

Assessing Financial Education Methods: Principles vs. Rules-of-Thumb Approaches

Evan Davies, Brian Mitchell, Dave Mun, & Bill Skimmyhorn

Department of Social Sciences, United States Military Academy

April 10, 2015

The views expressed herein are those of the authors and do not represent the U.S. Military Academy, the Department of the Army, or the Department of Defense.

Outline

- Motivation
- Background
- Experimental Design
- Data
- Empirical strategy
- Results
- Discussion

Motivation I: Current literature

- Challenges within the traditional financial education literature:
 - Typically lack “good” evaluations (Miller, Reichelstein, Salas and Zia 2014)
 - Little use of comprehensive pre/post assessments (Lusardi and Mitchell, 2009)
 - Frequently small sample sizes among those with pre/post assessments (Boarden, Lee, Serido, and Collins 2008; Bowen and Jones 2006)
- Some optimism given more recent experimental work using different methods (Lusardi et al. 2014; Heinberg et al. 2010)
- Uncertainty remains over the optimal use of behavioral economic insights in designing teaching methods (Willis 2011, Drexler, Fischer and Schoar 2014)

Motivation II: Our research

- Research Questions:
 - What is the general effectiveness of financial education?
 - What is the relative effectiveness of different teaching methods ?
- Our contribution:
 - A **large field experiment** with relatively **comprehensive measures** of effectiveness for a variety of outcomes using a **pre/post assessment** in a setting of wide interest
- Main Findings:
 - Both the principles and rules-of-thumb methods work
 - With some differences in the relative effectiveness of each
 - And few differential treatment effects

Background: The West Point Program

- USMA (n=4500 students) provides comprehensive academic, military, physical and character development
- Graduates serve as officers in the Active Duty Army
- The academic program includes a robust, STEM-intensive core curriculum including an engineering sequence and small classes (sample mean = 15 students / class)
- Two required courses supported our research design:
 - American Politics
 - Principles of Economics (with 4 x 2 hour Personal Finance Labs)

Personal Finance Lab Topics

- **Lab 1:** Personal Finance for Servicemembers
- **Lab 2:** Personal Finance Basics/Major Financial Decisions
- **Lab 3:** Investing for your future
- **Lab 4:** Retirement Planning and Insurance

Experimental Design: Two Teaching Methodologies

Treatment Group 1
Principles Based (PB)
N=289

Traditional

Often implemented
with MS Excel

Treatment Group 2
Rules-of-Thumb (ROT)
N=280

Basic Heuristics

Often implemented
with online resources

Concepts
Online Resources
Personal Finance Exercise

Examples

Take Home Pay and Budget

Calculate take-home pay
Allocate all money

Find take-home pay from paystub
ROT1: 20/30/50 Rule

Time Value of Money

TVM Equations / Calculations

Online calculators

*** Control Group: Students exclusively enrolled in American Politics N=422**

Data

1. Administrative data on student characteristics (age, gender, race, SAT Math, prior military service, 1st year GPA)
2. Outcome data gathered using mandatory *Pre* and *Post* course assessments
 - A. **N=994 (Pre)** and **N=986 (Post)** completed (**99.2%**)
 - B. **Outcomes** (See App. Table 1 for items):
 1. Objective knowledge measures (topical, Big 5, self-assessed)
 2. Self-reported behavioral measures (self-efficacy, motivation, likeliness to seek advice, risk preference, patience)
 3. New problem solving (3 items, collected in *Post* assessment only)

Note: 10 point Likert-Scale responses converted to binary measures of “High” (for answers ≥ 7) to support OLS regressions

Summary Statistics

Table 2. Summary Statistics

	Full	Control	Treatment Group:						
	Sample	Group	Combined		Principles (PB)		Rules of Thumb (ROT)		
	Mean (SD) (1)	Mean (SD) (2)	Mean (SD) (3)	Diff. from Control (4)	Mean (SD) (5)	Diff. from Control (6)	Mean (SD) (7)	Diff. from Control (8)	Diff. from PB (9)
Panel A. Individual Characteristics									
Age	18.60 (1.11)	18.57 (1.05)	18.63 (1.16)	0.06 [0.37]	18.66 (1.15)	0.09 [0.29]	18.60 (1.17)	0.04 [0.67]	-0.05 [0.588]
Female	0.15 (0.36)	0.14 (0.35)	0.16 (0.37)	0.02 [0.31]	0.15 (0.35)	0.01 [0.79]	0.18 (0.38)	0.04 [0.16]	0.03 [0.291]
Black	0.08 (0.28)	0.10 (0.29)	0.08 (0.26)	-0.02 [0.27]	0.08 (0.27)	-0.02 [0.36]	0.08 (0.26)	-0.02 [0.33]	-0.001 [0.950]
Hispanic	0.12 (0.32)	0.11 (0.31)	0.12 (0.33)	0.01 [0.50]	0.15 (0.35)	0.04 [0.14]	0.10 (0.30)	-0.01 [0.63]	-0.05 [0.071]
Other Race	0.11 (0.31)	0.12 (0.32)	0.10 (0.30)	-0.02 [0.40]	0.08 (0.28)	-0.03 [0.14]	0.12 (0.32)	0.001 [0.98]	0.03 [0.173]
White	0.69 (0.46)	0.68 (0.47)	0.70 (0.46)	0.02 [0.44]	0.69 (0.46)	0.02 [0.67]	0.71 (0.45)	0.03 [0.38]	0.02 [0.672]
SAT Score	1317 (124.97)	1304 (122.08)	1326 (126.33)	22.32 [0.01]	1318 (120.89)	14.12 [0.13]	1335 (131.37)	30.74 [0.002]	16.62 [0.117]
Prior Enlisted	0.21 (0.41)	0.22 (0.41)	0.20 (0.40)	-0.01 [0.62]	0.20 (0.40)	-0.01 [0.74]	0.20 (0.40)	-0.02 [0.62]	-0.005 [0.886]
First year GPA	2.98 (0.59)	2.92 (0.58)	3.03 (0.59)	0.11 [0.003]	2.97 (0.53)	0.05 [0.21]	3.09 (0.63)	0.17 [0.0003]	0.12 [0.014]
Observations	991	422	569		289		280		
Classes (Sections)	73	35	38		19		19		
Instructors	24	11	13		12		13		

Empirical Strategy: Difference-in-Differences (DD)

- Identification of a causal estimate (β_3) requires parallel trends

$$Y_i = \beta_0 + \beta_1 Post_i + \beta_2 T_i + \beta_3 Post_i \times T_i + X_i' \gamma + \delta_s + \varepsilon_i$$

- We provide evidence suggesting random assignment:
 - Institutional details suggest:
 - Random assignment of students to each course
 - Random assignment of sections to treatment method (balanced by instructor, Appendix Table 2)
 - t-tests of means reveals few differences (Table 2 Panel A)
 - Covariate regressions suggest that other potential determinants of the outcomes are unrelated to treatment condition (Table 2 Panel B)

Experimental Validity

Table 2. Summary Statistics

	Full	Control	Treatment Group:						
	Sample	Group	Combined		Principles (PB)		Rules of Thumb (ROT)		
	Mean	Mean	Mean	Diff. from	Mean	Diff. from	Mean	Diff. from	Diff. from
	(SD)	(SD)	(SD)	Control	(SD)	Control	(SD)	Control	PB
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
Panel A. Individual Characteristics									
Age	18.60 (1.11)	18.57 (1.05)	18.63 (1.16)	0.06 [0.37]	18.66 (1.15)	0.09 [0.29]	18.60 (1.17)	0.04 [0.67]	-0.05 [0.588]
Female	0.15 (0.36)	0.14 (0.35)	0.16 (0.37)	0.02 [0.31]	0.15 (0.35)	0.01 [0.79]	0.18 (0.38)	0.04 [0.16]	0.03 [0.291]
Black	0.08 (0.28)	0.10 (0.29)	0.08 (0.26)	-0.02 [0.27]	0.08 (0.27)	-0.02 [0.36]	0.08 (0.26)	-0.02 [0.33]	-0.001 [0.950]
Hispanic	0.12 (0.32)	0.11 (0.31)	0.12 (0.33)	0.01 [0.50]	0.15 (0.35)	0.04 [0.14]	0.10 (0.30)	-0.01 [0.63]	-0.05 [0.071]
Other Race	0.11 (0.31)	0.12 (0.32)	0.10 (0.30)	-0.02 [0.40]	0.08 (0.28)	-0.03 [0.14]	0.12 (0.32)	0.001 [0.98]	0.03 [0.173]
White	0.69 (0.46)	0.68 (0.47)	0.70 (0.46)	0.02 [0.44]	0.69 (0.46)	0.02 [0.67]	0.71 (0.45)	0.03 [0.38]	0.02 [0.672]
SAT Score	1317 (124.97)	1304 (122.08)	1326 (126.33)	22.32 [0.01]	1318 (120.89)	14.12 [0.13]	1335 (131.37)	30.74 [0.002]	16.62 [0.117]
Prior Enlisted	0.21 (0.41)	0.22 (0.41)	0.20 (0.40)	-0.01 [0.62]	0.20 (0.40)	-0.01 [0.74]	0.20 (0.40)	-0.02 [0.62]	-0.005 [0.886]
First year GPA	2.98 (0.59)	2.92 (0.58)	3.03 (0.59)	0.11 [0.003]	2.97 (0.53)	0.05 [0.21]	3.09 (0.63)	0.17 [0.0003]	0.12 [0.014]
Observations	991	422	569		289		280		
Classes (Sections)	73	35	38		19		19		
Instructors	24	11	13		12		13		
Panel B. Covariate Regression Results									
Partial R2 for Individ. Char.				0.017			0.016	0.024	0.020
p-value for F-Test of Joint Sig. of Individ. Char.				0.511			0.862	0.249	0.255
Observations				986			706	698	568

Empirical Strategy: Difference-in-Differences (DD)

- Recall our estimating equation:

$$Y_i = \beta_0 + \beta_1 Post_i + \beta_2 T_i + \beta_3 Post_i \times T_i + X_i' \gamma + \delta_s + \varepsilon_i$$

Age, Gender, Race
SAT Math, Prior Mil
Service, 1st Yr GPA

Hour
Fixed
Effects

- We report only the main DD estimates (β_3) for outcomes 1-8 and a binary treatment indicator for outcome 9
- We complete four comparisons:
 - Treatment group (Combined) vs. Control group
 - Principles-based group vs. Control group
 - Rules-of-thumb group vs. Control group
 - Rules-of-thumb group vs. Principles-based group
- We cluster std errors at the instructor level (N=24)

Main Effects: Combined Methods vs. Control Group

Table 4. OLS Estimates of Main Program Effects

Variable	Outcomes								
	(1) Topical Knowledge	(2) Big 5 Knowledge	(3) Self-Assessed Knowledge	(4) Self-Efficacy	(5) Motivation to Learn	(6) Likeliness to Seek Advice	(7) Self-Assessed Risk Pref	(8) Patience	(9) New Problem Solving
Panel A. Combined Treatment vs. Control									
Control Mean	0.5894	0.7153	0.2165	0.5251	0.7512	0.8373	0.4486	0.3973	0.4234
PostxT	0.0907*** (0.0111)	0.0566*** (0.0156)	0.2463*** (0.0267)	0.1543*** (0.0282)	0.0840** (0.0322)	-0.0283 (0.0329)	0.1454*** (0.0362)	0.0153 (0.0147)	T 0.1568*** (0.0210)
R2	0.3019	0.1298	0.0711	0.0431	0.0500	0.0233	0.0633	0.0217	0.2208
Obs	1972	1972	1972	1972	1972	1972	1972	1972	986
	15.4%	7.9%	113.8%	29.4%	11.2%	-3.3%	32.4%	3.9%	37.4%

- Large and statistically significant effects for 7/9 outcomes
- Over-confidence concerns?
 - *Topical / Big 5 vs. Self-Assessed Knowledge differences*

Main Effects: PB Method vs. Control Group

Table 4. OLS Estimates of Main Program Effects

Variable	Outcomes								
	(1) Topical Knowledge	(2) Big 5 Knowledge	(3) Self-Assessed Knowledge	(4) Self-Efficacy	(5) Motivation to Learn	(6) Likeliness to Seek Advice	(7) Self-Assessed Risk Pref	(8) Patience	(9) New Problem Solving
Panel B. PB Method vs. Control									
Control Mean	0.5894	0.7153	0.2165	0.5251	0.7512	0.8373	0.4486	0.3973	0.4234
PostxPB	0.0917*** (0.0126)	0.0521*** (0.0184)	0.2589*** (0.0297)	0.1908*** (0.0320)	0.0821** (0.0342)	0.0052 (0.0322)	0.1538*** (0.0359)	0.0409** (0.0170)	PB 0.1474*** (0.0226)
R2	0.2840	0.1435	0.0680	0.0522	0.0616	0.0225	0.0715	0.0218	0.2247
Obs	1412	1412	1412	1412	1412	1412	1412	1412	706
	15.6%	7.3%	119.6%	36.3%	10.9%	0.6%	34.3%	10.3%	34.8%

- Large and statistically significant effects for 8/9 outcomes
- *Increased Patience* (portion of hypothetical loan allocated to long-term goals)

Main Effects: ROT Method vs. Control Group

Table 4. OLS Estimates of Main Program Effects

Variable	Outcomes									
	(1) Topical Knowledge	(2) Big 5 Knowledge	(3) Self-Assessed Knowledge	(4) Self-Efficacy	(5) Motivation to Learn	(6) Likeliness to Seek Advice	(7) Self-Assessed Risk Pref	(8) Patience	(9) New Problem Solving	
Panel C. ROT Method vs. Control										
Control Mean	0.5894	0.7153	0.2165	0.5251	0.7512	0.8373	0.4486	0.3973		0.4234
PostxROT	0.0897*** (0.0121)	0.0613*** (0.0163)	0.2334*** (0.0299)	0.1168*** (0.0356)	0.0860** (0.0328)	-0.0630 (0.0372)	0.1368** (0.0557)	-0.0109 (0.0181)	ROT	0.1625*** (0.0340)
R2	0.2963	0.1290	0.0514	0.0387	0.0441	0.0280	0.0589	0.0212		0.2177
Obs	1396	1396	1396	1396	1396	1396	1396	1396		698
	15.2%	8.6%	107.8%	22.2%	11.5%	-7.5%	30.5%	-2.5%		38.4%

- Large and statistically significant effects for 7/9 outcomes
- Results suggest that this method may decrease the *Likeliness to Seek Advice* ($p=0.105$)

Main Effects: ROT Method vs. PB Method

Table 4. OLS Estimates of Main Program Effects

Variable	Outcomes									
	(1) Topical Knowledge	(2) Big 5 Knowledge	(3) Self-Assessed Knowledge	(4) Self-Efficacy	(5) Motivation to Learn	(6) Likeliness to Seek Advice	(7) Self-Assessed Risk Pref	(8) Patience	(9) New Problem Solving	
Panel D. ROT Method vs. PB Method										
PB Mean	0.6775	0.7594	0.3038	0.5799	0.8455	0.9167	0.4670	0.3762		0.3022
PostxROT	-0.0020 (0.0110)	0.0091 (0.0154)	-0.0254 (0.0267)	-0.0740* (0.0378)	0.0038 (0.0188)	-0.0682*** (0.0216)	-0.016 (0.0599)	-0.0518** (0.0198)	ROT	0.0087 (0.0392)
R2	0.2777	0.1137	0.0955	0.0480	0.0412	0.0209	0.0734	0.0319		0.1750
	-0.3%	1.2%	-8.2%	-12.8%	0.5%	-7.4%	-3.4%	-13.6%		2.9%

- PB and ROT comparably effective for 6/9 outcomes
- But the ROT method appears to:
 - Generate less *Self-Efficacy*
 - Reduce *Likeliness to Seek Advice*
 - Reduce a measure of *Patience*

Differential Effects: Motivation

- We analyze the treatment effects (PB vs. ROT) within four groups:
 1. Females
 2. Low quantitative ability (SAT Math ≤ 600)
 3. Low initial knowledge (*Pre-Score* $\leq 55\%$)
 4. Low initial motivation to learn (*Pre-Score* ≤ 6)
- And focus on the PB vs. ROT comparison

Differential Effects: Results

Table 5. OLS Estimates of Heterogeneous Program Effects

Variable	Outcomes								
	(1) Topical Knowledge	(2) Big 5 Knowledge	(3) Self-Assessed Knowledge	(4) Self-Efficacy	(5) Motivation to Learn	(6) Likeliness to Seek Advice	(7) Self-Assessed Risk Pref	(8) Patience	(9) New Problem Solving
Panel A: Females									
Control Mean	0.6345	0.6976	0.0952	0.5238	0.6310	0.9643	0.1905	0.3640	0.4444
PostxROT	-0.018 (0.0215)	-0.038 (0.0406)	0.0323 (0.0616)	-0.022 (0.1342)	-0.199 (0.1303)	-0.103 (0.0691)	-0.002 (0.0777)	-0.023 (0.0490)	ROT 0.0658 (0.0456)
R2	0.3221	0.2557	0.1190	0.1437	0.0825	0.0509	0.1736	0.0625	0.2602
Obs	184	184	184	184	184	184	184	184	92

- Very similar results ... and encouraging evidence for PB

Robustness Checks

- DD results may reflect control group outcome declines:
 - N/A to PB vs. ROT analysis
 - Comparable motivations (mandatory but ungraded)
 - Motivation/attention outcomes suggest no differential effects (See Appendix Table 3)
- Cluster-Wild Bootstrap results alleviates concerns that our small number of clusters ($N < 30$) might produce downward biased standard errors. (Appendix Table 4A)
- Alternate specifications support main findings:
 - Binary Treatment vs. DD (Appendix Table 4B)
 - Logit marginal effects (Appendix Table 4C)
 - Alternate Likert-scale outcome thresholds (Appendix Table 5)

Discussion I

- Both teaching methods prove effective in this sample:
 - Cognitive measures : 7-16% knowledge effects
 - Non-cognitive measures: 22-36% self-efficacy effects
- Few differences in relative effectiveness of teaching methods, though PB weakly dominates ROT
- Limited heterogeneous treatment effects, but none suggesting beneficial effects for ROT over PB
 - ROT is slightly less effective for: female students
 - ROT is slightly less effective for individuals with: low quantitative abilities and low initial knowledge scores

Discussion II

- Internal validity concerns (e.g., student absences, sharing of course materials/knowledge, “John Henry” effects) suggest that our estimates may be lower bounds for cadets
- External validity concerns (e.g., high human capital, certainty of low-interest \$40k loan, and professional motivations) suggest that our estimates are likely upper bounds for other undergraduates
- Future work might include:
 - Analyzing these teaching method effects on behavioral outcomes
 - Evaluating PB and ROT methods in other school/workplace settings

william.skimmyhorn@usma.edu

david.mun@usma.edu

BACKUP SLIDES

Results: Summary Statistics

Table 3. Financial Outcome Summary Statistics

Outcome	Description	Panel A. Control		Panel B. PB Treatment		Panel C. ROT Treatment	
		Pre Mean (SD)	Post Mean (SD)	Pre Mean (SD)	Post Mean (SD)	Pre Mean (SD)	Post Mean (SD)
1	Topical Knowledge, %	59.93 (14.84)	57.95 (17.14)	64.24 (12.06)	71.35 (12.63)	64.34 (11.63)	71.34 (13.11)
2	Big 5 Knowledge, %	73.06 (21.20)	70.00 (22.47)	74.40 (16.65)	77.01 (15.78)	73.93 (16.90)	77.00 (17.17)
3	Pr(Self-Assessed Knowledge \geq 7), %	22.97 (42.11)	20.33 (40.30)	19.72 (39.10)	41.73 (49.44)	20.71 (40.60)	41.43 (49.35)
4	Pr (Self-Efficacy \geq 7), %	54.07 (49.89)	50.96 (50.05)	51.76 (50.09)	64.08 (47.46)	53.57 (49.96)	62.14 (48.59)
5	Pr (Motivation to Learn \geq 7), %	78.71 (40.99)	71.53 (45.18)	82.57 (36.70)	83.80 (35.70)	81.07 (39.24)	82.50 (38.06)
6	Pr(Likeliness to Seek Advice \geq 7), %	84.69 (36.05)	82.78 (37.80)	93.31 (26.61)	88.56 (28.71)	94.29 (23.25)	86.07 (34.69)
7	Pr(Willingness to Take Risk \geq 7), %	44.74 (49.78)	44.98 (49.81)	38.20 (48.83)	52.99 (49.88)	37.50 (48.50)	51.43 (50.07)
8	Loan Allocation to Long Term Savings, %	38.97 (24.77)	40.49 (23.50)	37.20 (22.09)	40.25 (20.79)	39.64 (24.12)	40.07 (20.76)
9	New Problem Solving, %	-	42.34 (32.09)	-	59.98 (30.22)	-	61.19 (30.82)

Note: DOD data. N=986. Outcomes described in Section 2. The final outcome (new knowledge) was only collected during the post assessment.

Outcomes I: Knowledge

1. Topical Knowledge

- a) As you approach retirement, your investments should become _____? (*Less Risky*)
- b) A budget is important for all of the following reasons (*Both A [Spend less than you earn] and C [Track your expenses over time]*)
 - a) Do you think that the following statement is true or false: Buying a single company stock usually provides a safer return than a stock mutual fund? (*False*)
 - b) Suppose you had \$100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow? (*More than \$102*)

2. Self-Assessed Knowledge

How would you assess your overall financial knowledge? (*1=Very Low & 10=Very High*)

Outcomes II: Self-reported behavioral measures

4. Self-Efficacy

I am good at dealing with day-to-day financial matters, such as checking accounts, credit and debit cards, and tracking expenses. *(1=Strongly Disagree & 10=Strongly Agree)*

5. Motivation to Learn

I am motivated to learn about personal finance topics on my own. *(1=Strongly Disagree & 10=Strongly Agree)*

6. Likelihood to Seek Advice

When facing an important financial decision, how likely are you to seek assistance or advice? *(1=Not At All Likely & 10=Very Likely)*

Outcomes III: Self-reported behavioral measures

7. Willingness to Take Risk

When thinking of financial investments, how willing are you to take risks? (*1=Not At All Willing & 10=Very Willing*)

8. Time Preference (Patience)

Assume that you have just received \$40,000 for your Cow Loan and that you have no other debts. Write the amount that you would allocate to each option in the space provided:

- ***Present consumption*** (0-6 months after receipt of loan), such as spring break, car, gifts, and clothing
- ***Short-term savings*** (6 - 18 months after receipt of loan), such as class ring, uniforms, and furnishings;
- ***Medium-term savings*** (within 10 yrs of graduation), such as wedding, real estate, and graduate school tuition;
- ***Long-term savings*** (more than 10 yrs in the future), such as children's education and retirement.

Outcomes IV: New Knowledge (*Post* only)

9. New Problem Solving

One of your Soldiers asks your advice regarding what he should do with \$1,000 he recently inherited. The Soldier has \$2,500 in credit card debt with an APR of 18%, a \$5,000 car loan with an APR of 6%, \$500 in pay-day loan debt with an APR of 260%, and a \$10,000 loan from his credit union at 7%.

- How much money from his inheritance do you recommend he allocate to each type of debt? (*\$500 to Payday Loan, \$500 to Credit Card*)

Experimental Design: Instructor Assignments

Appendix Table 2. Teaching Methods by Time and Instructor

Instructor	Building	Course Time			
		PM1	PM2	AM1	AM2
1	A	-	-	ROT	PB
2	B	ROT	ROT	PB	PB
3	B	PB	PB	ROT	ROT
4	B	PB	ROT	ROT	PB
5	B	ROT	PB	ROT	PB
6	B	PB	ROT	ROT	PB
7	B	PB	PB	ROT	ROT
8	B	ROT	ROT	PB	PB
9	B	PB	ROT	-	-
10	B	ROT	PB	-	-
11	A	-	-	-	PB
12	A	-	-	PB	ROT
13	B	-	-	-	ROT
PB Total		5	4	3	7
ROT Total		4	5	6	4
Total		9	9	9	11

Note: We describe the PB and ROT methods in Section 2.

Robustness I: Effects on Student Attention

Appendix Table 3. OLS Estimates of Program Effects on Student Motivation/Attention

Variable	Outcomes				
	(1) Overall Attention, %	(2) Academy Mascots (Q7)	(3) Algebra Problem (Q15)	(4) Word Puzzle (Q35)	(5) West Point Trivia (Q47)
Panel A. Combined Treatment vs. Control					
Control Mean	0.9031	0.9139	0.9342	0.8170	0.9474
PostxPB	0.0118 (0.0114)	0.0183 (0.0190)	0.0124 (0.0137)	0.0187 (0.0239)	-0.0020 (0.0189)
R2	0.0528	0.0211	0.0299	0.0258	0.0186
Obs	1972	1972	1972	1972	1972
Panel B. PB Method vs. Control					
Control Mean	0.9031	0.9139	0.9342	0.8170	0.9474
PostxPB	0.0138 (0.0145)	0.0199 (0.0220)	0.0037 (0.0135)	0.0358 (0.0316)	-0.0039 (0.0236)
R2	0.0586	0.0253	0.0427	0.0261	0.0181
Obs	1412	1412	1412	1412	1412
Panel C. ROT Method vs. Control					
Control Mean	0.9031	0.9139	0.9342	0.8170	0.9474
PostxROT	0.0098 (0.0144)	0.0167 (0.0207)	0.0214 (0.0170)	0.0011 (0.0286)	-0.0001 (0.0245)
R2	0.0576	0.0199	0.0282	0.0375	0.0241
Obs	1396	1396	1396	1396	1396
Panel D. ROT Method vs. PB Method					
Control Mean	0.9366	0.9566	0.9740	0.8628	0.9531
PostxROT	-0.004 (0.0180)	-0.0032 (0.0200)	0.0177 (0.0138)	-0.0346 (0.0378)	0.0038 (0.0301)
R2	0.0307	0.0246	0.0148	0.0171	0.0240
Obs	1136	1136	1136	1136	1136

Robustness I: Alternate standard error computations

Appendix Table 1. Alternative Estimates of Program Effects

Variable	Outcomes								
	(1) Topical Knowledge	(2) Big 5 Knowledge	(3) Self-Assessed Knowledge	(4) Self-Efficacy	(5) Motivation to Learn	(6) Likeliness to Seek Advice	(7) Self-Assessed Risk Pref	(8) Patience	(9) New Problem Solving
Control Mean	0.5894	0.7153	0.2165	0.5251	0.7512	0.8373	0.4486	0.1568	0.4234
Panel A. Main Estimates									
PostxPB	0.0917***	0.0521***	0.2589***	0.1908***	0.0821**	0.0052	0.1538***	0.0409**	0.1528***
	(0.0125)	(0.0184)	(0.0296)	(0.0319)	(0.0341)	(0.0321)	(0.0358)	(0.0170)	(0.0216)
Cluster SE p-value	0.000	0.009	0.000	0.000	0.025	0.872	0.000	0.024	0.000
Wild Bootstrap SE p-value	0.000	0.009	0.000	0.000	0.026	0.861	0.000	0.031	0.000
PostxROT	0.0897***	0.0613***	0.2334***	0.1168***	0.0860**	-0.063	0.1368**	-0.010	0.1611***
	(0.0121)	(0.0163)	(0.0298)	(0.0355)	(0.0327)	(0.0372)	(0.0556)	(0.0181)	(0.0339)
Cluster SE p-value	0.000	0.001	0.000	0.003	0.015	0.103	0.022	0.552	0.000
Wild Bootstrap SE p-value	0.000	0.001	0.000	0.005	0.016	0.124	0.024	0.544	0.001
R2	0.3019	0.1298	0.0711	0.0431	0.0500	0.0233	0.0633	0.0217	0.2208
Obs	1972	1972	1972	1972	1972	1972	1972	1972	986
Clusters	24	24	24	24	24	24	24	24	24
PostxPB - PostxROT	0.0020	-0.0092	0.0255	0.0740	-0.0039	0.0682	0.0170	0.0509	-0.0083
p-value	0.8508	0.5492	0.3392	0.0575	0.8358	0.0037	0.7749	0.5944	0.8290

Note: DoD Data. The table reports several robustness checks for the main DD estimates using different functional forms. All regressions (except Panel B) include the following covariates: age, gender indicator, race indicators, total SAT score, prior service indicator, first year academic performance score and section fixed effects. Heteroskedasticity robust standard errors, clustered at the instructor level, are depicted in parentheses. Column 9 reports OLS estimates of the new knowledge outcome that was only included on the final assessment. The bottom panel reports the differences in the two method coefficients and the p-values for a test of coefficient equality. ***, **, and * reflect p<0.01, 0.05, and 0.10 respectively.

Robustness Ib: Alternate standard error computations

Appendix Table 4A. Alternate Standard Error Estimates for Main Program Effects

Variable	Outcomes								
	(1) Topical Knowledge	(2) Big 5 Knowledge	(3) Self-Assessed Knowledge	(4) Self-Efficacy	(5) Motivation to Learn	(6) Likeliness to Seek Advice	(7) Self-Assessed Risk Pref	(8) Patience	(9) New Problem Solving
Panel A. Combined Treatment vs. Control									
Control Mean	0.5894	0.7153	0.2165	0.5251	0.7512	0.8373	0.4486	0.3973	0.4234
PostxT	0.0907*** (0.0111)	0.0566*** (0.0156)	0.2463*** (0.0267)	0.1543*** (0.0282)	0.0840** (0.0322)	-0.0283 (0.0329)	0.1454*** (0.0362)	0.0153 (0.0147)	T 0.1568*** (0.0210)
<i>Instr. Cluster SE p-value</i>	0.0000	0.0014	0.0000	0.0000	0.0155	0.3973	0.0005	0.3079	0.0000
<i>InstrxHour Cluster SE p-value</i>	0.0000	0.0002	0.0000	0.0000	0.0061	0.2577	0.0003	0.3374	0.0000
<i>Wild Bootstrap SE p-value</i>	0.0000	0.0020	0.0000	0.0000	0.0220	0.3940	0.0000	0.3240	0.0000
R2	0.3019	0.1298	0.0711	0.0431	0.0500	0.0233	0.0633	0.0217	0.2208
Obs	1972	1972	1972	1972	1972	1972	1972	1972	986
Clusters	24	24	24	24	24	24	24	24	24
Panel B. PB Method vs. Control									
Control Mean	0.5894	0.7153	0.2165	0.5251	0.7512	0.8373	0.4486	0.3973	0.4234
PostxPB	0.0917*** (0.0126)	0.0521*** (0.0184)	0.2589*** (0.0297)	0.1908*** (0.0320)	0.0821** (0.0342)	0.0052 (0.0322)	0.1538*** (0.0359)	0.0409** (0.0170)	PB 0.1474*** (0.0226)
<i>Instr. Cluster SE p-value</i>	0.0000	0.0095	0.0000	0.0000	0.0248	0.8718	0.0003	0.0246	0.0000
<i>InstrxHour Cluster SE p-value</i>	0.0000	0.0038	0.0000	0.0000	0.0302	0.8371	0.0003	0.0274	0.0000
<i>Wild Bootstrap SE p-value</i>	0.0000	0.0060	0.0000	0.0000	0.0280	0.8880	0.0000	0.0320	0.0000
R2	0.2840	0.1435	0.0680	0.0522	0.0616	0.0225	0.0715	0.0218	0.2247
Obs	1412	1412	1412	1412	1412	1412	1412	1412	706
Panel C. ROT Method vs. Control									
Control Mean	0.5894	0.7153	0.2165	0.5251	0.7512	0.8373	0.4486	0.3973	0.4234
PostxROT	0.0897*** (0.0121)	0.0613*** (0.0163)	0.2334*** (0.0299)	0.1168*** (0.0356)	0.0860** (0.0328)	-0.063 (0.0372)	0.1368** (0.0557)	-0.010 (0.0181)	ROT 0.1625*** (0.0340)
<i>Instr. Cluster SE p-value</i>	0.0000	0.0010	0.0000	0.0033	0.0152	0.1041	0.0219	0.5529	0.0001
<i>InstrxHour Cluster SE p-value</i>	0.0000	0.0010	0.0000	0.0032	0.0147	0.0351	0.0107	0.5469	0.0000
<i>Wild Bootstrap SE p-value</i>	0.0000	0.0000	0.0000	0.0040	0.0180	0.1180	0.0260	0.5360	0.0000
R2	0.2963	0.1290	0.0514	0.0387	0.0441	0.0280	0.0589	0.0212	0.2177
Obs	1396	1396	1396	1396	1396	1396	1396	1396	698
Panel D. ROT Method vs. PB Method									
PB Mean	0.6775	0.7594	0.3038	0.5799	0.8455	0.9167	0.4670	0.3762	0.0000
PostxROT	-0.002 (0.0110)	0.0091 (0.0154)	-0.025 (0.0267)	-0.074* (0.0378)	0.0038 (0.0188)	-0.068*** (0.0216)	-0.016 (0.0599)	-0.051** (0.0198)	ROT 0.0087 (0.0392)
<i>Instr. Cluster SE p-value</i>	0.8538	0.5573	0.3489	0.0624	0.8391	0.0044	0.7794	0.0155	0.8257
<i>InstrxHour Cluster SE p-value</i>	0.8329	0.6396	0.5889	0.0869	0.9221	0.0088	0.7545	0.0097	0.7815
<i>Wild Bootstrap SE p-value</i>	0.7900	0.5880	0.3180	0.0760	0.8680	0.0060	0.7860	0.0180	0.8000
R2	0.2777	0.1137	0.0955	0.0480	0.0412	0.0209	0.0734	0.0319	0.1750
Obs	1136	1136	1136	1136	1136	1136	1136	1136	568

Robustness II: Binary Treatment Estimates

Appendix Table 4B. Alternate Estimation Models (Binary Treatment Controlling for Initial Scores) for Main Program Effects

Variable	Outcomes							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Topical Knowledge	Big 5 Knowledge	Self-Assessed Knowledge	Self-Efficacy	Motivation to Learn	Likelihood to Seek Advice	Self-Assessed Risk Pref	Patience
Panel A. Combined Treatment vs. Control								
Control Mean	0.5795	0.7000	0.2033	0.5096	0.7153	0.8278	0.4498	0.4049
T	0.1076*** (0.0142)	0.0672*** (0.0173)	0.2247*** (0.0308)	0.1315*** (0.0279)	0.1086*** (0.0320)	0.0137 (0.0230)	0.0981*** (0.0313)	0.0016 (0.0151)
R2	0.4885	0.2417	0.1996	0.1702	0.2012	0.1270	0.2389	0.1327
Obs	986	986	986	986	986	986	986	986
Panel B. PB Method vs. Control								
Control Mean	0.5795	0.7000	0.2033	0.5096	0.7153	0.8278	0.4498	0.4049
PB	0.1088*** (0.0143)	0.0710*** (0.0175)	0.2155*** (0.0337)	0.1583*** (0.0308)	0.1165*** (0.0298)	0.0419* (0.0237)	0.1056*** (0.0311)	0.0129 (0.0151)
R2	0.4781	0.2768	0.2069	0.1889	0.2137	0.1483	0.2574	0.1478
Obs	706	706	706	706	706	706	706	706
Panel C. ROT Method vs. Control								
Control Mean	0.5795	0.7000	0.2033	0.5096	0.7153	0.8278	0.4498	0.4049
ROT	0.1023*** (0.0149)	0.0612*** (0.0183)	0.2340*** (0.0332)	0.1021*** (0.0339)	0.0964** (0.0365)	-0.0147 (0.0244)	0.0855* (0.0493)	-0.0075 (0.0189)
R2	0.4926	0.2710	0.2123	0.1655	0.2088	0.1184	0.2396	0.1357
Obs	698	698	698	698	698	698	698	698
Panel D. ROT Method vs. PB Method								
PB Mean	0.7135	0.7701	0.4201	0.6597	0.8507	0.9097	0.5451	0.4043
ROT	-0.0080 (0.0058)	-0.0014 (0.0116)	-0.0231 (0.0237)	-0.0586** (0.0247)	-0.0193 (0.0181)	-0.0601*** (0.0160)	-0.0210 (0.0419)	-0.0179 (0.0151)
R2	0.3820	0.1363	0.1626	0.1567	0.1703	0.1420	0.2229	0.1284
Obs	568	568	568	568	568	568	568	568

Robustness III: Logit Marginal Effect Estimates

Appendix Table 4C. Alternate Estimation Models (Logit) for Select Main Program Effects

Variable	Outcomes								
	(1) Topical Knowledge	(2) Big 5 Knowledge	(3) Self-Assessed Knowledge	(4) Self-Efficacy	(5) Motivation to Learn	(6) Likeliness to Seek Advice	(7) Self-Assessed Risk Pref	(8) Patience	(9) New Problem Solving
Panel A. Combined Treatment vs. Control									
Control Mean			0.2033	0.5096	0.7153	0.8278	0.4498		
T	-	-	0.0828*** (0.0291)	0.1196*** (0.0328)	0.1266*** (0.0346)	0.0477* (0.0247)	0.0692* (0.0376)	-	-
Pseudo-R2	-	-	0.0319	0.0244	0.0442	0.0241	0.0388	-	-
Obs	-	-	1972	1972	1972	1972	1972	-	-
Panel B. PB Method vs. Control									
Control Mean			0.2033	0.5096	0.7153	0.8278	0.4498		
PB	-	-	0.0731** (0.0309)	0.0469 (0.0324)	0.0985*** (0.0234)	0.0742*** (0.0234)	-0.0002 (0.0316)	-	-
Pseudo-R2	-	-	0.0361	0.0305	0.0545	0.0291	0.0477	-	-
Obs	-	-	1412	1412	1412	1412	1412	-	-
Panel C. ROT Method vs. Control									
Control Mean			0.2033	0.5096	0.7153	0.8278	0.4498		
ROT	-	-	0.0930*** (0.0314)	0.0370 (0.0325)	0.0728*** (0.0275)	0.0536** (0.0275)	-0.0004 (0.0362)	-	-
Pseudo-R2	-	-	0.0254	0.0261	0.0371	0.0277	0.0392	-	-
Obs	-	-	1396	1396	1396	1396	1396	-	-
Panel D. ROT Method vs. PB Method									
PB Mean			0.4201	0.6597	0.8507	0.9097	0.5451		
ROT	-	-	-0.0020 (0.0209)	-0.0114 (0.0198)	-0.0216 (0.0189)	-0.0157 (0.0189)	-0.0136 (0.0212)	-	-
Pseudo-R2	-	-	0.0334	0.0231	0.0410	0.0184	0.0392	-	-
Obs	-	-	1136	1136	1136	1136	1136	-	-

Robustness IV: Alternate Likert-scale Thresholds

Appendix Table 5. OLS Estimates of Program Effects Using Alternate Thresholds for "High"

Variable	Outcomes				
	(1) Self-Assessed Knowledge	(2) Self-Efficacy	(3) Motivation to Learn	(4) Likeliness to Seek Advice	(5) Self-Assessed Risk Pref
Panel A: "High" Outcome ≥ 6					
Control Mean	0.4450	0.7201	0.8517	0.9139	0.6531
PostxT	0.3657*** (0.0376)	0.1396*** (0.0309)	0.0752** (0.0322)	0.0168 (0.0231)	0.1657*** (0.0333)
R2	0.1080	0.0502	0.0348	0.0191	0.0589
Obs	1972	1972	1972	1972	1972
Panel B: "High" Outcome ≥ 7 (Main Specification)					
Control Mean	0.2165	0.5251	0.7512	0.8373	0.4486
PostxT	0.2463*** (0.0267)	0.1543*** (0.0282)	0.0840** (0.0322)	-0.028 (0.0329)	0.1454*** (0.0362)
R2	0.0710	0.0423	0.0494	0.0212	0.0631
Obs	1972	1972	1972	1972	1972
Panel C: "High" Outcome ≥ 8					
Control Mean	0.0634	0.3218	0.5443	0.7356	0.2057
PostxT	0.0707*** (0.0202)	0.1468*** (0.0351)	0.0741** (0.0356)	-0.051 (0.0364)	0.0878*** (0.0283)
R2	0.0302	0.0289	0.0345	0.0141	0.0409
Obs	1972	1972	1972	1972	1972

Robustness V: DD estimates with Instructor FEs

- Original results:

Table 4. OLS Estimates of Main Program Effects

Variable	Outcomes									
	(1) Topical Knowledge	(2) Big 5 Knowledge	(3) Self-Assessed Knowledge	(4) Self-Efficacy	(5) Motivation to Learn	(6) Likelihood to Seek Advice	(7) Self-Assessed Risk Pref	(8) Patience	(9) New Problem Solving	
PB Mean	0.6775	0.7594	0.3038	0.5799	0.8455	0.9167	0.4670	0.3762	0.3022	
PostxROT	-0.0020 (0.0110)	0.0091 (0.0154)	-0.0254 (0.0267)	-0.0740* (0.0378)	0.0038 (0.0188)	-0.0682*** (0.0216)	-0.016 (0.0599)	-0.0518** (0.0198)	ROT (0.0392)	0.0087
R2	0.2777	0.1137	0.0955	0.0480	0.0412	0.0209	0.0734	0.0319	0.1750	
Obs	1136	1136	1136	1136	1136	1136	1136	1136	568	

Note: DoD Data. Columns 1-8 report the Difference-in-Differences estimates for Equation 2 for each outcome listed. Column 9 reports OLS estimates of Equation 2 for the new knowledge outcome that was only included on the final assessment. All regressions include section fixed effects. Heteroskedasticity robust standard errors, clustered at the instructor level, are depicted in parentheses. ***, **, and * reflect p<0.01, 0.05, and 0.10 respectively.

- Results with Instructor Fixed Effects:

Appendix Table 6. OLS Estimates of Main Program Effects (PB vs. ROT) with Instructor Fixed Effects

Variable	Outcomes									
	(1) Topical Knowledge	(2) Big 5 Knowledge	(3) Self-Assessed Knowledge	(4) Self-Efficacy	(5) Motivation to Learn	(6) Likelihood to Seek Advice	(7) Self-Assessed Risk Pref	(8) Patience	(9) New Problem Solving	
PB Mean	0.6775	0.7594	0.3038	0.5799	0.8455	0.9167	0.4670	0.3762	0.3022	
PostxROT	-0.0020 (0.0111)	0.0091 (0.0155)	-0.0254 (0.0268)	-0.0740* (0.0380)	0.0038 (0.0189)	-0.0682*** (0.0217)	-0.016 (0.0602)	-0.0518** (0.0199)	ROT (0.0421)	0.0159
R2	0.2932	0.1272	0.1087	0.0537	0.0608	0.0317	0.1024	0.0432	0.1995	
Obs	1136	1136	1136	1136	1136	1136	1136	1136	568	

Note: DoD Data. Columns 1-8 report the Difference-in-Differences estimates for Equation 2 for each outcome listed. Column 9 reports OLS estimates of Equation 2 for the new knowledge outcome that was only included on the final assessment. All regressions include section and instructor fixed effects. Heteroskedasticity robust standard errors, clustered at the instructor level, are depicted in parentheses. ***, **, and * reflect p<0.01, 0.05, and 0.10 respectively.

Discussion: Benchmarking vs. Economics Learning

Appendix Table. OLS Estimates of Main Program Effects on Principles of Economics Knowledge

Variable	Outcomes						
	(1) Overall Economics Knowledge	(2) Scarcity	(3) Opportunity Cost	(4) Capitalism & Inequality	(5) Comparative Advantage	(6) Marginal Analysis	(7) Long-Run Growth
Panel A. Combined Treatment vs. Control							
Control Mean	0.5146	0.7560	0.3505	0.5969	0.2560	0.3624	0.7656
PostxT	0.0476** (0.0186)	0.1098** (0.0435)	0.0823*** (0.0185)	0.0037 (0.0435)	-0.1156*** (0.0327)	0.1402*** (0.0394)	0.0651* (0.0337)
R2	0.3048	0.0912	0.2186	0.0385	0.1746	0.0621	0.0820
Obs	1972	1972	1972	1972	1972	1972	1972
Panel B. ROT Method vs. PB Method							
PB Mean	0.7078	0.8733	0.7292	0.6042	0.5816	0.5122	0.9462
PostxROT	0.0247 (0.0202)	0.0689** (0.0247)	-0.0146 (0.0376)	0.0121 (0.0464)	0.0074 (0.0755)	0.0875 (0.0605)	-0.0131 (0.0275)
R2	0.1909	0.0398	0.1448	0.0323	0.1213	0.0601	0.0259
Obs	1136	1136	1136	1136	1136	1136	1136

- Panel A results suggest the economics course increases knowledge by 9% (but not equal across measures)
 - Very similar to our main results (8-15%)
- Panel B results suggest that PB & ROT have equal effects
 - Placebo check ? ... or evidence of no externalities ?