technische universität
dortmund

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

Jakob Schwerter[1], Fani Lauermann[2], Thomas Dimpfl[3], Johannes Bleher[3], & Kou Murayama[4]

[1]TU Dortmund University, [2]University of Bonn,
[3]University of Hohenheim, [4]University of Tübingen

AERA – What works, when, for whom? – Chicago – XX.04.2023

FAIR

funded by:
Ministry of Culture and Science
of the State of
North Rhine-Westphalia

Self-Testing Opportunities

Self-Testing – Practice Testing – Self-Assessment

- Self-testing is one of the best-researched learning techniques in the educational sciences (Yang et al., 2021)
- It shows considerable promise in boosting student learning across diverse settings
- Evidence from laboratory studies (e.g., Karpicke & Blunt, 2011; Lim et al., 2015) and (less often) authentic educational settings (Roediger et al., 2011; Förster et al., 2018)

Self-Testing Opportunities

Self-Testing – Practice Testing – Self-Assessment

- Self-testing is one of the best-researched learning techniques in the educational sciences (Yang et al., 2021)
- It shows considerable promise in boosting student learning across diverse settings
- Evidence from laboratory studies (e.g., Karpicke & Blunt, 2011; Lim et al., 2015) and (less often) authentic educational settings (Roediger et al., 2011; Förster et al., 2018)

The Gap: Self-Testing in Math & Missing treatment heterogeneity

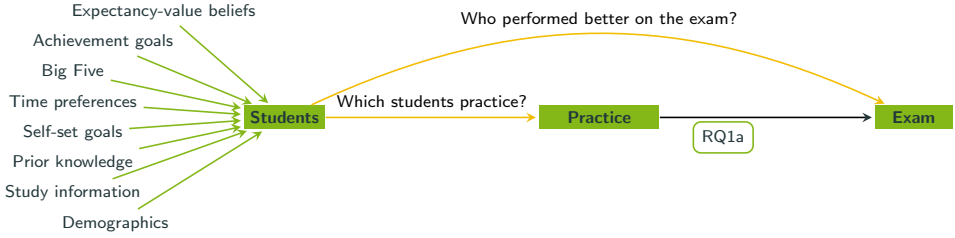
- Self-testing in mathematics is conceptually different from fact recall:
- Solution steps $\neq$ solutions $\Rightarrow$ "Derivative of the function $f(x) = 2 \cdot e^{4 \cdot x} + 4x + 2$"
- Little evidence for higher education math courses (Carvalho et al., 2022; Förster et al., 2018; Wong et al., 2017; Yeo & Fazio, 2019)
- Lack of evidence on possible heterogenous treatment effects

Present study

- RQ1: How effective are online self-testing opportunities for increasing student performance in a gateway math course?
  - RQ1a: Does increased self-testing relate to better exam grades?

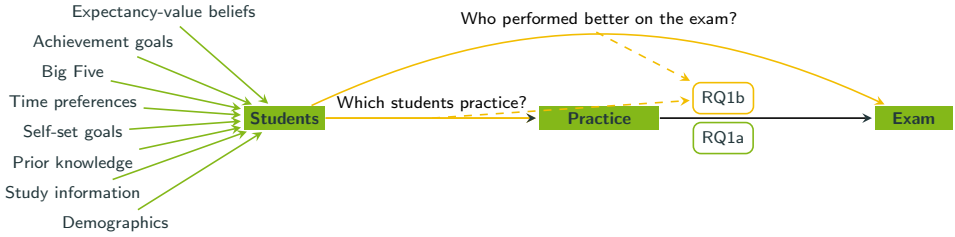
Present study

- RQ1: How effective are online self-testing opportunities for increasing student performance in a gateway math course?
  - RQ1a: Does increased self-testing relate to better exam grades?



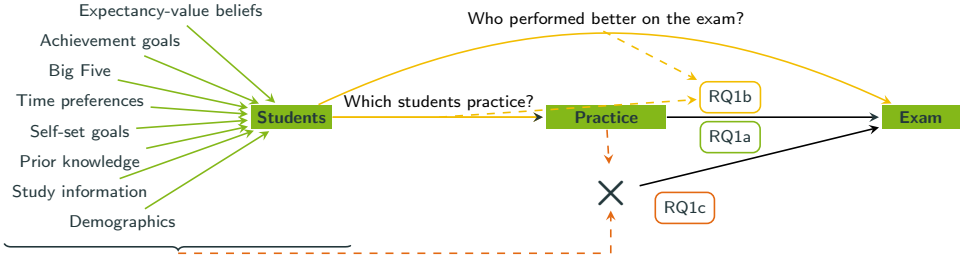
Present study

- RQ1: How effective are online self-testing opportunities for increasing student performance in a gateway math course?
  - RQ1a: Does increased self-testing relate to better exam grades?
  - RQ1b: Is the self-testing effect robust after including demographic, motivational, self-regulated, and personality variables?



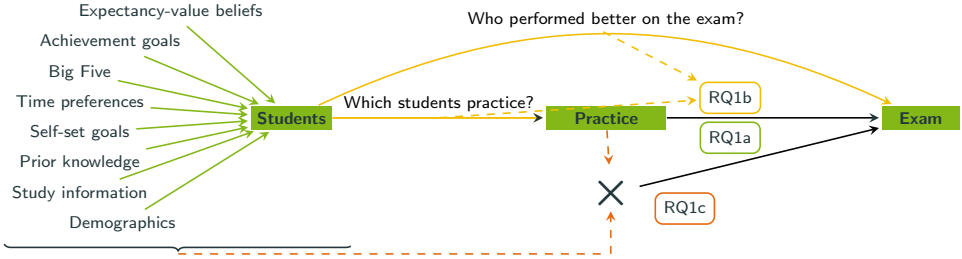
Present study

- RQ1: How effective are online self-testing opportunities for increasing student performance in a gateway math course?
  - RQ1a: Does increased self-testing relate to better exam grades?
  - RQ1b: Is the self-testing effect robust after including demographic, motivational, self-regulated, and personality variables?
  - RQ1c: Do demographic, motivational, self-regulated, and personality variables moderate the effects of online self-testing on performance?



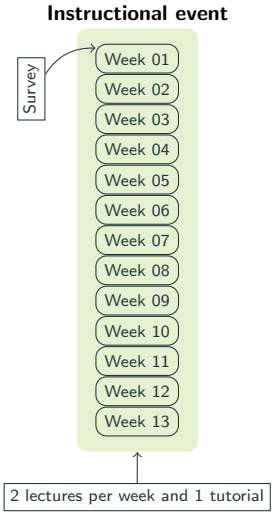
Present study

- RQ1: How effective are online self-testing opportunities for increasing student performance in a gateway math course?



- RQ2: Which exam performance group would have benefitted the most, if they have practiced more?

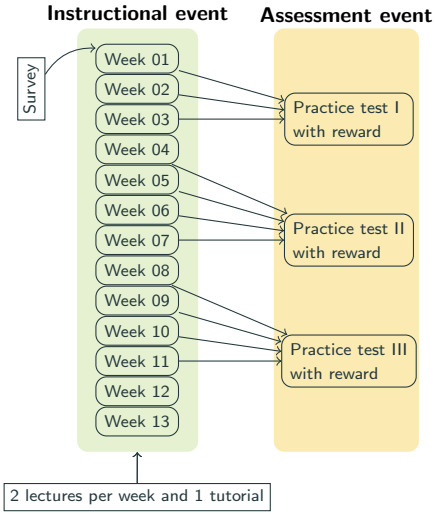
Research Design & Sample



Sample and Context

- Economics/business administration students (N=188, 59% female)
- 1st semester gateway math course

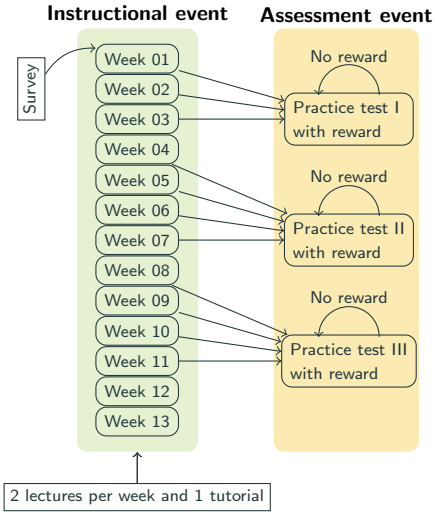
Research Design & Sample



Sample and Context

- Economics/business administration students (N=188, 59% female)
- 1st semester gateway math course

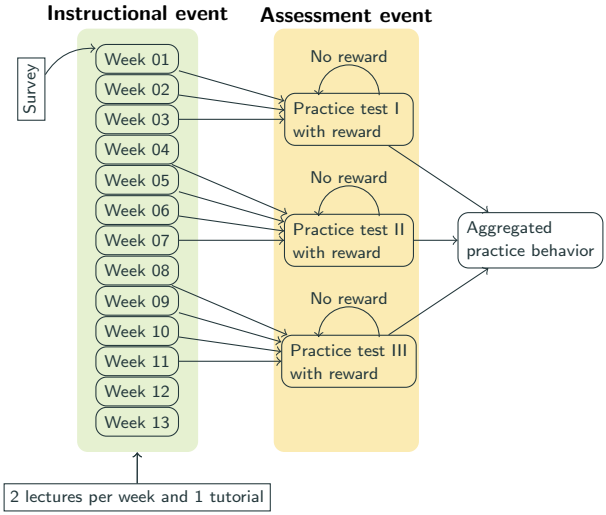
Research Design & Sample



Sample and Context

- Economics/business administration students (N=188, 59% female)
- 1st semester gateway math course
- Practice tests: Content from the previous 3-to-4 weeks
- Immediate corrective feedback

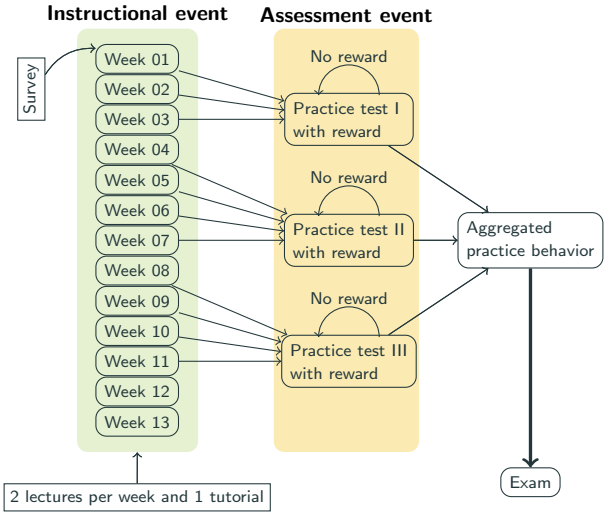
Research Design & Sample



Sample and Context

- Economics/business administration students (N=188, 59% female)
- 1st semester gateway math course
- Practice tests: Content from the previous 3-to-4 weeks
- Immediate corrective feedback

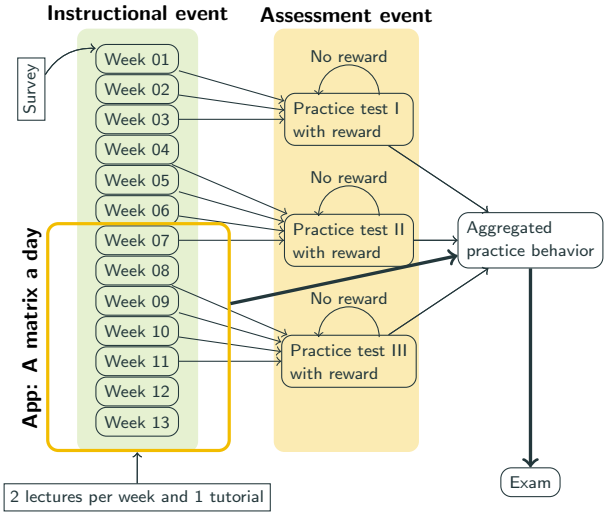
Research Design & Sample



Sample and Context

- Economics/business administration students (N=188, 59% female)
- 1st semester gateway math course
- Practice tests: Content from the previous 3-to-4 weeks
- Immediate corrective feedback

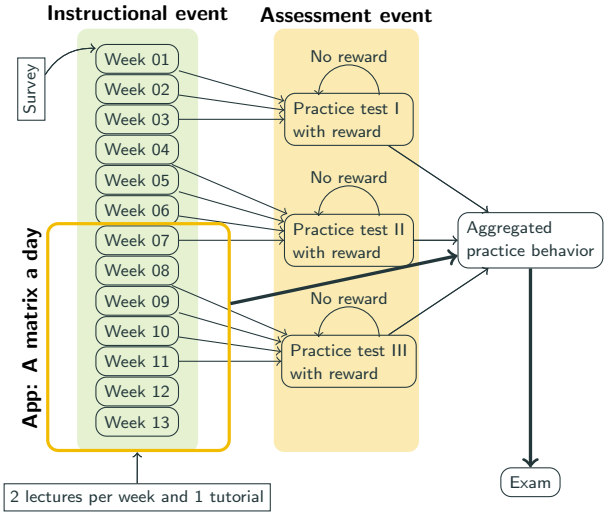
Research Design & Sample



Sample and Context

- Economics/business administration students (N=188, 59% female)
- 1st semester gateway math course
- Practice tests: Content from the previous 3-to-4 weeks
- Immediate corrective feedback

Research Design & Sample



Sample and Context

- Economics/business administration students (N=188, 59% female)
- 1st semester gateway math course
- Practice tests: Content from the previous 3-to-4 weeks
- Immediate corrective feedback
- Explanatory variables:
  - Practice test attempts
  - Practice test performance
  - App submissions
  - App performance
- Outcome: Points in final exam

Scales

Variables	Explanation	Mean	SD	α
Outcomes				
Exam	Grade points final exam	43.58	17.19	
Practice (participation and performance)				
Practice test attempts	Number of practice tests with and without reward	3.25	0.87	
Practice test performance	Average points in practice tests	71.10	15.85	
App submissions	Submission in "A Matrix A Day" App	26.65	36.92	
App performance	Average performance in "A Matrix A Day" App	2.53	7.91	
Potential moderators				
Demographics	Gender, HS GPA, Study information, Working, ...			
Expectancy value beliefs	Self-concept; Intrinsic, Attainment, Utility Value; Cost	2.4 – 2.8	0.3 – 0.6	.71 – .88
Achievement goals	Mastery & Performance approach & avoidance	5.0 – 6.1	0.7 – 1.6	.64 – .92
Big five	Conscientiousness, Extraversion, Agreeableness, Openness, Neuroticism	4.6 – 4.9	0.6 – 1.2	.65 – .82
Time preferences	Risk, Discount factor, Present bias			
Self-set course goals	Number and performance of practice tests with reward, intention to self-test without reward, exam grade expectation			

Note: Mathematical expectancy-value beliefs from Gaspard et al. (2017; Achievement goals from Elliot and Murayama (2008; Big Five personality traits from Schupp and Gerlitz (2014); Time preferences from Frederick and Loewenstein (2002)

RQ1a & RQ1b – OLS and Post-double selection results

	Dependent variable: Standardized points on final exam			
	Practice variables only	PDSR LASSO	PDSR Rand. For.	PDSR XGBoost
Constant	-2.401***	-0.824	-0.078	-0.195
	(0.328)	(0.847)	(1.256)	(0.753)
Practice test attempts	0.226***	0.215***	0.203***	0.205***
	(0.074)	(0.068)	(0.065)	(0.064)
Practice test performance	0.022***	0.010**	0.010**	0.010**
	(0.004)	(0.004)	(0.004)	(0.004)
App submissions	0.007	0.006	0.001	0.004
	(0.012)	(0.008)	(0.009)	(0.009)
App performance	0.004**	0.004**	0.005***	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
Additional controls	No	Yes	Yes	Yes
Observations	188	188	188	188
Adjusted R^2	0.213	0.446	0.464	0.410

PDSR: Post-double selection regression $\Rightarrow$ Selecting important variables for the dependent variable and the four practice variables using LASSO, RF, and XGBoost.

RQ1c – Prediction Rule Ensembles* – Importance and Rules

Table 1: Variable importance of selected variables

	Variables	Importance
1	HS GPA	0.42
2	Practice test performance	0.22
3	Math self-concept	0.18
4	HS math grade	0.16
5	Practice tests attempts	0.11
6	App submissions	0.09
7	App performance	0.08

Table 2: Prediction rules

Rules	Coefficient
$SC_{Math} \leq 3.25$ & App submissions ≤ 3	-0.37
HS GPA ≤ 3.7 & App performance ≤ 45.71	-0.35
PT attempts ≤ 3 & HS GPA ≤ 3.7	-0.34
HS GPA > 3.3 & HS math grade > 2.6	0.30
HS GPA > 4.1 & PT performance > 65.33	0.28
HS GPA > 3.6 & HS math grade > 2.6	-0.20
HS GPA > 3.7 & PT performance > 78.03	0.20
PT performance > 56.67 & $SC_{Math} > 2.67$	0.17
PT performance > 56.67 & HS math grade > 3	0.16
HS GPA > 3.3 & $SC_{Math} > 2.25$	0.16
PT attempts ≤ 3 & PT performance ≤ 72.41	-0.14

PT = Practice test

*Fokkema, M., & Strobl, C. (2020). Fitting Prediction Rule Ensembles to Psychological Research Data: An Introduction and Tutorial. *Psychological Methods*, 25(5), 636–652. https://doi.org/10.1037/met0000256

RQ1c – Prediction Rule Ensembles – Partial Dependence Plots

Figure 1: Interaction between self-testing attempts & HS GPA

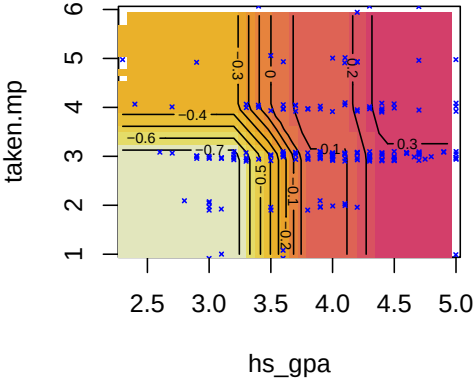
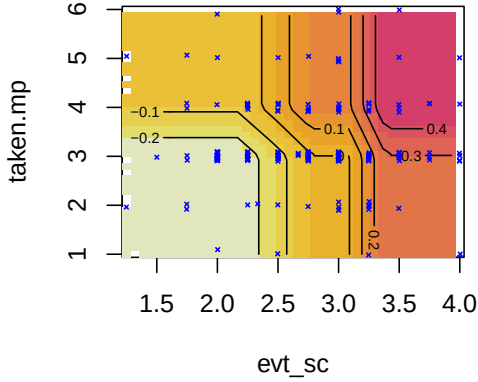
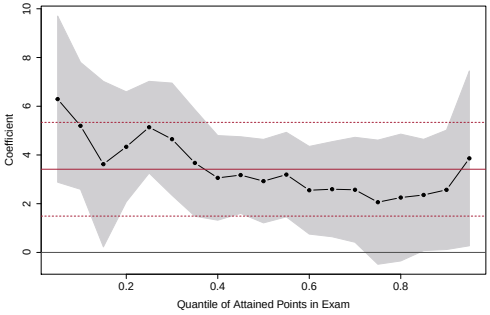


Figure 2: Interaction between self-testing attempts & Math self-concept



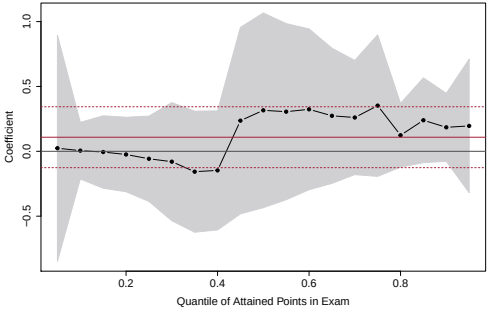
RQ2 – Quantile regressions

Figure 3: Practice test attempts



- Students with fewer exam points would have benefitted more than with more points

Figure 4: APP submissions



- Somewhat constant
- Good students would have benefitted from more practice in matrix algebra

Summary

Effectiveness of additional online self-testing

Practice tests:

- Effectiveness of self-testing with corrective feedback in a gateway higher education math course
  - Authentic learning environment
  - Higher order learning
  - Parameterized exercises $\rightarrow$ Near content transfer

$\Rightarrow$ Online practice tests are a promising intervention to support students

- Intervention with support for low-performing students

App 'Matrix a Day'

- Course-embedded practice tests were more effective than the daily self-testing app

$\Rightarrow$ The modality of implementation warrants further consideration

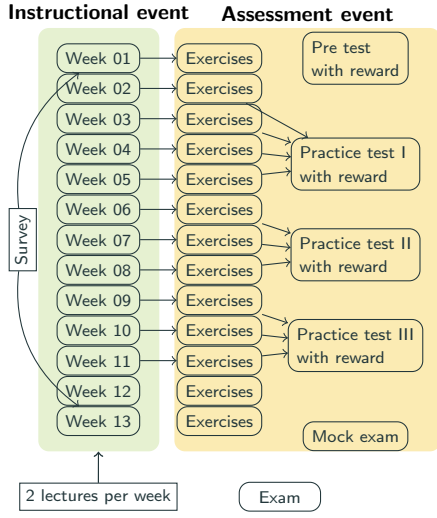
Outlook

Follow-up study

- Are the results content-specific vs. transfer effects to new content?
- Weekly online exercises & practice tests ($N = 606$)
- Within-person randomization: Variation in exercises students get

Preliminary results of practice test performance:

- The more students practiced the better in the practice tests
- Students performed better only on content they had practiced before



Preliminary results

Students are

- ... better on content that they had practiced
- ... worse on content that they had **not** practiced (but not stat. sig.!)

Performance in practice tests exercises of		
	Practiced	Not practiced
(Intercept)	0.668***	0.790***
	(0.040)	(0.037)
Group	0.054*	−0.022
	(0.026)	(0.024)
R^2	0.012	0.002
Adj. R^2	0.009	−0.000
Num. obs.	354	354

Note:

Thank you!

Dr. Jakob Schwerter
jakob.schwerter@tu-dortmund.de
https://www.researchgate.net/profile/Jakob-Schwerter

- Schwerter, J., Dimpfl, T., Bleher, J., & Murayama, K. (2022). Benefits of additional online practice opportunities in higher education. *Internet and Higher Education, 53*, 100834. https://doi.org/https://doi.org/10.1016/j.iheduc.2021.100834
- Schwerter, J., & Brahm, T. (2022). Voluntary e-learning exercises support students in mastering statistics. *Preprint at SocArXiv*. https://doi.org/10.31235/osf.io/8p2xh
- Schwerter, J., Wortha, F., & Gerjets, P. (2022). E-Learning with multiple-try-feedback: Can hints foster students' achievement during the semester? *Educational Technology Research and Development, 70*, 713-736. https://doi.org/10.1007/s11423-022-10105-z

Scales

Variables	Explanation	N	Mean	SD
Individual characteristics				
Female	Female vs Male	175	0.56	0.50
High school GPA	Numeric value from 1 to 5	175	2.07	0.59
Advanced math in HS	Indicator if yes or no	175	0.85	0.36
Last math grade in HS	Numeric value from 1 to 5	175	2.59	1.09
International studies	Indicator if yes or no	175	0.44	0.50
Sports degree	Indicator if yes or no	175	0.05	0.22
Minor	Indicator if yes or no	175	0.13	0.33
Work to finance studying	Indicator if yes or no	175	0.19	0.40
Semester	Number of semester	175	1.26	1.20
Re-taking exam	Indicator if yes or no	175	2.00	0.19
Present bias preferences				
Risk	Value bigger .5 implies risk aversion	175	0.68	0.20
Discount factor	Value below 1 indicates to be impatient	175	0.94	0.55
Present bias	Value below 1 makes individuals more impatient when the present is involved	175	1.05	0.18
Self-set goals				
How many midterms?	1 to 3	175	2.82	0.46
How good in midterms?	Percentage of correct answers in-between 0 and 100	175	0.79	0.14
Practice after midterms?	Aim to practice without rewards?	175	1.22	0.44
Which grade in exam?	Aim of the final grade	175	2.05	0.62

Scales

Variables	Explanation	N	Mean	SD	α
Expectancy value beliefs					
Self-concept	Math is easy for me	175	2.38	0.25	0.86
Intrinsic value/Dispositional Interest	Math is fun for me	175	2.79	0.60	0.87
Attainment value	It is personally important to me to master mathematical content	175	2.35	0.36	0.71
Utility value	For my professional future it will pay off to be good at math	175	3.54	0.54	0.88
Cost	Dealing with math costs me a lot of energy	175	2.40	0.55	0.75
Achievement goals					
Mastery approach	My goal is to learn as much as I can in this course	175	6.12	0.74	0.64
Mastery avoidance	My goal is to avoid learning less than I could in this course	175	5.62	0.98	0.71
Performance approach	I strive to do well compared to other students in this course	175	5.04	1.44	0.87
Performance avoidance	I strive to avoid being worse than others in this course	175	4.99	1.61	0.92
Big five					
Conscientiousness	... works thoroughly	175	4.87	0.55	0.65
Extraversion	... is communicative, talkative	175	4.76	0.65	0.82
Agreeableness	... can forgive	175	4.86	0.62	0.62
Openness	... is original, introduces new ideas	175	4.85	1.15	0.65
Neuroticism	... is slightly nervous	175	4.60	0.75	0.68

Why Self-Testing (in Higher Education)?

Improving students' performance

- A significant proportion of STEM students are plagued by low performance or high drop out rates (Benden & Lauermann, 2022; Chen & Soldner, 2013; Heublein & Schmelzer, 2018)
- Low performance influences labor market opportunties (REF)
- Dropout incurs significant personal and societal costs (Faas et al., 2018; OECD, 2019)
- Study behavior and performance is difficult to change/ improve
  (Jaggars & Xu, 2016; Oreopoulos & Petronijevic, 2020; Sussan & Son, 2014; Susser & McCabe, 2013)

Why online?

- Whenever & wherever (e.g., ...)
- Immediate knowledge of correct response feedback

Regression result summary

How effective are the additional self-testing opportunities?

- Self-testing with the practice tests improved students' exam scores by about 5 points (of 90)
- Self-testing effect remained significant but decreased to 2.5 points
- Use of the daily self-testing app only significant in simple OLS regression
- Risk-averse students, those who planned repeated practice, and students with a higher math self-concept were most likely to benefit from self-testing
- 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

The Gap (However)

Self-Testing in Mathematical Fields

- Self-testing in mathematics is conceptually different from fact recall:
- Solution steps $\neq$ solutions $\Rightarrow$ "Derivative of the function $f(x) = 2 \cdot e^{4 \cdot x} + 4x + 2$"
- Little evidence for higher education math courses

  (Carvalho et al., 2022; Förster et al., 2018; Wong et al., 2017; Yeo & Fazio, 2019)

  - Laboratory studies with exercises involving the Poisson distribution (Yeo & Fazio, 2019)
  - Online quizzes in a statistics course (Förster et al., 2018)

? Unclear how, when, and for whom practice testing in mathematics is required

Who benefits from self-testing opportunities?

Heterogeneity

- Most of the self-testing results are general treatment effects without much heterogeneity analysis
- Notable exceptions:
  - Students with lower prior knowledge benefit most from self-testing
  - Students with higher prior knowledge become less overconfident with self-testing (Cogliano et al., 2019)
  - Missing: Interactions with motivation or personality variables

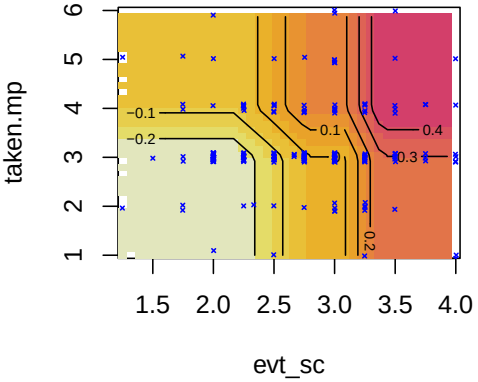
OLS and Post-double selection results

	Dependent variable: Standardized points on final exam			
	Practice variables only	PDSR LASSO	Practice × all moderators	Practice × EVT beliefs
Constant	-2.401***	-0.824	-1.335	-0.216
	(0.328)	(0.847)	(1.845)	(0.502)
Practice test attempts (PTA)	0.226***	0.215***	-0.060	-0.137
	(0.074)	(0.068)	(0.195)	(0.093)
Practice test performance (PTP)	0.022***	0.010**	-0.002	0.011***
	(0.004)	(0.004)	(0.010)	(0.004)
'Matrix a Day' submissions (MADS)	0.007	0.006	-0.003	0.004**
	(0.012)	(0.008)	(0.004)	(0.002)
'Matrix a Day' performance (MADP)	0.004**	0.004**	-0.017***	0.004**
	(0.002)	(0.002)	(0.006)	(0.002)
Practice test attempts × self-concept			0.005	0.114***
			(0.068)	(0.028)
Practice test attempts × risk-avers			0.221***	
			(0.061)	
Practice test attempts × Self-set goal: Additional practice			0.071**	
			(0.036)	
Practice test performance × self-concept			0.005	
			(0.003)	
App submissions × openness			0.005***	
			(0.002)	
App performance × neuroticism			0.002*	
			(0.001)	
Including linear terms	–	–	Yes	Yes
Additional controls	No	Yes	Yes	Yes
Observations	188	188	188	188
Adjusted R^2	0.213	0.446	0.422	0.432

Table 3: Variable importance of selected variables

	Variables	Importance
1	HS GPA	0.50
2	Practice test performance	0.23
3	HS math grade	0.18
4	App performance	0.15
5	Practice tests attempts	0.14
6	Math self-concept	0.09

Figure 5: Interaction between self-testing attempts & Math self-concept



Prediction Rule Ensembles – Importance and Rules

Table 4: Variable importance of selected variables

	Variables	Importance
1	HS GPA	0.50
2	Practice test performance	0.23
3	HS math grade	0.18
4	App performance	0.15
5	Practice tests attempts	0.14
6	Math self-concept	0.09

Figure 6: Interaction between self-testing attempts & HS GPA