Economic Education Interventions and Outcomes

Anxiety, Test-taking Aid, and Test Scores: Evidence from Economics Classes

By COLIN CANNONIER AND MONICA GALLOWAY BURKE[*]

Online Appendix

[*] Cannonier: Economics Department, College of Business, Jack C. Massey College of Business, Belmont University, Nashville, Tennessee 37212, USA (e-mail: colin.cannonier@belmont.edu); Burke: Department of Counseling and Student Affairs, Western Kentucky University, Bowling Green, Kentucky 42101, USA (e-mail: monica.burke@wku.edu).

APPENDIX FIGURES

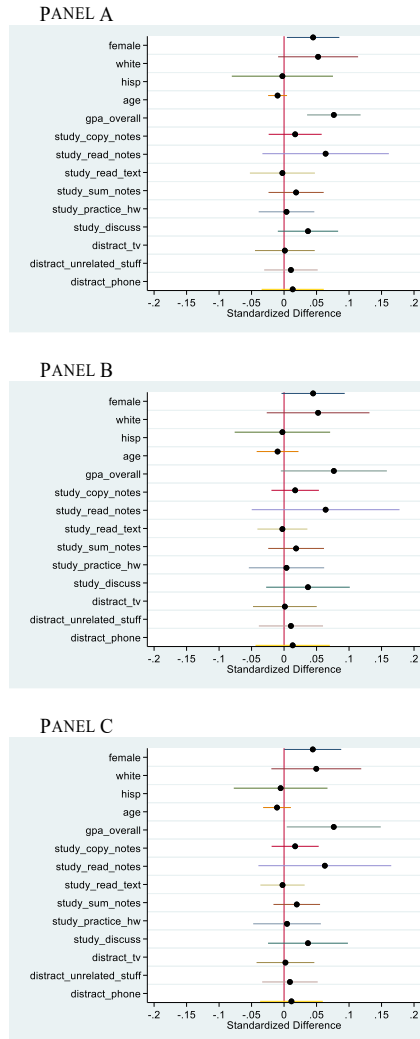PANEL A

PANEL B

PANEL C

FIGURE A1. DIFFERENCES BETWEEN TREATMENT AND CONTROL GROUP

Note: Figure plots the standardized differences between treatment and control groups (Panel A), conditional on clustering at the section level (Panel) and conditional on both clustering and fixed effects for test, year, semester (Panel C).

The plots are based on a test for non-randomness or sample selection in which we compare covariate characteristics that may simultaneously influence both test scores and the likelihood of using test aid. The process involves regressing each attribute on whether the student uses a test aid. We estimate a model of the form $x_i = \phi_0 + \phi_1 TestAid_i + \varepsilon_i$, where x_i is the individual-level attributes and $TestAid_i$ is an indicator for whether a test aid was used and zero otherwise. Next, we obtain the standard estimates from ϕ_1 and then compare the differences in the estimates obtained from a regression model in base form (Panel A); clustering at the section level (Panel B) and conditional on both clustering and the inclusion of fixed effects for test, semester, and year (Panel C).

The standardized differences fall within the range of -0.05 and 0.1 (i.e., between -5% and 10%), which is well within the small-size covariate imbalances across groups (Cohen 1988; Stuart 2010).

PANEL A

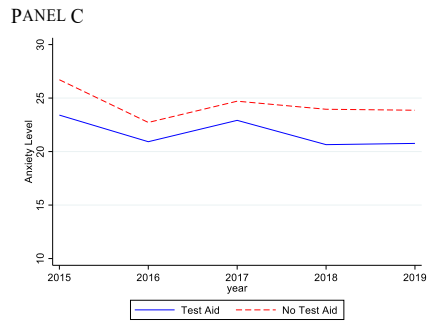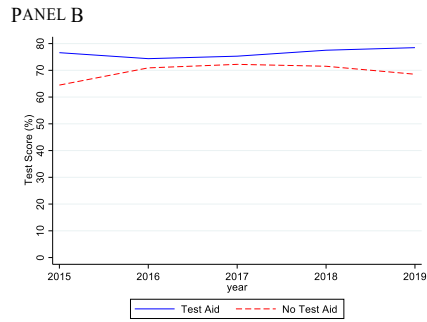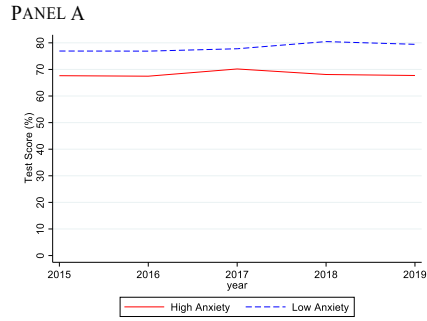PANEL B

PANEL C

FIGURE A2. ANXIETY LEVELS AND TEST SCORES
OVER TIME

Note: Figure plots the average test scores over the
period 2015 to 2019 according to the extent of
anxiety (panel A) and whether test aids are used
during tests (panel B), while panel B plots anxiety
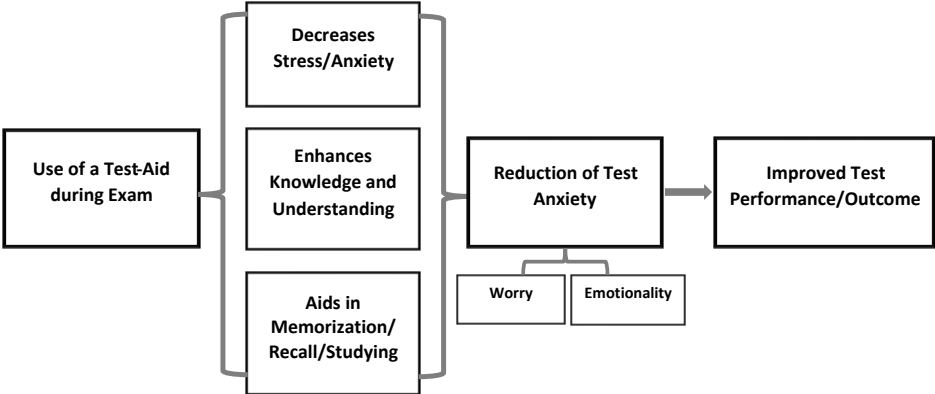levels based whether a test aid is used.

FIGURE A3. CONCEPTUAL FRAMEWORK: TEST AID, TEST ANXIETY , AND TEST PERFORMANCE

APPENDIX TABLES

TABLE A1— SUMMARY STATISTICS OF STUDENT ATTRIBUTES

Variables	Description	Mean	Std. Dev	Minimum	Maximum
Test score (points)	Number of correct answers out of 40 questions	29.568	6.603	7	40
Test score (%)	Student test score (in percent)	73.919	16.508	17.5	100
Standardized Test score	Student test score converted to a Z-score	-0.081	0.991	-3.468	1.485
Anxiety	Anxiety – combined emotionality and worry scales	22.562	8.565	10	50
High Anxiety	Equals one if respondent has anxiety levels higher than average,	0.445	0.497	0	1
No test aid	Equals one if respondent took a particular test using no cheat sheet, zero otherwise	0.387	0.487	0	1
Any test aid	Equals one if respondent took a particular test using no cheat sheet, zero otherwise	0.613	0.487	0	1
Own test aid	Equals one if respondent took a test using an own-prepared cheat sheet, zero otherwise	0.314	0.464	0	1
Group test aid	Equals one if respondent took a test using directed or group-prepared crib sheet, zero otherwise	0.300	0.458	0	1
Demographic variables					
Female	Equals one if respondent is female, zero otherwise	0.603	0.489	0	1
Male	Equals one if respondent is male, zero otherwise	0.397	0.489	0	1
White	Equals one if respondent is White, zero otherwise	0.880	0.325	0	1
Other race	Equals one if respondent is non-White, zero otherwise	0.120	0.325	0	1
Hispanic	Equals one if respondent is Hispanic of any race, zero otherwise	0.069	0.253	0	1
Age	Age of respondent in years	19.203	1.381	17	35
Overall GPA	Overall GPA of the student prior to entering the class	3.395	0.481	1.030	4.000
Other variables					
Freshman	Equals one if respondent is a freshman, 0 otherwise	0.395	0.489	0	1
Sophomore	Equals one if respondent is a sophomore, 0 otherwise	0.493	0.500	0	1
Junior	Equals one if respondent is a junior, 0 otherwise	0.077	0.266	0	1
Senior	Equals one if respondent is a senior, 0 otherwise	0.036	0.185	0	1
Spring	Equals one if spring semester, zero otherwise	0.440	0.496	0	1
Summer	Equals one if spring semester, zero otherwise	0.020	0.141	0	1
Fall	Equals one if spring semester, zero otherwise	0.540	0.498	0	1
Additional Controls					
Copy notes	Equals one if respondent is likely to copy notes in studying for the test, zero otherwise	0.628	0.483	0	1
Read notes	Equals one if respondent is likely to read notes in studying for the test, zero otherwise	0.955	0.206	0	1
Read text	Equals one if respondent is likely to read the text in studying for the test, zero otherwise	0.809	0.393	0	1
Summarize notes	Equals one if respondent is likely to summarize their notes in studying for the test, zero otherwise	0.684	0.465	0	1
Practice homework	Equals one if respondent is likely to redo homework assignments or practice problems in studying for the test, zero otherwise	0.696	0.460	0	1
Discuss with others	Equals one if respondent is likely to discuss the material with a classmate or friend in studying for the test, zero otherwise	0.759	0.428	0	1
Distraction – TV	Equals one if respondent ever watches some TV while studying, zero otherwise	0.241	0.428	0	1
Distraction–unrelated stuff	Equals one if respondent ever talks to others about topics unrelated to the test while studying, zero otherwise	0.644	0.479	0	1
Distraction - phone	Equals one if respondent ever text messages or make phone calls while studying, zero otherwise	0.782	0.413	0	1

Notes: There are 2,357 student-test observations for the period 2015 to 2019.

TABLE A2— SUMMARY STATISTICS OF STUDENT ATTRIBUTES AND MEAN DIFFERENCE BETWEEN CONTROL AND TREATMENT

Variables	Control			Treatment			Mean Difference (p-value)
	Obs.	Mean	Std. Dev	Obs.	Mean	Std. Dev	
Test score (%)	911	70.203	17.433	1,446	76.26	15.451	0.000
Standardized Test score	911	-0.304	1.047	1,446	0.059	0.928	0.000
Demographic variables							
Female	911	0.575	0.495	1,446	0.620	0.485	0.030
White	911	0.866	0.341	1,446	0.889	0.314	0.174
Hispanic	911	0.134	0.341	1,446	0.111	0.314	0.938
Age	911	19.252	1.361	1,446	19.172	1.394	0.807
Overall GPA	911	3.349	0.501	1,446	3.424	0.466	0.033
Other variables							
Freshman	911	0.378	0.485	1,446	0.405	0.491	0.649
Sophomore	911	0.491	0.500	1,446	0.494	0.500	0.596
Junior	911	0.091	0.288	1,446	0.068	0.251	0.235
Additional Controls							
Copy notes	911	0.618	0.486	1,446	0.635	0.482	0.228
Read notes	911	0.948	0.221	1,446	0.960	0.196	0.218
Read text	911	0.810	0.392	1,446	0.808	0.394	0.705
Summarize notes	911	0.674	0.469	1,446	0.691	0.462	0.349
Practice homework	911	0.694	0.461	1,446	0.697	0.460	0.811
Discuss with others	911	0.742	0.438	1,446	0.770	0.421	0.227
Distraction – TV	911	0.240	0.428	1,446	0.241	0.428	0.763
Distraction–unrelated stuff	911	0.638	0.481	1,446	0.648	0.478	0.478
Distraction - phone	911	0.776	0.417	1,446	0.786	0.411	0.475

Notes: The last column shows the p-value from a t-test of the difference in the mean of the treatment and control.

	(1)	(2)	(3)	(4)
Own Test Aid	-0.113***	-0.110***		
	(0.029)	(0.028)		
Group Test Aid	-0.138***	-0.133***		
	(0.028)	(0.027)		
Test Aid			-0.125***	-0.121***
			(0.025)	(0.024)
Test fixed effects	Yes	Yes	Yes	Yes
Semester fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Demographic controls	No	Yes	No	Yes
Other controls	No	Yes	No	Yes
Observations	2,357	2,357	2,357	2,357
R^2	0.037	0.115	0.037	0.115
Wald test: $\theta_{own} = \theta_{group}$	0.340	0.385		

Notes: Results are obtained on estimates from a model of the form:

$$Anxiety_{iecst} = \pi_0 + \pi_1 Own\_TestAid_{iecst} + \pi_2 Group\_TestAid_{iecst} + X_{iecst}\Omega + \kappa_e + \tau_c + \gamma_s + \delta_t + \varepsilon_{iecst}.$$

Observations are at the student $\times$ test level. Robust standard errors are in parentheses and clustered at the section level. The dependent variable is binary equal to one if anxiety level above the mean and equal to zero otherwise Demographic controls comprise gender, age, race, ethnicity, overall GPA while other control variables include various indicators of student study patterns and the distractions encountered when studying. Statistical levels of significance are: *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

We report the average treatment effects of test aid treatments on having high anxiety with and without controls. The respective point estimates (standard errors) for the treatment effects — *Own_TestAid* and *Group_TestAid* —are -0.113(0.029) and -0.138(0.028) without controls and -0.110(0.028) and -0.133(0.027) with controls. The point estimates (standard errors) for the combined treatment effects are -0.125(0.025) without controls and -0.121(0.024) with controls, respectively. In the case of the combined treatment effects estimates, these results imply that using a test aid is associated with a 12-percentage point reduction in the likelihood that a student will have high anxiety. These estimates are consistent with the implied treatment effects in Figure 1 (Panel A).

The table also reports the p-values of Wald tests for the hypotheses that the two test aid treatments have the same effect (i.e., $\theta_{own} = \theta_{group}$).

TABLE A4— IMPACT OF TESTING AID ON TEST PERFORMANCE

	(1)	(2)	(3)	(4)
Own Test Aid	0.313***	0.269***		
	(0.073)	(0.053)		
Group Test Aid	0.427***	0.362***		
	(0.077)	(0.056)		
Test Aid			0.368***	0.315***
			(0.068)	(0.049)
Test fixed effects	Yes	Yes	Yes	Yes
Semester fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Demographic controls	No	Yes	No	Yes
Other controls	No	Yes	No	Yes
Observations	2,357	2,357	2,357	2,357
R^2	0.204	0.425	0.202	0.424
Wald test:$\theta_{own} = \theta_{group}$	0.092	0.057		

Notes: Observations are at the student $\times$ test level. Robust standard errors are in parentheses and clustered at the section level. The dependent variable is the standardized test score. Demographic controls comprise gender, age, race, ethnicity, overall GPA while other control variables include various indicators of student study patterns and the distractions encountered when studying. The table reports the p-values of Wald tests for the hypotheses that the two test aid treatments have the same effect (i.e., $\theta_{own} = \theta_{group}$). Statistical levels of significance are: *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

The table reports the results from a model where treatment effects of test aid are regressed on test scores. The respective point estimates (standard errors) for the treatment effects are 0.313(0.073) and 0.427(0.077) without controls and 0.269(0.053) and 0.362(0.056) with controls. Wald F-tests reject the null hypotheses that the two test aid treatments have the same effect. The point estimates (standard errors) for the combined treatment effects are 0.368(0.068) without controls and 0.315(0.049) with controls, respectively. In the case of the combined model with controls, the coefficient of 0.315 indicates that using a test aid is associated with about a 0.32 standard deviation improvement in test scores. These estimates are consistent with the implied treatment effects in Figure 1 (Panel B).

TABLE A5— CORRELATIONS BETWEEN ANXIETY, TEST AID AND TEST SCORES

	Variables	1	2	3	4	5
1	Test score (%)	1.00				
2	Anxiety (10 to 50)	-0.33***	1.00			
3	Own Test Aid	0.05***	-0.06***	1.00		
4	Group Test Aid	0.14***	-0.10***	-0.44***	1.00	
5	Any Test Aid	0.18***	-0.14***	0.54***	0.52***	1.00

Notes: Statistical levels of significance are: *** significant at the 1 percent level, ** significant at the 5 percent level.

TABLE A6— ANXIETY QUESTIONS

Directions: To the left of each of the following statements, indicate your feelings, attitudes, or thoughts as they are *right now* in relation to this course examination. Use the following numerical scale:

1. The statement does not describe my present condition.

2. The condition is barely noticeable.

3. The condition is moderate.

4. The condition is strong.

5. The condition is very strong; the statement describes my present condition very well.

1. _____ I feel my heart beating fast. (emotionality)
2. _____ I feel regretful. (worry)
3. _____ I am so tense that my stomach is upset. (emotionality)
4. _____ I am afraid that I should have studied more for this test. (worry)
5. _____ I have an uneasy, upset feeling. (emotionality)
6. _____ I feel that others will be disappointed in me. (worry)
7. _____ I am nervous. (emotionality)
8. _____ I feel I may not do as well on this test as I could. (worry)
9. _____ I feel panicky. (emotionality)
10. _____ I do not feel very confident about my performance on this test. (worry)

Adopted from:

Morris, L.W., Davis, M.A., & Hutchings, C.H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology*, *73*(4), 541-555.Variables

Notes: The survey was based on the revised Worry-Emotionality Questionnaire (WEQ) originated with Liebert and Morris (1967) and modified by Morris, Davis, and Hutchings (1981). A copy of the questionnaire is available upon request. Five of these questions form a worry scale and the summation of the remaining five questions constitute an emotional scale. These three items had a scale reliability coefficient above 0.90.

	(1)	(2)	(3)	(4)	(5)	(6)
Own Test Aid	0.269***	0.235***	0.189***			
	(0.053)	(0.051)	(0.056)			
Group Test Aid	0.362***	0.322***	0.239***			
	(0.056)	(0.054)	(0.060)			
Test Aid				0.315***	0.277***	0.213***
				(0.049)	(0.047)	(0.050)
High Anxiety		-0.308***	-0.390***		-0.309***	-0.390***
		(0.039)	(0.064)		(0.039)	(0.064)
High Anxiety x Own Test			0.092			
			(0.094)			
High Anxiety x Group Test			0.185**			
			(0.084)			
High Anxiety x Test Aid						0.135*
						(0.078)
Female	-0.204***	-0.140***	-0.140***	-0.203***	-0.140***	-0.140***
	(0.035)	(0.033)	(0.033)	(0.035)	(0.033)	(0.033)
White	0.030	0.041	0.044	0.031	0.041	0.045
	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)
Hispanic	-0.040	-0.020	-0.021	-0.041	-0.021	-0.022
	(0.072)	(0.071)	(0.071)	(0.072)	(0.071)	(0.071)
Age	0.007	0.008	0.008	0.007	0.008	0.008
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Overall GPA	0.963***	0.900***	0.901***	0.965***	0.901***	0.904***
	(0.040)	(0.037)	(0.037)	(0.040)	(0.037)	(0.037)
Copy notes	-0.150***	-0.139***	-0.140***	-0.149***	-0.138***	-0.140***
	(0.028)	(0.028)	(0.029)	(0.028)	(0.029)	(0.029)
Read notes	-0.073	-0.078	-0.081	-0.071	-0.077	-0.079
	(0.069)	(0.067)	(0.067)	(0.070)	(0.067)	(0.067)
Read text	0.022	0.018	0.020	0.023	0.019	0.020
	(0.045)	(0.044)	(0.045)	(0.045)	(0.044)	(0.044)
Summarize notes	-0.022	-0.025	-0.026	-0.023	-0.026	-0.027
	(0.039)	(0.038)	(0.038)	(0.039)	(0.038)	(0.038)
Practice homework	0.154***	0.145***	0.144***	0.153***	0.144***	0.142***
	(0.036)	(0.034)	(0.034)	(0.036)	(0.034)	(0.034)
Discuss with others	-0.108**	-0.106**	-0.104**	-0.106**	-0.105**	-0.103**
	(0.041)	(0.042)	(0.042)	(0.041)	(0.042)	(0.042)
Distraction – TV	0.010	0.020	0.018	0.008	0.018	0.017
	(0.041)	(0.041)	(0.041)	(0.042)	(0.042)	(0.041)
Distraction – unrelated stuff	-0.068**	-0.049	-0.051	-0.068**	-0.049	-0.051
	(0.034)	(0.035)	(0.035)	(0.034)	(0.035)	(0.035)
Distraction – phone	0.035	0.050	0.053	0.035	0.050	0.052
	(0.044)	(0.042)	(0.041)	(0.044)	(0.042)	(0.041)
Test fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Semester fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,357	2,357	2,357	2,357	2,357	2,357
R^2	0.425	0.446	0.448	0.1424	0.445	0.446
Wald test:$\theta_{own} = \theta_{group}$	0.057	0.080	0.399			

Notes: Observations are at the student × test level. Robust standard errors are in parentheses and clustered at the section level. The dependent variable is the standardized test score. The table reports the p-values of Wald tests for the hypotheses that the two test aid treatments have the same effect (i.e. $\theta_{own} = \theta_{group}$). Statistical levels of significance are: *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

TABLE A8— IMPACT OF TEST AID ON TEST PERFORMANCE, MODELS WITH AND WITHOUT ANXIETY BY ACADEMIC GENDER

	Females (1)	Females (2)	Females (3)	Males (4)	Males (5)	Males (6)
Test Aid	0.296***	0.260***	0.217***	0.347***	0.310***	0.204***
	(0.058)	(0.056)	(0.062)	(0.059)	(0.057)	(0.068)
High Anxiety		-0.270***	-0.320***		-0.357***	-0.518***
		(0.041)	(0.074)		(0.060)	(0.096)
High Anxiety x Test Aid			0.078			0.294**
			(0.088)			(0.120)
Test fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Semester fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,421	1,421	1,421	936	936	936
R^2	0.462	0.478	0.478	0.400	0.427	0.432

Notes: Observations are at the student × test level. Robust standard errors are in parentheses and clustered at the section level. The dependent variable is the standardized test score. Demographic controls comprise gender, age, race, ethnicity, overall GPA while other control variables include various indicators of student study patterns and the distractions encountered when studying. Statistical levels of significance are: *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

TABLE A9— IMPACT OF TEST AID ON TEST PERFORMANCE, MODELS WITH AND WITHOUT ANXIETY BY ACADEMIC CLASSIFICATION

	Freshman (1)	Freshman (2)	Freshman (3)	Sophomore (4)	Sophomore (5)	Sophomore (6)	Junior & Senior (7)	Junior & Senior (8)	Junior & Senior (9)
Test Aid	0.311***	0.255***	0.159**	0.309***	0.285***	0.244***	0.302***	0.271***	0.144
	(0.081)	(0.076)	(0.074)	(0.058)	(0.054)	(0.069)	(0.096)	(0.091)	(0.126)
High Anxiety		-0.392***	-0.518***		-0.248***	-0.302***		-0.251***	-0.376***
		(0.072)	(0.132)		(0.051)	(0.096)		(0.083)	(0.111)
High Anxiety x Test Aid			0.208			0.089			0.218
			(0.161)			(0.114)			(0.161)
Test fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Semester fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	930	930	930	1,162	1,162	1,162	265	265	265
R^2	0.434	0.466	0.469	0.443	0.457	0.457	0.530	0.544	0.547

Notes: Observations are at the student × test level. Robust standard errors are in parentheses and clustered at the section level. The dependent variable is the standardized test score. Demographic controls comprise gender, age, race, ethnicity, overall GPA while other control variables include various indicators of student study patterns and the distractions encountered when studying. Statistical levels of significance are: *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

TABLE A10— IMPACT OF TEST AID ON TEST PERFORMANCE, MODELS WITH AND WITHOUT ANXIETY BY TEST

	Test #1 (1)	Test #1 (2)	Test #1 (3)	Test #2 (4)	Test #2 (5)	Test #2 (6)	Test #3 (7)	Test #3 (8)	Test #3 (9)
Test Aid	0.236***	0.213***	0.134*	0.265***	0.216***	0.258***	0.440***	0.408***	0.246***
	(0.078)	(0.077)	(0.075)	(0.063)	(0.056)	(0.083)	(0.076)	(0.075)	(0.087)
High Anxiety		-0.230***	-0.337***		-0.378***	-0.331***		-0.308***	-0.521***
		(0.069)	(0.097)		(0.067)	(0.118)		(0.058)	(0.071)
High Anxiety x Test Aid			0.206			-0.078			0.325***
			(0.125)			(0.146)			(0.091)
Test fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Semester fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	698	698	698	831	831	831	828	828	828
R^2	0.336	0.351	0.355	0.310	0.346	0.347	0.323	0.347	0.354

Notes: Observations are at the student × test level. Robust standard errors are in parentheses and clustered at the section level. The dependent variable is the standardized test score. Demographic controls comprise gender, age, race, ethnicity, overall GPA while other control variables include various indicators of student study patterns and the distractions encountered when studying. Statistical levels of significance are: *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

TABLE A11— IMPACT OF ANXIETY AND TESTING AID ON TEST PERFORMANCE BY GENDER, ACADEMIC CLASSIFICATION AND TESTS –
MODELS WITH ANXIETY

	Full Sample (1)	Females (2)	Males (3)	Freshman (4)	Sophomore (5)	Junior & Senior (6)	Test #1 (7)	Test #2 (8)	Test #3 (9)
Test Aid	0.213***	0.217***	0.204***	0.159**	0.244***	0.144	0.134*	0.258***	0.246***
	(0.050)	(0.062)	(0.068)	(0.074)	(0.069)	(0.126)	(0.075)	(0.083)	(0.087)
High Anxiety	-0.390***	-0.320***	-0.518***	-0.518***	-0.302***	-0.376***	-0.337***	-0.331***	-0.521***
	(0.064)	(0.074)	(0.096)	(0.132)	(0.096)	(0.111)	(0.097)	(0.118)	(0.071)
High Anxiety x Test Aid	0.135*	0.078	0.294**	0.208	0.089	0.218	0.206	-0.078	0.325***
	(0.078)	(0.088)	(0.120)	(0.161)	(0.114)	(0.161)	(0.125)	(0.146)	(0.091)
Test fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Semester fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,357	1,421	936	930	1,162	265	698	831	828
R^2	0.446	0.478	0.432	0.469	0.457	0.547	0.355	0.347	0.354

Notes: Observations are at the student × test level. Robust standard errors are in parentheses and clustered at the section level. The dependent variable is the standardized test score. Demographic controls comprise gender, age, race, ethnicity, overall GPA while other control variables include various indicators of student study patterns and the distractions encountered when studying. Statistical levels of significance are: *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

APPENDIX NOTES

Notes A1: Experiment Design

Participants

To obtain data for this analysis, we surveyed just over 900 students taking various sections of a course, principles of macroeconomics, at a medium-sized private liberal arts university from fall 2015 to fall 2019, inclusive of summer 2016 and 2017.[1] All students are taught by the same instructor who teaches approximate three sections of the course in the regular semesters of fall and spring. One section of the course is taught in summer. The number of students in each section ranges from 16 to 35. Each semester, students from each section are required to take three in-semester paper-based tests which are identical for each section. Students are taught in a traditional classroom setting two days a week for approximately 75 minutes each day in the fall and spring, while summer students meet for about 105 minutes each day for approximately 5 weeks.

Prior to taking each test, students were asked to voluntary complete a survey intended to gauge their levels of anxiety. The survey was based on the revised Worry-Emotionality Questionnaire (WEQ) originated with Liebert & Morris (1967) and modified by Morris, Davis & Hutchings (1981). From a total of 10 questions (which formed an overall anxiety scale), each with Likert Scale ranging from 1 to 5, we summed five of these questions to form a worry scale and the summation of the remaining five questions constituted an emotional scale. These three scales had a reliability coefficient well above 0.90. See the appendix for a copy of the questionnaire.

As an incentive to providing survey data, students were given the opportunity to use a potential learning aid (referred to as test aid) in the form of a cheat sheet (also referred to as crib sheets in Cannonier and Smith 2019) on two of the three tests. The following restrictions applied for preparing test aids: (1) the student could utilize information on one side of a sheet of paper being no more than 8.5 by 11 inches and (2) all graphs must be drawn by hand; otherwise, all other information on the test aid could be typed or handwritten. For one of the tests, students prepared their own test aid, while for the other test, students prepared a test aid within groups of 5 to 7 students. A random process was used to determine which of the sections would use a test aid for a particular test.

---

[1] All survey information, collected voluntarily and with confidentiality agreements, were approved for study by the university's internal review board (IRB).

Randomization Process

The experiment is based on a cluster-randomization design where random selection of students was carried out at the section-level for each of the three tests done during the semester. That is, in this case, students are assigned to the condition based on the section they are associated with. As stated earlier, each semester, there are three sections of the course being taught with each section consisting of about 25 students on average. Sections were randomly assigned to receive (1) crib sheets prepared individually, (2) crib sheets prepared by group, or (3) control. In the first test, we randomize a list consisting of the three test-aid options (own crib sheet, group crib sheet, no crib sheet). The first test aid option is assigned to the students in the first section; the second test aid option is assigned to students in the second section; and remaining test aid option is assigned to students in the third section. The table below helps to illustrate. Suppose a randomization process generated the test aid options in the following order: no test aid (i.e., control), own test aid and group test aid. For test 1, students in Section 1 will be assigned the control group (no test aid), students in Section 2 will be assigned own test aid and those in Section 3 will be assigned to group test aid. For the second test, a similar randomization process is done on the condition that each section is assigned a test aid different to the one assigned in the previous test and similarly for the third test.

ILLUSTRATION OF CLUSTER RANDOMIZATION PROCESS

Test #	Section 1	Section 2	Section 3
Test #1	None	Own Test aid	Group Test aid
Test #2	Group Test aid	None	Own Test aid
Test #3	Own Test aid	Group Test aid	None

For the sample period, there are a total of 90 sections with 60 assigned to treatment and 30 to the control. There are several advantages to using this cluster randomization procedure. First, cluster random designs have been known to possess greater external validity. Second, these experimental approaches are suitable in educational settings such as ours where the provision of test aids to students in the treatment group and providing no test aid to participants in a control group can run the risk of the test aids being shared with the control group during a test. Third, a unique feature of this experiment is that tests are taken by each of the sections on the same day with the important distinction that all test aids are collected from students following the test. This

mitigates against the possibility that later-day test takers in the control group are able to access test aids from their peers who used test aids in a different section earlier in the day. Fourth, this cluster randomization design is more cost effective as it relates to the ease of implementation and the efficiencies associated with administration. Finally, this study which is large in scale with multiple sections is likely to produce more precise estimates and correct statistical inferences when using clustered standard errors.[2]

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (FALL 2015)

Test #	Section 3	Section 4	Section 5	Section 7
Test #1	Own Test aid	None	Own Test aid	None
Test #2	None	Own Test aid	None	Own Test aid
Test #3	Group Test aid	Group Test aid	Group Test aid	Group Test aid

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (SPRING 2016)

Test #	Section 3	Section 51	Section 52
Test #1	None	Own Test aid	Group Test aid
Test #2	Group Test aid	None	Own Test aid
Test #3	Own Test aid	Group Test aid	None

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (SUMMER2016)

Test #	Section 1
Test #1	None
Test #2	Group Test aid
Test #3	Own Test aid

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (FALL 2016)

Test #	Section 5	Section 6	Section 7
Test #1	None	Own Test aid	Group Test aid
Test #2	Group Test aid	None	Own Test aid
Test #3	Own Test aid	Group Test aid	None

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (SPRING 2017)

Test #	Section 50L	Section 51L	Section 52L
Test #1	None	Own Test aid	Group Test aid
Test #2	Group Test aid	None	Own Test aid
Test #3	Own Test aid	Group Test aid	None

[2] See Schochet (2008) for a detailed examination of the theoretical and empirical issues associated with impact evaluation of education interventions using clustering randomized experiments.

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (SUMMER2017)

Test #	Section 1
Test #1	None
Test #2	Group Test aid
Test #3	Own Test aid

Notes: For test #2 and test #3, some students used own test aid and group test aid, respectively.

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (FALL 2017)

Test #	Section 6	Section 7	Section 51L	Section 52L
Test #1	Group Test aid	None	None	Own Test aid
Test #2	Own Test aid	Own Test aid	Group Test aid	Group Test aid
Test #3	None	Group Test aid	Own Test aid	None

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (SPRING 2018)

Test #	Section 6	Section 50L	Section 51L
Test #1	None	Own Test aid	Group Test aid
Test #2	Group Test aid	None	Own Test aid
Test #3	Own Test aid	Group Test aid	None

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (FALL 2018)

Test #	Section 7	Section 8	Section 51L
Test #1	Group Test aid	None	Own Test aid
Test #2	Own Test aid	Group Test aid	None
Test #3	None	Own Test aid	Group Test aid

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (SPRING 2019)

Test #	Section 4	Section 50L	Section 51L
Test #1	Own Test aid	Group Test aid	None
Test #2	None	Own Test aid	Group Test aid
Test #3	Group Test aid	None	Own Test aid

ILLUSTRATION OF QUASI-RANDOMIZATION PROCESS (FALL 2019)

Test #	Section 7	Section 51L
Test #1	Own Test aid	None
Test #2	None	Group Test aid
Test #3	Group Test aid	Own Test aid

Notes:

APPENDIX REFERENCES

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. New York: Routledge.

Liebert, Robert M., and Larry W. Morris. 1967. "Cognitive and Emotional Components of Test Anxiety: A Distinction and Some Initial Data." *Psychological Reports* 20 (3): 975-978.

Morris, Larry W., Mark A. Davis, and Calvin H. Hutchings. 1981. "Cognitive and Emotional Components of Anxiety: Literature Review and a Revised Worry–Emotionality Scale." *Journal of Educational Psychology* 73 (4): 541–555.

Schochet, Peter Z. 2008. "Statistical power for random assignment evaluations of education programs." *Journal of Educational and Behavioral Statistic* 33 (1): 62–87.

Stuart, Elizabeth A. 2010. "Matching methods for causal inference: a review and a look forward." *Stat. Sci.* 25 (1): 1–21.