Are the impacts of practice engagement (RQ2a) and practice performance (RQ2b) robust relative to demographic information, prior achievement, self-concept, task values, cost, achievement goals, personality traits, and present-bias preferences, or does the effect vanish once these additional controls are included?

- RQ3: Would low-performing students have benefitted more if they had practiced more with our e-learning opportunities than the high-performing students?

Given the inconsistent results of Brown and Liedholm (2002) and Panus et al. (2014) and the findings on procrastination presented in Bisin and Hyndman (2020), it was unclear at the beginning of this study whether students in the course would engage in and benefit from the additional e-learning practice. However, based on the previously described literature on self-testing, spacing, and feedback, we expected that the newly introduced digital practice opportunities would have a positive effect if students engaged with them. Further, it was unclear whether this effect would remain once different sets of control variables were included; it was also unclear whether low-performing students would benefit from additional practice. Hence, this part of our investigation is exploratory.

4. Materials, Setting, and Methods

4.1 Design of Online Exercises and Participants

Mathematics for Economics and Business Administration is a compulsory module in the first semester of all bachelor’s degree programs in economics and business administration (majors and minors) at the German public university where the study was conducted. The course has three voluntary practice tests, which are offered through the built-in test feature in the university's open-source online learning management system (ILIAS). These practice tests are designed primarily to give students the opportunity to self-test their knowledge throughout the semester. In order to incentivize participation, students could earn up to two extra points based on practice test performance; these points were added to their final exam score, which was worth a total of 90 points. Dobrow, Smith, and Posner (2010) and Garaus, Furtmuller, and Güttel (2016) show that small external rewards can increase intrinsic motivation and autonomous forms of extrinsic motivation.

To provide additional practice opportunities, the practice tests were kept available online. ILIAS automatically changed the values used in the math problems after each attempt. This enabled students to complete the tests multiple times to test themselves. No additional external rewards were linked to the test after the first attempt. The practice tests offered an easily accessible way for students to space learning sequences throughout the semester, preventing cramming right before the exam. Participating in the optional practice tests could also help students keep track of their learning status throughout the semester. Ideally, this would help students keep up with the course material and avoid falling behind. An additional tool that covers content from the first half of the semester is A Matrix a Day (MAD), a web-based linear algebra application. This app allowed students to practice tasks covered during the first half of the semester during the second half. The app encourages students to self-test their knowledge of the content from the first half of the semester throughout the course. Since linear algebra is a new topic to the vast majority of students enrolled in the course, they need to devote additional time to learning it. Using the MAD app was voluntary for the students, but they could register, earn a certain number of points for their solutions, and opt-in to receive a ranking. While no extra exam credits were awarded, the students with the six top scores were rewarded with a 20€ shopping voucher. Students were able to enter the competition at any point during the semester; however, they could not submit solutions for previous days. We also provided all matrices from the previous week on the app's website as additional study material.

The conceptual design of the e-learning exercises that accompanied the course *Mathematics for Economics and Business Administration* is graphically depicted in Figure 1. The figure illustrates possible ways to go through the course to the final exam. While 336 students registered for the course, only 280 attended the final exam. The line with red dashes shows the theoretical path of students who did not use any of the self-testing opportunities provided but took the survey on the covariates. However, no student took the survey and skipped the practice opportunities altogether. The solid line represents students who completed all the practice tests for extra credit (Practice Test I-III with reward in Figure 1) and used the practice tests without extra credit (Practice Test I-III without reward in Figure 1) to prepare for the exam as much as possible. The dotted line at the bottom represents the path for students who participated in the practice tests with rewards but did not complete any additional practice test for which they did not get any additional credits. It was also possible for students to participate partially, skipping one or more practice tests. These possibilities are not shown in Figure 1. The linear algebra app (MAD app) was made available on November 8, 2019, and introduced in class on November 13, 2019. Students were free to use it at their leisure. The shading in Figure 1 indicates that most students looked at the app immediately after it came online; frequency of use declined as the exam approached.

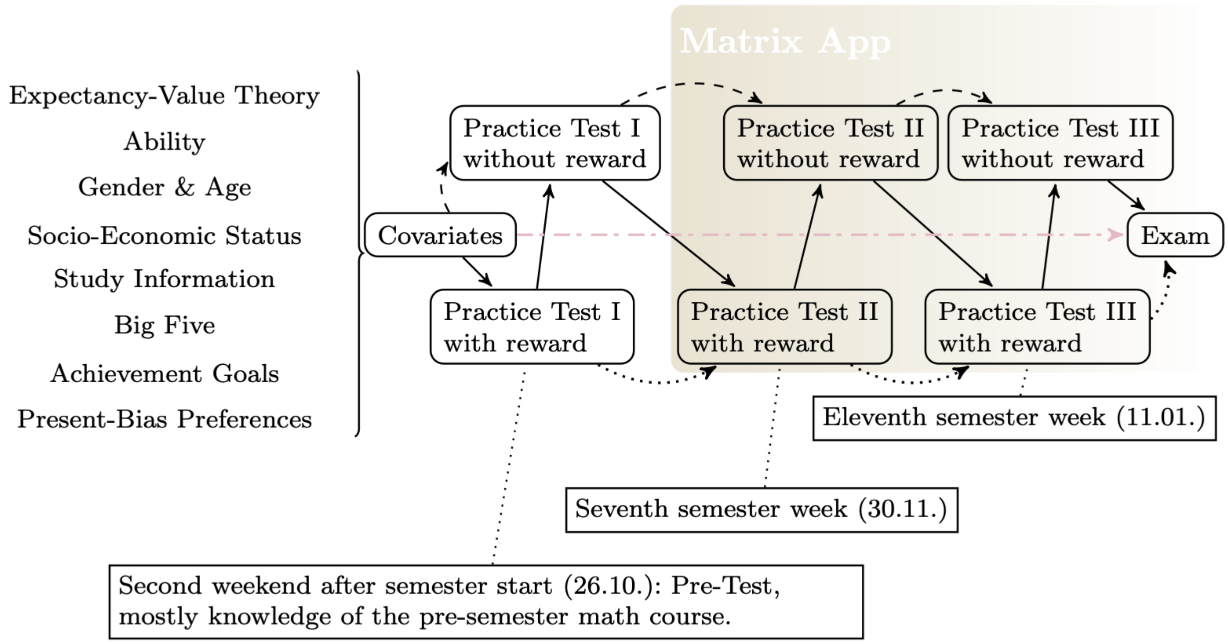


Figure 1. *Design of the practice part of the course* Mathematics for Economics and Business Administration. *The dates refer to the winter semester 2019–20.*

To collect the necessary data to construct the covariates, we surveyed the students at the beginning of the lecture. The study and the survey were announced at the first lecture on October 17, 2019. Students were then given until the following Sunday (October 20) to complete the online questionnaires. We used a raffle of three iPads and 75 shopping vouchers worth 20€ each to encourage a high participation rate. The likelihood of winning at least one of the prices was roughly 25%, based on the number of registrations on ILIAS. Ultimately, 325 students (at least partially) filled out at least one of the questionnaires.

4.2 Sample Description and Measures

Table 1 summarizes the statistics for the final exam, practice tests, additional practice opportunities, the linear algebra app, and demographic information. Table 1 also shows the students' own goals for the semester. As the table shows, 56% of the students were female. Prior achievement in high school was rather good, with an average high school GPA above 2.0 and an average math grade of 2.6 on the German scale, which ranges from 1 (best) to 5 (worst, failing grade). Forty-four percent of the students were seeking an internationally focused degree (BSc in International Business Administration or BSc in International Economics). The largest share of students minoring in business were pursuing a sports management degree. It is noteworthy that the average number of practice tests students planned to complete was less than 3 at the beginning of the semester; the number of envisaged self-testing runs was slightly above 1. Students aimed for a grade of 2 on the final exam, on average.

By the end of the semester, 280 individuals had taken the exam. Of these, we have complete information for 188 students, which we will hereafter refer to as the complete cases sample. At least one variable was missing for the other 92 students. Differences between the two samples are minor and do not qualitatively influence the results. To ensure transparency, we include a discussion of these differences and the robustness checks in Appendix C.

Overall, the exam was difficult, as the average score was less than 45 out of 90 (the passing threshold). The maximum score achieved by any individual student was 82. It should be noted, however, that extra credit points from completing the practice tests were added afterward. Students received an additional three points on average. Of the students who sat for the exam, 33.35% did not pass the exam. Our practice variables indicate that the average student completed three practice tests and submitted around three solutions on the MAD app.

Table 1: *Descriptive statistics for outcomes, variables of interest, and demographic information*

	Full sample			Complete cases sample				
	N	**Mean**	**SD**	**N**	**Mean**	**SD**	**Min**	**Max**
Exam (outcome)								
Final exam score	280	41.19	17.06	188	43.97	17.16	20.00	82.00
Standardized final exam score	280	0.00	1.00	188	0.00	1.00	-2.45	2.22
Practice (participation and performance)								
Practice test attempts	280	3.10	1.00	188	3.23	0.87	1.00	6.00
Practice test performance	280	67.00	19.18	188	71.37	15.82	6.90	100.00
MAD submissions	280	2.74	8.38	188	2.71	8.26	0.00	77.00

MAD percentage	280	24.74	37.05	188	26.82	36.97	0.00	100.00
Individual characteristics								
Female*	280	0.56	0.50	188	0.59	0.49	0.00	1.00
High school GPA	226	2.08	0.60	188	2.07	0.60	1.00	3.70
Advanced math in HS	219	0.83	0.38	188	0.86	0.35	0.00	1.00
Last math grade in HS	226	2.62	1.10	188	2.57	1.08	1.00	5.00
International degree program*	280	0.41	0.49	188	0.44	0.50	0.00	1.00
Sports management degree*	280	0.08	0.26	188	0.05	0.21	0.00	1.00
Minor*	280	0.16	0.37	188	0.12	0.32	0.00	1.00
Working to finance studies*	210	0.22	0.41	188	0.20	0.40	0.00	1.00
Semester of studies	225	1.23	1.10	188	1.24	1.16	1.00	13.00
Re-taking course*	225	2.01	0.21	188	2.01	0.19	0.00	1.00
Students' goals								
Number of practice tests	223	2.81	0.46	188	2.82	0.45	0.00	3.00
Practice tests score	223	0.79	0.13	188	0.79	0.14	0.00	1.00
Practice after practice tests	223	1.24	0.45	188	1.20	0.43	1.00	3.00
Exam grade	223	2.05	0.62	188	2.05	0.62	1.00	4.00

Notes: The table shows the number of observations, means, and standard deviations per variable for three different subsamples. The first of these is the raw sample, which includes all individuals who took the exam. The number of observations varies because some students did not complete the survey or did not answer all survey questions. Next, we looked at the complete case sample. Therefore, we only included individuals for whom all variables were available. Thus, the number of observations in this sample is fixed across all variables. In the appendix, we include a short discussion (and additional analyses) showing that the following regression results are unlikely to be affected by selection bias. We only present the minimum and maximum for the complete cases as they are mostly identical in the full sample.

[*] indicates a dichotomous variable. These are equal to one if the realization is equal to the name of the variable and zero otherwise (e.g., Female = 1 if students are female and zero otherwise).

We constructed the practice variables presented in Table 1 by summing the number of practice tests (with and without an external reward) attempted by the students (practice test attempts). This variable ranged from zero to six. Theoretically, it could go to infinity because students were allowed to repeat the practice tests as often as they wanted (with no extra credit). Only one student retook one of the no-credit practice test twice. A few students repeated each of the three practice opportunities once for a total of six attempts. We converted the points obtained in each practice trial (whether for extra credit or no credit) to derive a relative score in percentages. We then took the mean of this score, considering only the practice attempts that a student actually took (practice test performance). Students who did not complete any practice tests were assigned zero points. For the linear algebra app, we used the number of submitted solutions (MAD submissions). Students could attain up to 11 points for each submitted solution, and we used the median percentage score in the study (MAD percentage). Thus, we obtained four variables: two for participation (practice test attempts and MAD submissions) and two for performance (practice test performance and MAD percentage).

Table 2 summarizes the additional control variables used in the analysis. To account for differences in motivation, we included the following variables from expectancy-value theory (Eccles et al., 1983): self-concept, intrinsic value, attainment value, utility value, and cost. The respective items for each variable were taken from Gaspard, Häfner, Parrisius, Trautwein, and Nagengast (2017), since they are available in German and have been adapted for the university context. We also included the achievement goals mastery–approach, mastery–achievement, performance–approach, and performance–achievement from Elliot and Murayama (2008). We translated these items into German and adjusted them to fit the course context. To measure students' personality, we assessed the Big Five personality traits using the short questionnaire by Schupp and Gerlitz (2014). For all items measuring motivation and personality, students had to indicate to what extent they agreed or disagreed with a statement using a scale ranging from 1 to 7. We calculated Cronbach's α in R using the *psych* package (Table 2) (Revelle, 2020). We also measured present-bias preferences following Frederick and Loewenstein (2002). Here, students had to answer questions about their preferences; we used their answers to calculate their risk-taking propensities, discount rates, and present bias.

Table 2: *Descriptive statistics for additional control variables*

	Full sample			Complete cases sample			
	N	**Mean**	**SD**	**N**	**Mean**	**SD**	**Cron. α**
Expectancy-value theory							
Self-concept	227	2.70	0.62	188	2.74	0.61	0.86
Intrinsic value/dispositional interest	226	2.74	0.61	188	2.76	0.60	0.87
Attainment value	225	2.02	0.53	188	2.00	0.54	0.71
Utility value	224	3.51	0.52	188	3.51	0.52	0.88

Cost	225	2.38	0.55	188	2.38	0.55	0.75
Big Five							
Conscientiousness	225	4.94	1.11	188	4.97	1.03	0.65
Extraversion	225	4.94	1.30	188	5.02	1.24	0.82
Agreeableness	225	5.49	1.09	188	5.52	0.97	0.62
Openness	223	4.87	1.15	188	4.84	1.18	0.65
Neuroticism	225	4.39	1.22	188	4.42	1.16	0.68
Achievement goals							
Mastery–approach	221	6.15	0.71	188	5.66	0.98	0.64
Mastery–avoidance	221	5.64	0.99	188	4.97	1.52	0.71
Performance–approach	218	4.97	1.51	188	4.95	1.65	0.87
Performance–avoidance	218	4.96	1.62	188	2.82	0.45	0.92
Present bias preferences							
Risk	222	0.68	0.20	188	0.69	0.20	
Discount factor	217	0.98	0.68	188	0.95	0.55	
Present bias	216	1.06	0.28	188	1.05	0.18	

Notes: The table shows the number of observations, means, and standard deviations per variable for two samples. The full sample includes incomplete cases, while the complete cases sample only consists of observations for students who fully completed all questions on both questionnaires.

4.3 Methods

4.3.1 Linear Model Setup

To analyze the effect of practice on exam outcomes, we used the following linear model:

$$\begin{aligned} points_i = \mu + \boldsymbol{\rho}'\mathbf{p}_i + \boldsymbol{\beta}'_1\,\mathbf{char}_i + \boldsymbol{\beta}'_2\,\mathbf{EVT}_i + \boldsymbol{\beta}'_4\,\mathbf{agoals}_i \\ + \boldsymbol{\beta}'_4\,\mathbf{bigfive}_i + \boldsymbol{\beta}'_5\,\mathbf{pbp}_i + \boldsymbol{\beta}'_6\,\mathbf{sgoals}_i + \eta_i, \end{aligned} \tag{1}$$

where $\boldsymbol{p}$ denotes a (4×1) vector that contains the four practice variables *practice test attempts, practice test performance*, *MAD submissions*, and *MAD percentage.* The index i stands for individuals, and η_i is an idiosyncratic error term. $\boldsymbol{\rho}$ and $\boldsymbol{\beta'}_1$ through $\boldsymbol{\beta'}_6$ are vectors of parameters; the length for each is determined by the number of factors included in each category. Students' characteristics, which are listed in Table 1, are subsumed in the vector $\boldsymbol{char_i}$. Measures from expectancy-value theory, which are listed in Table 2, are represented with the vector $\boldsymbol{EVT_i}$. Personality traits captured in the survey are grouped into five factors and included in the vector $\boldsymbol{bigfive_i}$. The measures identifying present bias preferences are contained in $\boldsymbol{pbp_i}$. Achievement goals are captured in $\boldsymbol{agoals_i}$, while $\boldsymbol{sgoals_i}$ represents students' own goals at the beginning of the semester. In the first step, we considered a baseline model that only contains the regressors in $\boldsymbol{p}$.

The practice variables in this baseline model could be endogenously determined. Nonetheless, this model is interesting as it can be used to analyze the effect of engagement with the additional self-testing opportunities. In the second step, we aimed to isolate the potential effect of the practice variables on the final exam score by adding the set of remaining control variables shown in Equation (1). In essence, all factors included in our control variables may affect both the exam outcome and students' testing engagement and performance. As the control variables were (temporally) predetermined, students' behavior could not be affected by them.

We estimated Equation (1) using OLS and calculated heteroskedasticity-robust standard errors. However, given that we have only 188 students for which all variables are available, we have too few observations for the rich set of control variables for valid inference. To address this issue, we decided to use machine-learning techniques to select variables. We employed three different techniques to determine and select essential features for predicting an outcome variable: the least absolute shrinkage and selection operator (LASSO), random forest, and xgBoost (an algorithm for extreme gradient boosting). These techniques can also rank variables by importance. Each method identified a set of control variables deemed most important by that method. A short description can be found in the online appendix.

After selecting variables via these three machine-learning methods, we followed the recommendation of Belloni et al. (2013) and ran post-selection OLS regressions with the selected variables. This method selects predictors for the respective outcome (in our case, the exam score) and the main explanatory variables (the four practice variables). Usually, a feature selection algorithm would only be run for variables that predict the exam score. However, having identified the variables with the highest predictive power does not necessarily allow consistent estimation of the marginal effect of the variable of interest. Thus, according to Belloni et al.

(2013), in order to achieve an unbiased estimation of an independent variable of interest, that independent variable (here, the practice variables) should be included in the algorithm. It is not sufficient to only include variables that are important predictors of the outcome (here, the exam score). Therefore, we also ran a feature selection algorithm for control variables that predicted the practice variables. We then used the set of variables that significantly predicted the exam score and the four practice variables in a post-OLS regression, that is, a normal OLS regression, after the variables were selected. This enabled us to identify the strongest predictive variables in the model and the variables that could reduce possible bias related to our explanatory variables. We use the term “double selection” to refer to this two-step or (in our case) five-step selection process.

While, in principle, the practice variables could correlate with confounding variables, we included measures to reduce the impact of the most important confounders identified in previous literature on predicting student achievement. Clearly, we could not completely eliminate the possibility that a significant variable might be omitted. The same holds for possible selection bias. However, given the rich set of controls and the findings of previous literature, and assuming the approximate representativeness of the sample, we are confident that our study and model setup can reasonably approximate a causal treatment effect of practice on students’ final exam performance. Furthermore, comparing the estimation results of the basic model with those of the full model offers insights into the amount of bias when the additional control variables are not included in the model.

4.3.2 A Quantile Model for Exam Performance

The model discussed in Equation (1) is designed to identify the average effect of practice on exam outcomes for all students. In this subsection, we go one step further and investigate how the marginal effects differ for students who performed better (or worse) on the exam. Therefore, we estimated a quantile-specific version of Equation (1) using the quantile regression method proposed by Koenker and Bassett (1978) to analyze how the effect of practice varied across performance groups. The post-selection quantile regression allowed us to investigate whether students at the bottom of the distribution benefitted more from additional practice than those at the top. Hence, quantile regression provides additional insights into the heterogeneity of the practice effects and indicates who benefited most from practice.

Following Koenker and Bassett (1978), the conditional quantile model for exam points is

$$Q_{\mathrm{Points}_i}(\tau|\mathbf{p}_i,\mathbf{x}_i)=\boldsymbol{\alpha}_\tau+\boldsymbol{\rho}'_\tau\mathbf{p}_i+\boldsymbol{\beta}'_\tau\mathbf{x}_i \tag{2}$$

for quantiles $\tau \in (0,1)$ in intervals of 0.05. In contrast to Equation (1), quantile regression specifies a linear model for every quantile of the dependent variable. Hence, the parameter estimates $\boldsymbol{\rho}$ and $\boldsymbol{\beta}$ may vary across τs. To simplify the notation, $\mathbf{x}_i$ contains all control variables. We used the R package *quantreg* developed by Koenker (2019) to obtain the estimates.

5. Results

5.1 Regression Results

Table 3 presents the regression results for the baseline and full model presented in Equation (1). The full model includes the covariates identified via double selection OLS using the LASSO, random forest, xgBoost, or all available covariates. The outcome variable is the standardized final exam score without the extra credit points students received for completing the practice tests. While the complete model might have too many control variables to allow sufficient degrees of freedom for making statistical inferences, we include it here to check whether the estimates of the respective practice variables would be entirely different (e.g., carry the opposite sign), which is not the case.

Table 3: *Main regression results*

	Dependent variable: Standardized points on final exam					
	Practice variables only	LASSO	Random forest	xgBoost	All covariates	No performance variables
Practice test attempts	0.226***	0.215***	0.203***	0.205***	0.236***	0.273***
	(0.074)	(0.068)	(0.065)	(0.064)	(0.068)	(0.084)
Practice test performance	0.022***	0.010**	0.010**	0.010**	0.009*	
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	
MAD submissions	0.007	0.006	0.001	0.004	0.001	0.020*
	(0.012)	(0.008)	(0.009)	(0.009)	(0.009)	(0.011)
MAD percentage	0.004**	0.004**	0.005***	0.004**	0.004**	
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
Constant	-2.401***	-0.824	-0.078	-0.195	-1.335	-0.937***
	(0.328)	(0.847)	(1.256)	(0.753)	(1.845)	(0.285)
Additional controls	No	Yes†	Yes‡	Yes§	All	No
Observations	188	188	188	188	188	188
Adjusted R^2	0.213	0.446	0.464	0.410	0.453	0.085

Note: Heteroskedasticity-robust standard errors are in parentheses. Columns 3 to 5 show the post-selection results for each algorithm (LASSO, random forest, and xgBoost). For all three columns, we used a double selection process to select variables that are important for (i) the outcome and (ii) the practice variables. † Includes high school GPA, advanced math in HS, last math grade in HS, international degree program, sports management program, minor, work to finance studies, semester of study, retaking course, self-concept, planned number of practice tests. ‡ Includes high school GPA, advanced math in HS, last math grade in HS, international degree program, minor, semester of study, retaking course, self-concept, attainment value, performance–approach, performance–avoidance, risk, present bias. § Includes female, high school GPA, international degree program, mastery–avoidance, performance–approach, agreeableness, openness, neuroticism, risk, discount factor, present bias.
*p<0.1; **p<0.05; ***p<0.01

The first estimation results shown in Table 3 (practice variables only) answer RQ1: When no control variables are included, the coefficient for practice test attempts is 0.226, which is significant at the 1% level. This estimated coefficient is only reduced slightly when different sets of control variables are included. Variable selection via LASSO reduces the coefficient to 0.215, selection via random forest reduces it to 0.203, and selection via xgBoost reduces it to 0.205. Including all control variables increases the coefficient to 0.236. For all specifications, the coefficient is significant at the 1% level. Thus, practice test attempts have a robust positive impact, even when including a rich set of control variables.

The coefficient for practice test performance is 0.022 in the second column (practice variables only), which is significant at the 1% level. This coefficient is more than halved when additional control variables are included, although it remains significant at the 5% level. Failing to include the control variables would lead to an overestimation of the coefficient for practice performance.

The coefficient for the number of MAD submissions is small and nonsignificant. Only the MAD percentage of correct answers on the MAD app is significant at the 5% or 1% level, with a robust coefficient of 0.004 or 0.005.

The coefficients in the second column of Table 3, where only the four practice variables are included, imply that a student who—like the majority of participants—completed the three practice tests for extra credit is expected to attain an additional 13 points out of 90 on the final exam compared to a student who did not complete any practice tests. These points do not include the extra credit points awarded for successful completion of the practice tests. In addition, a student who submitted her results to the MAD app on ten different days is predicted to attain roughly three extra points on the final exam. Note that students who engaged with the MAD app submitted an average of approximately eight solutions.

As measured by the adjusted R^2, the sparse models have approximately the same explanatory power as the full model above. Among the linear models considered, the LASSO selects the sparsest and best-fitting model. Thus, employing machine-learning techniques, we have shown that the complete model in Equation (1) can be equivalently represented with the sparse variable sets presented in Table 3; this approach maintains the effect size of the practice variables. Hence, the additional results in Table 3 answer RQ2: The effect of practice test engagement was consistently significant even after additional control variables were included (RQ2a). While the performance effect (RQ2b) was reduced when control variables were included, this coefficient is still highly significant. Thus, this performance coefficient had an upward bias due to the omitted variables.

The significant effects of the practice variables indicate that practice has a direct positive effect on exam results. Ability, motivation, and other individual characteristics have an indirect effect on exam results; this effect is moderated by practice. The existence of this indirect effect can be deduced from the changes to the coefficient for practice test performance when the control variables are included. This will be discussed in more detail in Section 6.

The last column of Table 3 shows the regression results excluding the two performance variables. This shows the direction of the bias if the performance measurement is omitted. The coefficient for practice test attempts increases from 0.226 to 0.273, and the coefficient for the number of MAD submissions increases from 0.007 to 0.020 (which is significant at the 10% level). Thus, not differentiating attempts from performance on additional practice opportunities would result in an upward bias, especially for MAD.

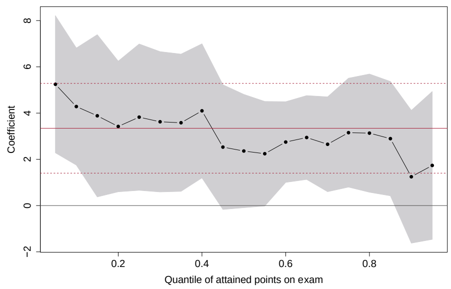
5.2 Quantile Regressions Results

To explore who benefits most from practice, we present the results of the quantile regression model specified in Equation (2). The covariates in $\mathbf{x}_i$ are specified as either the union or the intersection of all relevant covariates identified in Section 5.1. The union set includes all variables selected by at least one machine-learning algorithm for at least one outcome or explanatory variable (specification a). The intersection set includes only the variables selected by all algorithms (specification b). The results, including only the practice variables of interest, are displayed graphically in Figure 3. The confidence intervals of the quantile regression coefficients are based on standard errors that were calculated using a wild bootstrap procedure with 5,000 replications. It should be stressed that quantile regression models the outcome distribution given a certain set of characteristics, that is, the conditional distribution of outcomes. Therefore, at least in hindsight, quantile regression indicates whether specific characteristics decisively shift the outcome range. On the other hand, having a certain set of characteristics does not automatically mean that an individual will land in a particular performance quantile; the characteristics only indicate that a certain performance range is more likely.

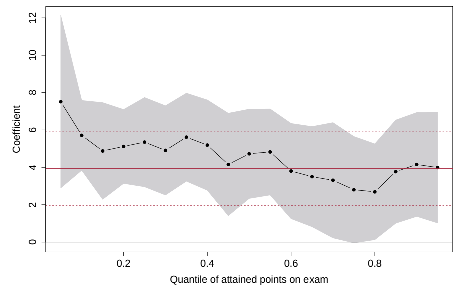
Figure 3i presents the coefficient estimates for the number of practice test attempts. In specifications (a) and (b), the parameters are statistically significant, with some exceptions for quantiles above 0.9, shown in (a). Figure 3ia shows that students who were expected to achieve a low number of points on the exam, that is, those in the lower quantiles, benefit more from practice than students who were expected to score well on the exam. Additional practice can add six points to the final exam score for students in the lowest quantile, while this effect drops to roughly two points for students in the highest quantile. When all control variables are included (see Figure 3ib), this pattern is less pronounced. The lower quantile estimates are still as high as before, but the upper quantile estimates are now around 4 as well. This indicates that practicing is helpful in general and that low-performing students seem to benefit from it even more than high-performing students.

The effect of practice test performance is less clear. Figure 3iia shows that, when the set of variables selected by all algorithms is included, there is a statistically significant relationship in the very low quantiles (below 20%). This suggests that good performance on the practice tests predicts improved exam performance among students who are expected to achieve a low number of points. However, this effect vanishes in both (a) and (b) just below the 40% quantile. Above the median, up to the 80% quantile, good performance on the practice tests also predicts a higher number of points on the final exam. There is no significant relationship among very strong students (upper quantiles). Hence, these students are expected to achieve high scores on the final exam regardless of their practice test performance, which seems plausible.

How often a student used the MAD app does not seem to affect the specifications presented in Figures 3iiia and 3iiib, emphasizing the findings concerning the OLS regression coefficient in Section 5.1 above. This holds across all quantiles. Nevertheless, good performance on the MAD app seems to be a good predictor of exam performance, particularly among students who are expected to achieve a low number of points on the exam. This holds for both specifications presented in Figures 3iva and 3ivb. More correct submissions were especially beneficial for the lowest 5% of students. Students in the middle quartiles also benefited from successful engagement with the MAD app. A coefficient of 0.08 implies that a student who achieved an average of 80% of the available points on their MAD submissions is expected to score 6.4 points higher on the exam than a student whose submissions earned 0% of the available points.

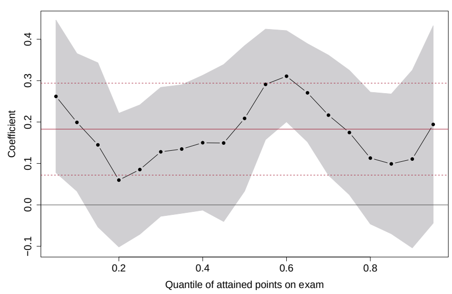


(a) Intersection

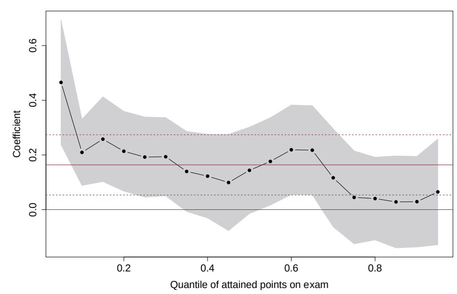


(b) Union

(i) Practice Test Attempts

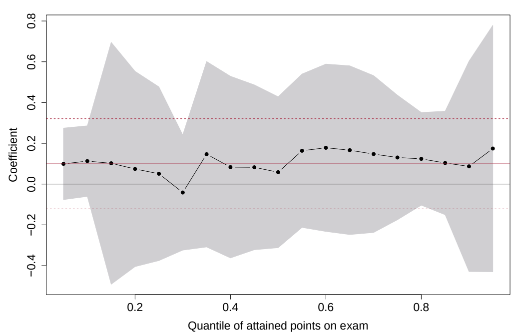


(a) Intersection

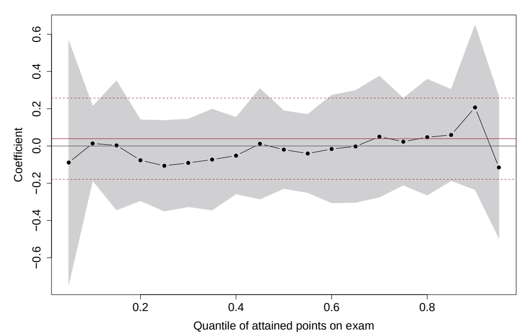


(b) Union

(ii) Practice Test Performance

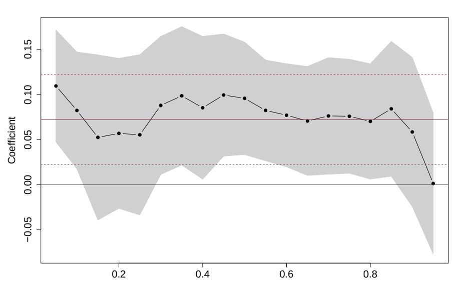


(a) Intersection

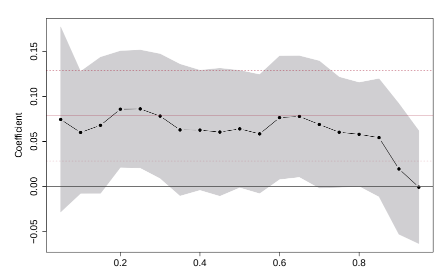


(b) Union

(iii) MAD submissions



(a) Intersection



(b) Union

(iv) MAD percentage

Figure 3. *Quantile Regression Results.* The panels show the coefficient estimates mentioned in the headings from (i) to (iv) across quantiles. The quantile-specific estimates are obtained from a quantile regression on the exam

points. Figure 3a shows the coefficients across quantiles when the intersection of all control variables selected by the machine-learning techniques are included. Figure 3b shows the coefficients across quantiles when the union of all selected control variables are included. The solid red horizontal line shows the value obtained in an equivalent OLS regression, while the dotted red lines indicate the 90% confidence bounds of the OLS estimate. The shaded areas identify the 90% confidence bounds of the quantile regression estimates. The standard errors of the quantile regression coefficients have been calculated using a wild bootstrap procedure with 5,000 replications.

6. Discussion

In this section, we discuss the mechanisms and channels through which students might have benefited from the practice options. We also explore the effect of including control variables when measuring the effect of practice. In addition, we discuss potential reasons for the differences between submissions on the MAD app and practice exam attempts. Finally, we comment on the results of the quantile regressions.

6.1 How Practice Opportunities Improve Performance

Our results show that completing one of the e-learning practice tests improved final exam performance by 2.5 to 5 points out of 90. Why did the students benefit from this additional practice? The primary channel through which students profited cannot be directly assigned to a single feature. Since we developed our practice opportunities based on the rich body of literature on self-testing, spacing, and feedback, the effects of our practice variables in our study may capture the effects of several channels. First, both the practice tests and the MAD were released with enough time between releases and before the final exam. This should have encouraged students to start studying at the beginning of the semester, enabling them to benefit from spaced learning (see, e.g., Ericson, 2017; Rodriguez et al., 2021). This also connects to the literature on procrastination in higher education: students are known to procrastinate (Denny et al., 2018), and reduced procrastination increases student learning (Bisin & Hyndman, 2020; Horn & Kiss, 2018). Since the external rewards for the first trial of each practice exam could only be earned on fixed dates, students who wanted to obtain these extra credits were “forced” to study early in the semester. This automatically led to spaced learning and reduced procrastination. The spaced practice tests might also have encouraged students to set their own goals, which, again, could reduce procrastination (D. Clark et al., 2019; Ericson, 2017).

Since the practice opportunities were, in essence, self-tests, our findings are also supported by Rodriguez et al. (2021), who document a positive effect of self-testing. Students were able to observe their current level of knowledge and assess whether they had mastered the topics. Transferring the results of Attali and van der Kleij (2017), we infer that knowledge of correct response feedback should have further strengthened self-observation, thus increasing learning. Furthermore, providing solutions on the practice tests made it easier for students to find their own errors, as demonstrated by Attali (2015). For the practice exams, we find that the number of attempts has a positive effect when we control for mean performance. Thus, it seems plausible that students also benefited from the *potentiation effect* (Izawa, 1971), the *error generation effect* (Kornell et al., 2009), and/or the *hyper-correction effect* (Butler, Karpicke, & Roediger, 2008). These results illustrate that laboratory results on self-testing, spacing, and feedback can be used to design practice opportunities using an e-learning environment in a higher education course without specifically promoting these learning methods. These results are noteworthy because they can help higher education institutions implement similar designs in math-related courses, and perhaps other courses, to promote student learning.

6.2 Why Additional Controls and Performance Should be Included in the Analysis

Our results emphasize the need to measure both self-testing attempts and respective performance. The number of attempts is significant for the practice tests, even when students’ performance during these self-tests is included in the analysis. As our study does not fit the classical randomized controlled trial framework, one might argue that an ability bias remains even after we have controlled for prior achievement. However, since attempts and performance are both measured by the e-learning tests, this ability bias should be captured by the performance features. More able students score higher on the practice tests and on the final exam. Consistent with this idea, the effect of practice test attempts increases when we control for performance. This effect, which suggests that students learned from practice attempts regardless of performance (or especially when performance was poor), aligns with Clark and Bjork (2014), who state that "students can be learning even when a current test shows little or no progress" (Clark & Bjork, 2014, p. 21). This consistent with the effects of *potentiation, error generation,* and *hyper-correction* mentioned above. Self-testing—even if the self-test scores are poor—helps students learn.

Given the extensive set of control variables, our estimation results have more confidence and may approach the causal effect of practice more closely than those of Panus et al. (2014), who look only at correlations between self-testing attempts and exam outcomes. The control variables allow us to avoid a pure selection effect, which could occur if only *better*, more motivated, more conscious, or more risk-averse students took the practice tests and used the MAD app. Furthermore, Panus et al. (2014) studied students in a pathophysiology course; participants could attempt each quiz 17 to 30 times, which supports low-level learning. Our results show that even a few attempts help students achieve high-level learning and improve their grades in a course.

Future similar research could change some aspects of our study. The external reward could be changed; shorter weekly self-tests could be offered instead of three longer, less frequent self-tests; feedback could consist only of knowledge of the correct response; or more elaborate feedback could be offered. It might also be useful to compare the effects of explicitly encouraging self-testing and spacing in a high-level learning course, as in Rodriguez et al. (2021), to those of inducing self-testing and spacing as we did with the practice tests. It would be interesting to see which method is more beneficial in terms of increasing intrinsic motivation and student learning outcomes and to explore whether explicitly encouraging self-testing and spacing would increase the number of attempts in our design.

6.3 Effect Difference Between Practice Tests and MAD

Our results show that the MAD app was less effective than the practice tests. This could be due to the differences in the respective settings of these practice opportunities. Since external rewards were offered for the practice tests and since the first attempt counted only when it was completed on the release date, students might have perceived the practice tests as a priority, while MAD, which was introduced as a daily routine, was seen as an additional feature. As mentioned earlier, Ericson (2017) shows that practice tests create intermediate deadlines and reduce the risk that students will procrastinate until the week before the exam. Since the MAD app offered daily practice opportunities, this might not have reduced procrastination behavior. In addition, students might take the term "practice test" more seriously than an application called "A Matrix A Day." Students have busy weekly schedules, and MAD might not have been a priority. Finally, Garaus et al. (2016) mention that small rewards can make up for a lack of face-to-face feedback. Because the course is too large for students to receive personal feedback, the automatic feedback may have supported students and (at least partially) replaced personal feedback from the instructor. This is important because Karlen, Suter, Hirt, and Maag Merki (2019) show that positive motivational patterns must be supported and encouraged for students to achieve academic improvement. Thus, our external rewards might have helped motivate students to participate in the practice tests and thereby weakened the appeal of MAD. Even the MAD rankings and the vouchers offered to top students did not lead to daily participation. Although Denny et al. (2018) show that gamification elements such as a points system or a badge system can increase motivation, our settings were insufficient to motivate daily use.

6.4 Can Low-Performing Students Benefit from Practice?

Our quantile regression results indicate that students could improve their performance on the final exam by using our additional practice options, regardless of other individual characteristics. The additional practice helped, not only high-achieving students, but also—or especially—students at risk of failing the exam. This is also in line with the significant effect of practice test attempts, which was independent of performance on these self-tests: self-testing during the semester enhanced retention of the learning material for the final exam, regardless of performance. This is also in line with research conducted by Roediger III and Karpicke (2006) and Sloboda, Davidson, Howe, and Moore (1996). Even though Joyce et al. (2015) show that low-performing students are more likely to suffer in a blended learning format, we have demonstrated that these students benefitted from the added e-learning features. Thus, it is not inevitable that low-performing students suffer when a face-to-face format changes to a blended one. The key component of the blended learning in Joyce et al.'s (2015) study was that students had only one face-to-face class instead of two per week. The material not covered in in the face-to-face class was presented in videos. They also provided additional practice material and quizzes before and after the lectures, but these practice materials were compulsory for both groups. Thereby, our blended format differs from that used by Joyce et al. (2015): All students could attend all in-person lectures, and they also had the option to engage in additional practice in an e-learning environment. While the additional video material in Joyce et al.'s (2015) study might have increased procrastination, our design should have reduced it, which seems to be especially important for low-performing students.

6.5 Limitations of the Study

This study has some limitations. First, participants included only students enrolled in the math course. These students had self-selected into an economics or business administration bachelor's degree program for which the studied course is mandatory. Furthermore, there was some minor self-selection concerning students dropping out of the course during the semester. We could not control for this (and it happens every year in real educational settings). However, this did not affect our results, as shown in Appendix C. Second, the effect of online practice may differ in other subjects or if the design of the practice material includes other features (such as no external rewards, no feedback, or quizzes). More research is needed to see if the estimated effects can be generalized to other subjects and to identify the most effective features of our approach. Third, omitted variable bias cannot be ruled out completely. We know whether or not participants had to work to finance their studies, but we collected no other information about students' social environments. Students might have different levels of distractions from their studies due to factors such as friends and family (background); we could not control for these variables. In addition, it is theoretically impossible to preempt and assess all potentially omitted variables in a survey. A randomized controlled trial was not an option for this study due to ethical concerns. Nonetheless, a randomized controlled trial with varying amounts of additional practice opportunities would help assess whether the results obtained here, namely, that more practice improves students' learning outcomes, can be generalized.

7. Conclusion

We evaluated two types of voluntary e-learning opportunities offered to students taking a mathematics lecture course designed for first-year business administration and economics students. The first practice option consisted of three online practice tests with repeated access and extra credit for the first attempt on a specific date; the second consisted of the “Matrix a Day” app. We find that students who took the optional practice tests in the e-learning environment generally improved their exam performance (RQ1a), regardless of their performance on these practice tests. This result is robust even when controlled for demographic characteristics, prior achievement, self-concept, value tasks, cost, achievement goals, personality traits, present bias preferences, and personal course goals (as stated at the beginning of the semester) (RQ2a). Performance on practice test attempts added to the positive effect of the attempts (RQ1b), but the inclusion of additional control variables halved the coefficient (RQ2b). To address concerns about overfitting, we employed variable selection techniques from the machine-learning domain, namely LASSO, random forest, and xgBoost. None of the different sets of selected variables qualitatively changed the estimated effects of practice test attempts or of performance on the practice tests. We find that the number of solutions submitted (RQ1a and RQ2a) on the MAD app did have an effect, but no additional performance effect (RQ1b and RQ2b). In addition, the quantile regressions show that particularly low-performing students benefited the most from additional practice (RQ3). This indicates that students’ final exam grades were not predetermined by past performance or personal characteristics but could be altered by their learning behavior. Hence, higher education institutions should find ways to encourage low-performing students to engage with optional practice possibilities, even if this approach contradicts the conventional wisdom of some instructors, because no students should be left behind.

References

Alpert, W. T., Couch, K. A., Harmon, O. R., & Gpa, P. (2016). A Randomized Assessment of Online Learning. *American Economic Review: Papers & Proceedings*, *106*(5), 378–382.

Ardac, D., & Sezen, A. H. (2002). Effectiveness of computer-based chemistry instruction in enhancing the learning of content and variable control under guided versus unguided conditions. *Journal of Science Education and Technology*, *11*(1), 39–48. https://doi.org/10.1023/A:1013995314094

Attali, Y. (2015). Effects of multiple-try feedback and question type during mathematics problem solving on performance in similar problems. *Computers and Education*, *86*, 260–267. https://doi.org/10.1016/j.compedu.2015.08.011

Attali, Y., & van der Kleij, F. (2017). Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers and Education*, *110*, 154–169. https://doi.org/10.1016/j.compedu.2017.03.012

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist*, *40*(4), 199–209. https://doi.org/10.4324/9781315866239-2

Bailey, T. H., & Phillips, L. J. (2016). The influence of motivation and adaptation on students' subjective well-being, meaning in life and academic performance. *Higher Education Research and Development*, *35*(2), 201–216.

Baker, R., Xu, D., Park, J., Yu, R., Li, Q., Cung, B., Fischer, C., Rodriguez, F., Warschauer, M., & Smyth, P. (2020). The benefits and caveats of using clickstream data to understand student self-regulatory behaviors: opening the black box of learning processes. *International Journal of Educational Technology in Higher Education*, *17*(1), 1–24. https://doi.org/10.1186/s41239-020-00187-1

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research*, *61*(2), 213–238. https://doi.org/10.3102/00346543061002213

Baranik, L. E., Stanley, L. J., Bynum, B. H., & Lance, C. E. (2010). Examining the construct validity of mastery-avoidance achievement goals: A meta-analysis. *Human Performance*, *23*(3), 265–282. https://doi.org/10.1080/08959285.2010.488463

Becker, A., Deckers, T., Dohmen, T., Falk, A., & Kosse, F. (2012). The Relationship Between Economic Preferences and Psychological Personality Measures. *Annual Review of Economics*, *4*(1), 453–478. https://doi.org/10.1146/annurev-economics-080511-110922

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, *81*(2), 608–650. https://doi.org/10.1093/restud/rdt044

Bisin, A., & Hyndman, K. (2020). Present-bias, procrastination and deadlines in a field experiment. *Games and Economic Behavior*, *119*, 339–357. https://doi.org/10.1016/j.geb.2019.11.010

Bjork, R. A. (1975). Retrieval as a memory modifier: An Interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144).

Brown, B. W., & Liedholm, C. E. (2002). Can Web Courses Replace the Classroom in Principles of Microeconomics? *American Economic Review*, *92*(2), 444–448.

Brown, M. G. (2016). Blended instructional practice: A review of the empirical literature on instructors' adoption and use of online tools in face-to-face teaching. *Internet and Higher Education*, *31*, 1–10.

Butler, A. C. (2010). Repeated Testing Produces Superior Transfer of Learning Relative to Repeated Studying. *Journal of Experimental Psychology*, *36*(5), 1118–1133.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a Metacognitive Error: Feedback Increases Retention of Low-Confidence Correct Responses. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a Metacognitive Error: Feedback Increases Retention of Low-Confidence Correct Responses. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Clark, C. M., & Bjork, R. A. (2014). When and why introducing difficulties and errors can enhance instruction. In A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum* (pp. 20–30). Society for the Teaching of Psychology. https://doi.org/10.4324/9780203817421

Clark, D., Gill, D., Prowse, V., & Rush, M. (2019). Using Goals to Motivate College Students: Theory and Evidence from Field Experiments. *Review of Economics and Statistics*. http://conference.iza.org/conference_files/Cognitive_2017/prowse_v2102.pdf

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*(3), 215–235. https://doi.org/https://doi.org/10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1

Denny, P., McDonald, F., Empson, R., Kelly, P., & Petersen, A. (2018). Empirical Support for a Causal Relationship Between Gamification and Learning Outcomes. *CHI*, *311*, 1–13. https://doi.org/10.1007/978-3-030-37386-3_30

Deschacht, N., & Goeman, K. (2015). The effect of blended learning on course persistence and performance of adult learners: A difference-in-differences analysis. *Computers and Education*, *87*(2015), 83–89. https://doi.org/10.1016/j.compedu.2015.03.020

Digman, J. M. (1990). Personality Structure: Emergence if the Five-Factor Model. *Annual Review of Psychology*, *41*, 417–440.

Dobrow, S. R., Smith, W. K., & Posner, M. A. (2010). Managing the grading paradox: Leveraging the power of choice in the classroom. *Academy of Management 2010 Annual Meeting - Dare to Care: Passion and Compassion in Management Practice and Research, AOM 2010*, *10*(2), 261–276. https://doi.org/10.5465/ambpp.2010.54492060

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, Supplement*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Dunn, T. J., & Kennedy, M. (2019). Technology Enhanced Learning in higher education; motivations, engagement and academic achievement. *Computers and Education*,

137(March), 104–113. https://doi.org/10.1016/j.compedu.2019.04.004
Dusi, D., & Huisman, J. (2021). It's more complex than it seems! Employing the concept of prosumption to grasp the heterogeneity and complexity of student roles in higher education. *Higher Education*, *81*(5), 935–948. https://doi.org/10.1007/s10734-020-00588-1
Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie (Memory: A Contribution to Experimental Psychology)*. Duncker & Humboldt.
Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, Values, and Academic Behaviors. In J. T. Spence (Ed.), *Achievement and Achievement Motivation* (pp. 75–146). Freeman, W. H.
Elliot, A. J., & McGregor, H. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*, 501–519. https://doi.org/10.1037//0022-3514.80.3.501
Elliot, A. J., McGregor, H. A., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology*, *91*(3), 549–563. https://doi.org/10.1037/0022-0663.91.3.549
Elliot, A. J., & Murayama, K. (2008). On the Measurement of Achievement Goals: Critique, Illustration, and Application. *Journal of Educational Psychology*, *100*(3), 613–628. https://doi.org/10.1037/0022-0663.100.3.613
Ericson, K. M. (2017). On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation. *European Economic Association*, *15*(3), 692–719. https://doi.org/10.1093/jeea/jvw015
Feather, N. T. (1992). Values, Valences, Expectations, and Actions. *Journal of Social Issues*, *48*(2), 109–124. https://doi.org/https://doi.org/10.1111/j.1540-4560.1992.tb00887.x
Finn, B., Thomas, R., & Rawson, K. A. (2018). Learning more from feedback: Elaborating feedback with examples enhances concept learning. *Learning and Instruction*, *54*, 104–113. https://doi.org/https://doi.org/10.1016/j.learninstruc.2017.08.007
Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, *28*(2), 161–186.
Frederick, S., & Loewenstein, G. (2002). Time Discounting and Time Preference : A Critical Review. *Journal of Economic Literature*, *XL*, 351–401.
Garaus, C., Furtmuller, G., & Güttel, W. H. (2016). The hidden power of small rewards: The effects of insufficient external rewards on autonomous motivation to learn. *Academy of Management Learning and Education*, *15*(1), 45–59. https://doi.org/10.5465/amle.2012.0284
Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology*, *48*, 67–84. https://doi.org/10.1016/j.cedpsych.2016.09.003
Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34. https://doi.org/doi.org/10.1037//0003-066x.48.1.26
Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology*, *94*(3), 562–575. https://doi.org/10.1037/0022-0663.94.3.562
Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin and Review*, *19*(1), 126–134.

https://doi.org/10.3758/s13423-011-0181-y

Hattie, J., & Gan, M. (2011). Instruction Based on Feedback. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of Research on Learning and Instruction*. Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Horn, D., & Kiss, H. J. (2018). Which preferences associate with school performance?—Lessons from an exploratory study with university students. In *PLoS ONE* (Vol. 13, Issue 2). https://doi.org/10.1371/journal.pone.0190163

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A Meta-Analytic Review of Achievement Goal Measures: Different Labels for the Same Constructs or Different Constructs With Similar Labels? *Psychological Bulletin*, *136*(3), 422–449. https://doi.org/10.1037/a0018947

Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology*, *8*(2), 200–224. https://doi.org/https://doi.org/10.1016/0022-2496(71)90012-5

Jaehnig, W., & Miller, M. L. (2007). Feedback types in programmed instruction: A systematic review. *Psychological Record*, *57*(2), 219–232. https://doi.org/10.1007/BF03395573

Jonides, J. (2004). How does practice makes perfect? *Nature Neuroscience*, *7*(1), 10–11. https://doi.org/10.1038/nn0104-10

Joyce, T., Crockett, S., Jaeger, D. A., Altindag, O., & O'Connell, S. D. (2015). Does classroom time matter? *Economics of Education Review*, *46*, 64–77. https://doi.org/10.1016/j.econedurev.2015.02.007

Karlen, Y., Suter, F., Hirt, C., & Maag Merki, K. (2019). The role of implicit theories in students' grit, achievement goals, intrinsic and extrinsic motivation, and achievement in the context a long-term challenging task. *Learning and Individual Differences*, *74*(December 2018), 101757. https://doi.org/10.1016/j.lindif.2019.101757

Karpicke, J. D. (2016). *A powerful way to improve learning and memory.* . Psychological Science Agenda. http://www.apa.org/science/about/psa/2016/06/learning-memory

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced Retrieval : Absolute Spacing Enhances Learning Regardless of Relative Spacing. *Journal of Experimental Psychology*, *37*(5), 1250–1257. https://doi.org/10.1037/a0023436

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science*, *139*(February), 772–774.

Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471–479. https://doi.org/10.1080/09658210802647009

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102

Kirschner, P., Sweller, J., & Clark, R. (2006). Why Unguided Learning Does Not Work: An Analysis of the Failure of Discovery Learning, Problem-Based Learning, Experiential Learning and Inquiry-Based Learning. *Educational Psychologist*, *41*(2), 75–86.

Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. https://doi.org/10.1037//0033-2909.119.2.254

Koenker, R. (2019). *quantreg: Quantile Regression*. https://cran.r-project.org/package=quantreg

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, *46*(1), 33–50.
Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences*, *19*(1), 47–52. https://doi.org/10.1016/j.lindif.2008.07.001
Komarraju, M., Karau, S. J., Schmeck, R. R., & Avdic, A. (2011). The Big Five personality traits, learning styles, and academic achievement. *Personality and Individual Differences*, *51*(4), 472–477. https://doi.org/10.1016/j.paid.2011.04.019
Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful Retrieval Attempts Enhance Subsequent Learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(4), 989–998. https://doi.org/10.1037/a0015729
Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory*, *17*(5), 493–501. https://doi.org/10.1080/09658210902832915
Krause, A., Rinne, U., & Zimmermann, K. F. (2012). Anonymous job applications of fresh Ph.D. economists. *Economics Letters*, *117*, 441–444. https://doi.org/10.1016/j.econlet.2012.06.029
Kulhavy, R. W. (1977). Feedback in Written Instruction. *Review of Educational Research*, *47*(2), 211–232. https://doi.org/10.3102/00346543047002211
Kulik, J. A., & Kulik, C. L. C. (1988). Timing of Feedback and Verbal Learning. *Review of Educational Research*, *58*(1), 79–97. https://doi.org/10.3102/00346543058001079
Lane, A. M., Hall, R., & Lane, J. (2004). Self-efficacy and statistics performance among Sport Studies students. *Teaching in Higher Education*, *9*(4), 435–448.
Linnenbrink-Garcia, L., Tyson, D. F., & Patall, E. A. (2008). When are achievement goal orientations beneficial for academic achievement? A closer look at main effects and moderating factors. *Revue Internationale de Psychologie Sociale*, *21*(1–2), 19–70.
Macher, D., Papousek, I., Ruggeri, K., & Paechter, M. (2015). Statistics anxiety and performance: Blessings in disguise. *Frontiers in Psychology*, *6*(AUG), 4–7.
Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501
McKenzie, K., & Schweitzer, R. (2001). Who Succeeds at University? Factors predicting academic performance in first year Australian university students. *Higher Education Research & Development*, *20*(1), 21–33.
Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596–606. https://doi.org/https://doi.org/10.1016/S0022-5371(70)80107-4
Morrison, G. R., & Anglin, G. J. (2005). Research on cognitive load theory: Application to e-learning. *Educational Technology Research and Development*, *53*(3), 94–104. https://doi.org/10.1007/BF02504801
Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLoS ONE*, *10*(7), 1–23. https://doi.org/10.1371/journal.pone.0120644
O'Donoghue, T., & Rabin, M. (1999). *Doing It Now or Later By*. https://doi.org/10.1007/s11606-008-0688-8
Panus, P. C., Stewart, D. W., Hagemeier, N. E., Thigpen, J. C., & Brooks, L. (2014). A subgroup analysis of the impact of self-testing frequency on examination scores in a pathophysiology course. *American Journal of Pharmaceutical Education*, *78*(9). https://doi.org/10.5688/ajpe789165

Park, J., Yu, R., Rodriguez, F., Baker, R., Smyth, P., & Warschauer, M. (2018). Understanding Student Procrastination via Mixture Models. *Proceedings of the 11th International Conference on Educational Data Mining (EDM)*, 187–197.

Payne, S. C., Youngcourt, S. S., & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. In *Journal of Applied Psychology* (Vol. 92, Issue 1, pp. 128–150). American Psychological Association. https://doi.org/10.1037/0021-9010.92.1.128

Pintrich, P. R. (2000). An Achievement Goal Theory Perspective on Issues in Motivation Terminology, Theory, and Research. *Contemporary Educational Psychology*, *25*(1), 92–104. https://doi.org/https://doi.org/10.1006/ceps.1999.1017

Pintrich, P. R. (2003). A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts. *Journal of Educational Psychology, 95*(4), 667–686. https://doi.org/10.1037/0022-0663.95.4.667

Plante, I., O'Keefe, P. A., & Théorêt, M. (2013). The relation between achievement goal and expectancy-value theories in predicting achievement-related outcomes: A test of four theoretical conceptions. *Motivation and Emotion*, *37*(1), 65–78. https://doi.org/10.1007/s11031-012-9282-9

Rawson, K. A., & Dunlosky, J. (2012). When Is Practice Testing Most Effective for Improving the Durability and Efficiency of Student Learning? *Educational Psychology Review*, *24*(3), 419–435. https://doi.org/10.1007/s10648-012-9203-1

Revelle, W. (2020). *psych: Procedures for Psychological, Psychometric, and Personality Research*. https://cran.r-project.org/package=psych

Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2016). True grit and genetics: Predicting academic achievement from personality. *Journal of Personality and Social Psychology, 111*(5), 780–789.

Rodriguez, F., Fischer, C., Zhou, N., Warschauer, M., & Massimelli, J. (2021). Student spacing and self-testing strategies and their associations with learning in an upper division microbiology course. *SN Social Sciences*, *1*(38), 1–21. https://doi.org/10.1007/s43545-020-00013-5

Rodriguez, F., Kataoka, S., Janet Rivas, M., Kadandale, P., Nili, A., & Warschauer, M. (2018). Do spacing and self-testing predict learning outcomes? *Active Learning in Higher Education*.

Rodriguez, F., Rivas, M. J., Matsumura, L. H., Warschauer, M., & Sato, B. K. (2018). How do students study in STEM courses? Findings from a light-touch intervention and its relevance for underrepresented students. *PLOS ONE*, *13*(7), 1–20.

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning. Taking Memory Tests Improves Long-Term Retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, *1*(4), 242–248. https://doi.org/https://doi.org/10.1016/j.jarmac.2012.09.002

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Ross, S. M., & Morrison, G. R. (1989). In search of a happy medium in instructional technology

research: Issues concerning external validity, media replications, and learner control. *Educational Technology Research and Development*, *37*(1), 19–33.
Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology*, *73*(1), 93–105. https://doi.org/10.1037/0022-0663.73.1.93
Schupp, J., & Gerlitz, J. (2014). Big Five Inventory-SOEP (BFI-S). *Zusammenstellung Sozialwissenschaftlicher Items Und Skalen (ZIS)*.
Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795
Sloboda, J. A., Davidson, J. W., Howe, M. J. A., & Moore, D. G. (1996). The role of practice in the development of performing musicians. *British Journal of Psychology*, *87*, 287–309. https://doi.org/10.1017/CBO9781107415324.004
Spitzer, H. F. (1939). Educational psychology. *Journal of Educational Psychology*, *XXX*(9), 641–656.
Stanger-Hall, K. F., Shockley, F. W., & Wilson, R. E. (2011). Teaching students how to study: A workshop on information processing and self-testing helps students learn. *CBE Life Sciences Education*, *10*(2), 187–198. https://doi.org/10.1187/cbe.10-11-0142
Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. In *Journal of Personality and Social Psychology* (Vol. 97, Issue 6, pp. 1115–1128). American Psychological Association. https://doi.org/10.1037/a0017048
Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences*, *16*(1), 1–12. https://doi.org/10.1016/j.lindif.2005.06.004
Wang, Z., Gong, S. Y., Xu, S., & Hu, X. E. (2019). Elaborated feedback and learning: Examining cognitive and motivational influences. *Computers and Education*, *136*(January), 130–140. https://doi.org/10.1016/j.compedu.2019.04.003
Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81. https://doi.org/10.1006/ceps.1999.1015
Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology*, *10*(3087), 1–14. https://doi.org/10.3389/fpsyg.2019.03087
Yperen, N. W. Van, Blaga, M., & Postmes, T. (2014). *A Meta-Analysis of Self-Reported Achievement Goals and Nonself-Report Performance across Three Achievement Domains (Work , Sports , and Education)*. *9*(4). https://doi.org/10.1371/journal.pone.0093594
Zimmerman, B. J. (1989). A Social Cognitive View of Self-Regulated Academic Learning. *Journal of Educational Psychology*, *81*(3), 329–339. https://doi.org/10.1037/0022-0663.81.3.329