Who Benefits from Additional Online Practice Opportunities in a Gateway Math Course?

Authors:

Name	Affiliation	E-Mail
Jakob Schwerter	*TU Dortmund University*	Jakob.schwerter@tu-dortmund.de
Thomas Dimpfl	*University of Hohenheim*	thomas.dimpfl@uni-hohenheim.de
Johannes Bleher	*University of Hohenheim*	johannes.bleher@uni-hohenheim.de
Kou Murayama	*University of Tübingen*	k.murayama@uni-tuebingen.de
Fani Lauermann	*TU Dortmund University*	Fani.Lauermann@tu-dortmund.de

Background/Aims

Practice testing is one of the best-researched learning techniques in the educational sciences and has shown considerable promise in boosting student learning across diverse settings (Yang et al., 2021). However, surprisingly little research has evaluated the effectiveness of self-testing in demanding math courses in higher education (Carvalho et al., 2022; Wong et al., 2017), which are often plagued by high dropout rates and low performance (Benden & Lauermann, 2022). Moreover, self-testing in mathematics is conceptually different from fact recall in that it is about knowing *how* to solve a problem rather than memorizing the solution. The present study thus examined (*a*) the effectiveness of self-testing via two online practice opportunities in a gateway math course, and (*b*) the extent to which different subgroups of students benefitted from self-testing.

Method

Economics/business administration students (N=188, 59% female) enrolled in *a gateway math course* participated in the study. Students participated in three 30-minute online self-testing sessions, each covering content from the previous 3-to-4 weeks. A daily self-testing app was introduced in the second half of the course to practice content from the first half. Students received immediate corrective feedback. Students' gender, prior performance, motivational beliefs, personality traits, time preferences, and course goals were assessed at the beginning of the semester (**Table 1)**.

Results

First, OLS regressions revealed that self-testing with the practice tests improved students' exam scores by about 5 points (of 90). A double-selection regression (Belloni et al., 2014) was then used to select relevant control variables, reduce a potential omitted variable bias, and avoid overfitting in the prediction of exam performance; the self-testing effect remained significant but decreased to 2.5 points (**Table 2**). Use of the daily self-testing app had a significant effect on students' exam performance in the simple OLS regression, but not in the post-double-selection regression.

Second, heterogeneous practice effects were examined for a set of LASSO-selected control variables (**Table 2**). A post-LASSO regression showed that risk-averse students, those who planned repeated practice, and students with a higher math self-concept were most likely to benefit from self-testing. Notably, students' gender, achievement goals, and personality traits did not contribute to differential practice effects. Only one significant interaction emerged for self-testing via the daily app; higher open-mindedness corresponded to greater benefits from self-testing.

Finally, a quantile regression revealed that self-testing via the practice tests was most effective for students who obtained a lower end-of-term exam grade (see **Figure 1**).

Discussion

Our results show that self-testing is a promising intervention to support students' academic success in demanding math courses. This intervention appears to support those who need it most—low-performing students. In contrast, students' gender, achievement goals, and personality traits were not associated with differential self-testing effects, which supports the broad applicability of self-testing. The modality of implementation, however, warrants further consideration, as course-embedded practice tests were more effective than a daily self-testing app.

New data from an ongoing follow-up study will also be presented, which uses within-person randomization to assess content- and time-specific effects of self-testing on later performance.

References

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

Benden, D. K., & Lauermann, F. (2022). Students' motivational trajectories and academic success in math-intensive study programs: Why short-term motivational assessments matter. *Journal of Educational Psychology 114*(5), 1062–1085. https://doi.org/10.1037/edu0000708

Carvalho, P. F., Mclaughlin, E. A., Koedinger, K. R. (2022). Varied practice testing is associated with better learning outcomes in self-regulated online learning. *Journal of Educational Psychology*. *Advance online publication*. https://doi.org/http://dx.doi.org/10.1037/edu0000754

Wong, S. S. H., Ng, G. J. P., Tempel, T., & Lim, S. W. H. (2017). Retrieval practice enhances analogical problem solving. *Journal of Experimental Education*, *87*(1), 128–138. https://doi.org/10.1080/00220973.2017.1409185

Yang, C., Zhao, W., Luo, L., Sun, B., Potts, R., & Shanks, D. R. (2021). Testing potential mechanisms underlying test-potentiated new learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition. Adcance online publication*. https://doi.org/10.1037/xlm0001021

Tables and Figures for Paper #1:

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	α
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 variable (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 high school	219	0.83	0.38	188	0.86	0.35	0.00	1.00	-
Last math grade in high school	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	-
Additional 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	-
Expectancy-value constructs (Gaspard et al., 2017)									
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 personality traits (Schupp and Gerlitz, 2014)									
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 (Elliot and Murayama, 2008)									
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

(continues on next page)
(continued from previous page)

Present bias preferences (Frederick and Loewenstein, 2002)									
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	-	-	-

Note. The table shows the number of observations (N), means, and standard deviations (SD) per variable for the full sample, including incomplete cases, and for a subsample with complete data. Supplemental analyses show that the reported regression results are unlikely to be affected by selection bias. MAD = The app 'a matrix a day'. * 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).

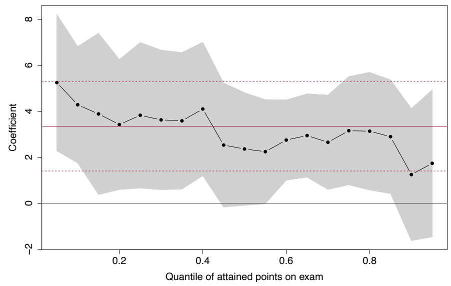
Table 2: *Main OLS and double-selection regression results*

	Dependent variable: Standardized points on final exam					
	Practice variables only	PDSR - LASSO	PDSR - Random Forest	PDSR - xgBoost	Selection of all possible interaction terms	Selection of interactions terms of EVT beliefs
Constant	-2.401***	-0.824	-0.078	-0.195	-1.335	-0.216
	(0.328)	(0.847)	(1.256)	(0.753)	(1.845)	(0.502)
Practice test attempts (PTA)	0.226***	0.215***	0.203***	0.205***	-0.060	-0.137
	(0.074)	(0.068)	(0.065)	(0.064)	(0.195)	(0.093)
Practice test performance (PTP)	0.022***	0.010**	0.010**	0.010**	-0.002	0.011***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.010)	(0.004)
MAD submissions (MADS)	0.007	0.006	0.001	0.004	-0.003	0.004**
	(0.012)	(0.008)	(0.009)	(0.009)	(0.004)	(0.002)
MAD percentage (MADP)	0.004**	0.004**	0.005***	0.004**	-0.017***	
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	
PTA × self-concept					0.005	0.114***
					(0.068)	(0.028)
PTA × risk					0.221***	
					(0.061)	
PTA × additional practice after practice tests					0.071**	
					(0.036)	
PTP × self-concept					0.005	
					(0.003)	
MADS × openness					0.005***	
					(0.002)	
MADP × neuroticism					0.002*	
					(0.001)	
Additional controls	No	Yes†	Yes‡	Yes§	Yes∝	YesΓ
Observations	188	188	188	188	188	188
Adjusted R^2	0.213	0.446	0.464	0.410	0.422	0.432

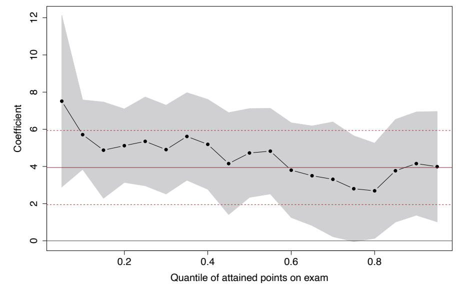
Note. Heteroskedasticity-robust standard errors are shown in parentheses. MAD = The app 'a matrix a day'. Columns 2 to 4 show the post-double selection regression (PDSR) results for each algorithm (LASSO, random forest, and xgBoost). For all three columns, we used a double selection process to select relevant control variables 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. Columns 5 and 6 show simple selection results using LASSO, including interaction terms with the practice variables. Column 5 includes all possible interactions between the practice variables (see Table 1) and the additional variables listed below the practice variables in Table 1, while column 6 focuses on EVT beliefs. ∝ and Γ Include, apart from the interactions, the high school GPA.

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Figure 1: *Quantile Regression Results*

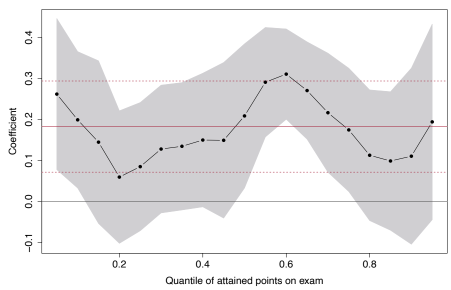


(a) Intersection

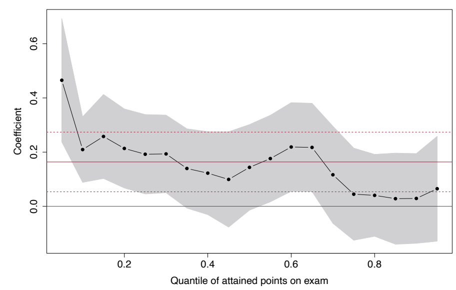


(b) Union

(i) Practice Test Attempts

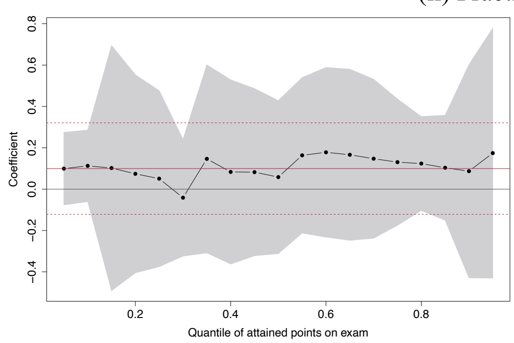


(a) Intersection

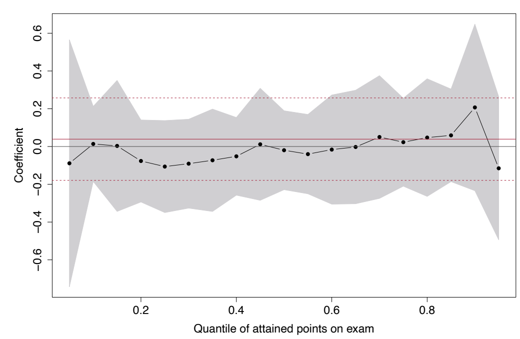


(b) Union

(ii) Practice Test Performance

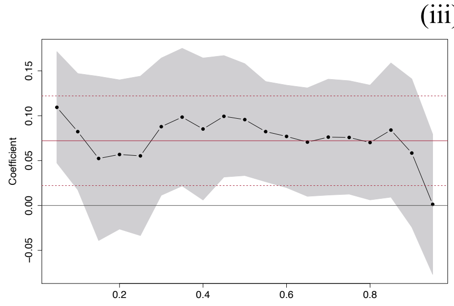


(a) Intersection

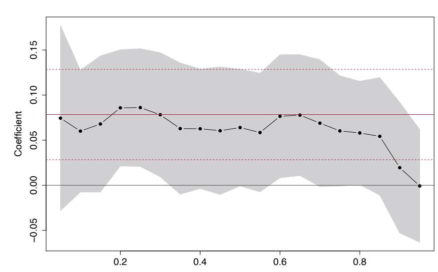


(b) Union

(iii) MAD submissions



(a) Intersection



(b) Union

(iv) MAD percentage

Note. The panels show the coefficient estimates mentioned in the headings from (i) to (iv) across quantiles and estimated based on a quantile regression on the final exam points. The coefficients across quantiles were estimated either including (a) the intersection of all machine-learning selected control variables or (b) the union of all selected control variables. The solid red horizontal line shows the value obtained in an equivalent OLS regression, and the dotted red lines show 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 were calculated using a wild bootstrap procedure with 5,000 replications.