Engaging reflective thinking during exam-like situations:

Slowing students down on short-answer questions increases performance

jared@phas.ubc.ca
@StangJared

Jared Stang, *Science Teaching and Learning Fellow*

Joss Ives, *Instructor I*

Department of Physics and Astronomy, UBC

Carl Wieman Science Education Initiative

joss@phas.ubc.ca
@jossives

Motivation:

Which (possibly extraneous) factors impact student performance on exams or in exam-like situations?

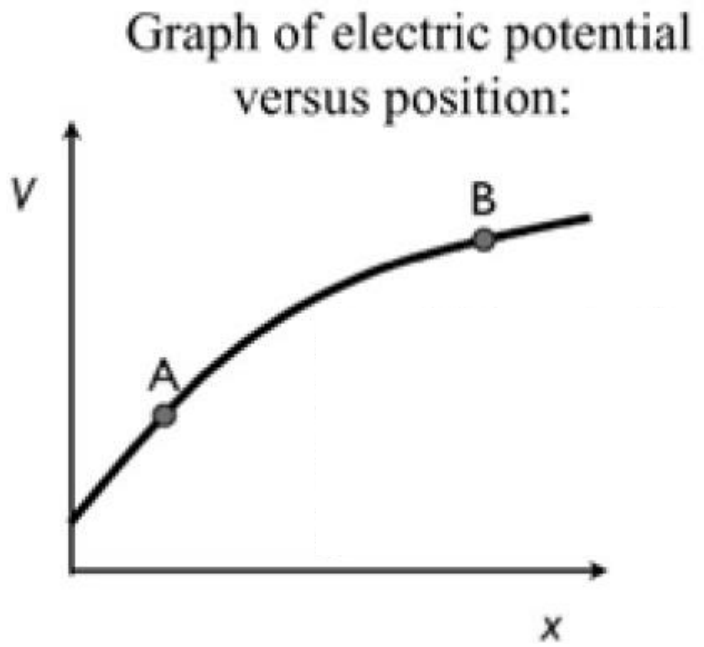
Background: Automatic cognitive processes play a role in student answers

- Dual processing theory describes two modes of thinking:

System 1	**System 2**
Unconscious, intuitive judgements	Deliberate, conscious mental effort

Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

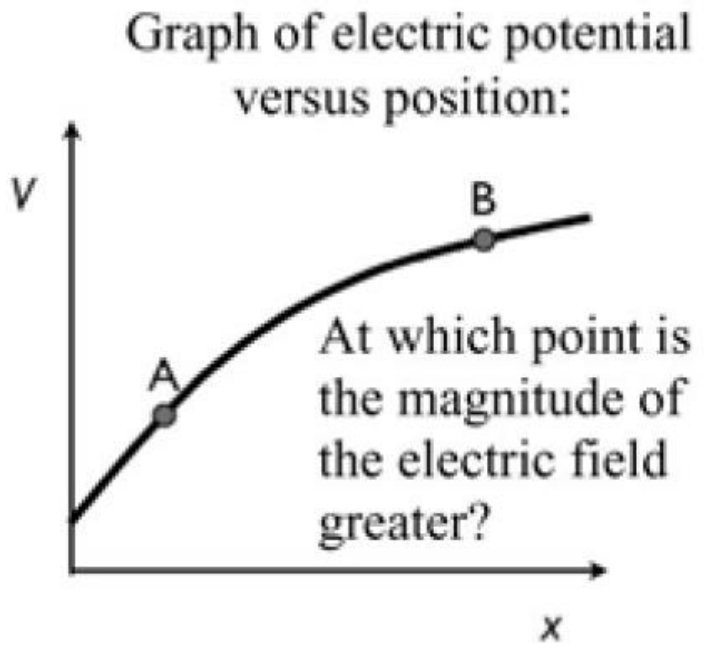
Background: Automatic cognitive processes play a role in student answers



- Students process height (of A vs B) faster than slope

Heckler and Scaife. *Cog. Sci.* 39.3 (2015): 496-537.
Heckler, Andrew F. Psychology of Learning and Motivation-Advances in Research and Theory 55 (2011): 227.

Background: Automatic cognitive processes play a role in student answers



- Students process height (of A vs B) faster than slope
- Imposing a delay increased performance (47% to 70%)

Heckler and Scaife. *Cog. Sci.* 39.3 (2015): 496-537.
Heckler, Andrew F. Psychology of Learning and Motivation-Advances in Research and Theory 55 (2011): 227.

The Cognitive Reflection Test (CRT) measures tendency to engage system 2

- 3-item test
- Each item has an incorrect intuitive response

Frederick. *J. Econ. Perspect.* 19.4 (2005): 25-42.

The Cognitive Reflection Test (CRT) measures tendency to engage system 2

- 3-item test
- Each item has an incorrect intuitive response

E.g:

A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

Frederick. *J. Econ. Perspect.* 19.4 (2005): 25-42.

Motivation:

Which (possibly extraneous) factors impact student performance on exams or in exam-like situations?

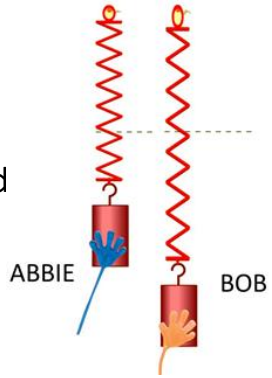
Motivation:

Which (possibly extraneous) factors impact student performance on exams or in exam-like situations?

Does this correlate with the CRT?

Extraneous example: Question pairs

Question Pair 1: Two friends, Abbie and Bob, are using a setup consisting of two identical springs and two identical masses the behaviour of masses oscillating up and down on springs. They both pull their masses down, but Bob pulls his down further than Abbie.



Q1a: If they both release their masses at the exact same time, which mass will oscillate up and then back down and return to that person's hand in the shortest time?

A) Abbie's mass will return to her hand in the shortest time
B) Bob's mass will return to his hand in the shortest time
C) Each mass will take the same amount of time to return to the hand of the person that released it

Q1b: If they both release their masses at the exact same time, which mass will have the highest maximum speed while it travels up and then back down again?

A) Abbie's mass will have the highest maximum speed
B) Bob's mass will have the highest maximum speed
C) Both masses will have the same maximum speed

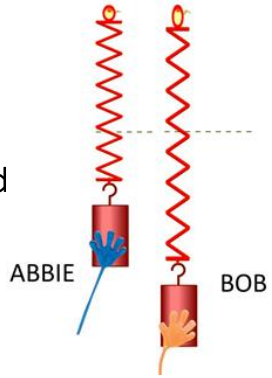
Question pairs crossover design

- Hypothesis: Student response can be influenced by asking about related concepts together ("Treatment") versus in isolation ("Control").
- Used on 9 question pairs on an end-of-term diagnostic
- First-year physics, N = 650 students

	Question Pair 1		Question Pair 2		Question Pair 3		…	Question Pair 9	
	Q1a	Q1b	Q2a	Q2b	Q3a	Q3b		Q9a	Q9b
Study Group A	Treat	Treat		Control	Control			Control	
Study Group B	Control		Treat	Treat		Control			Control
Study Group C		Control	Control		Treat	Treat		Treat	Treat

Extraneous example: Question pairs

Question Pair 1: Two friends, Abbie and Bob, are using a setup consisting of two identical springs and two identical masses the behaviour of masses oscillating up and down on springs. They both pull their masses down, but Bob pulls his down further than Abbie.



Q1a: If they both release their masses at the exact same time, which mass will oscillate up and then back down and return to that person's hand in the shortest time?

A) Abbie's mass will return to her hand in the shortest time
B) Bob's mass will return to his hand in the shortest time
C) Each mass will take the same amount of time to return to the hand of the person that released it

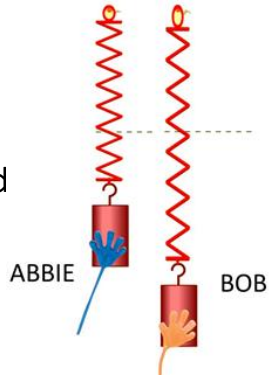
Q1b: If they both release their masses at the exact same time, which mass will have the highest maximum speed while it travels up and then back down again?

A) Abbie's mass will have the highest maximum speed
B) Bob's mass will have the highest maximum speed
C) Both masses will have the same maximum speed

	Question Pair 1	
	Q1a	Q1b
Treatment - Study Group A: Sees both questions as presented to the left	Treat	Treat
Control - Study Group B: Sees only the question stem and Q1a.	Control	
Control - Study Group C: Sees only the question stem and Q1b.		Control

Extraneous example: Question pairs

Question Pair 1: Two friends, Abbie and Bob, are using a setup consisting of two identical springs and two identical masses the behaviour of masses oscillating up and down on springs. They both pull their masses down, but Bob pulls his down further than Abbie.



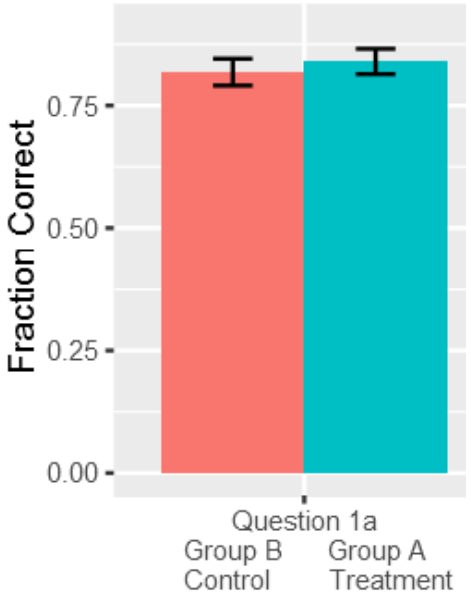
Q1a: If they both release their masses at the exact same time, which mass will oscillate up and then back down and return to that person's hand in the shortest time?

A) Abbie's mass will return to her hand in the shortest time
B) Bob's mass will return to his hand in the shortest time
C) Each mass will take the same amount of time to return to the hand of the person that released it

Q1b: If they both release their masses at the exact same time, which mass will have the highest maximum speed while it travels up and then back down again?

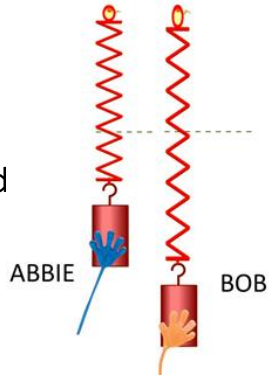
A) Abbie's mass will have the highest maximum speed
B) Bob's mass will have the highest maximum speed
C) Both masses will have the same maximum speed

	Question Pair 1	
	Q1a	Q1b
Treatment - Study Group A: Sees both questions as presented to the left	Treat	Treat
Control - Study Group B: Sees only the question stem and Q1a.	Control	
Control - Study Group C: Sees only the question stem and Q1b.		Control



Extraneous example: Question pairs

Question Pair 1: Two friends, Abbie and Bob, are using a setup consisting of two identical springs and two identical masses the behaviour of masses oscillating up and down on springs. They both pull their masses down, but Bob pulls his down further than Abbie.



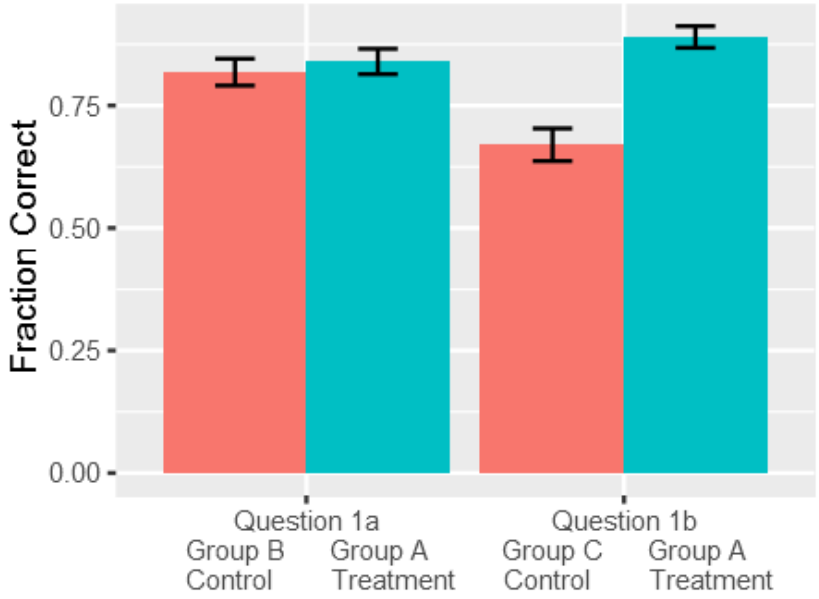
Q1a: If they both release their masses at the exact same time, which mass will oscillate up and then back down and return to that person's hand in the shortest time?

A) Abbie's mass will return to her hand in the shortest time
B) Bob's mass will return to his hand in the shortest time
C) Each mass will take the same amount of time to return to the hand of the person that released it

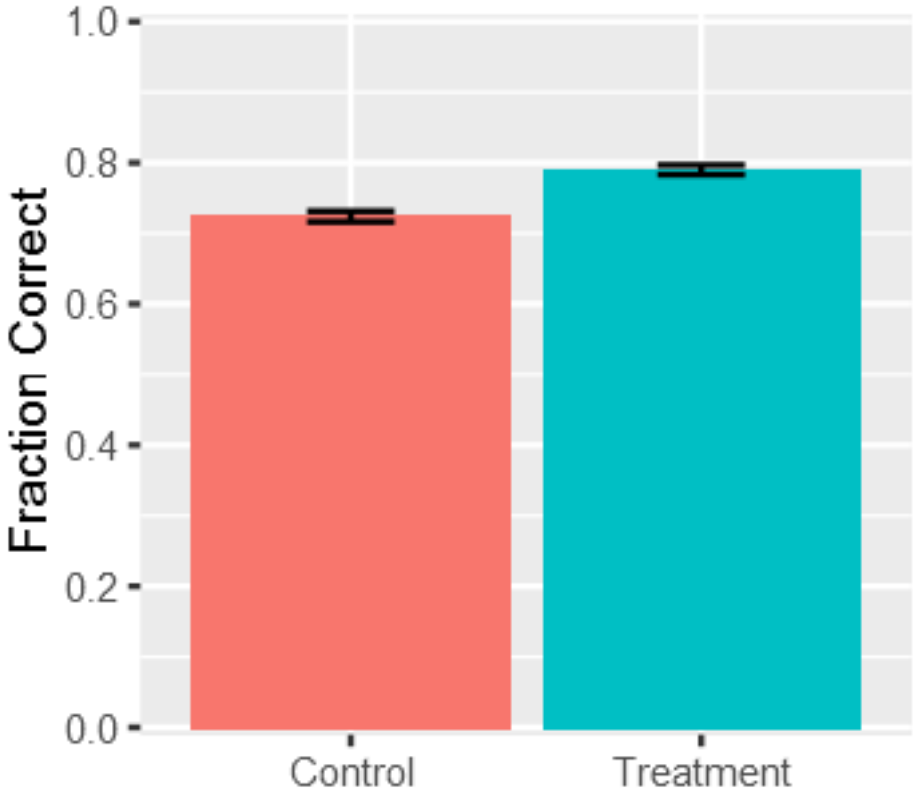
Q1b: If they both release their masses at the exact same time, which mass will have the highest maximum speed while it travels up and then back down again?

A) Abbie's mass will have the highest maximum speed
B) Bob's mass will have the highest maximum speed
C) Both masses will have the same maximum speed

	Question Pair 1	
	Q1a	Q1b
Treatment - Study Group A: Sees both questions as presented to the left	Treat	Treat
Control - Study Group B: Sees only the question stem and Q1a.	Control	
Control - Study Group C: Sees only the question stem and Q1b.		Control

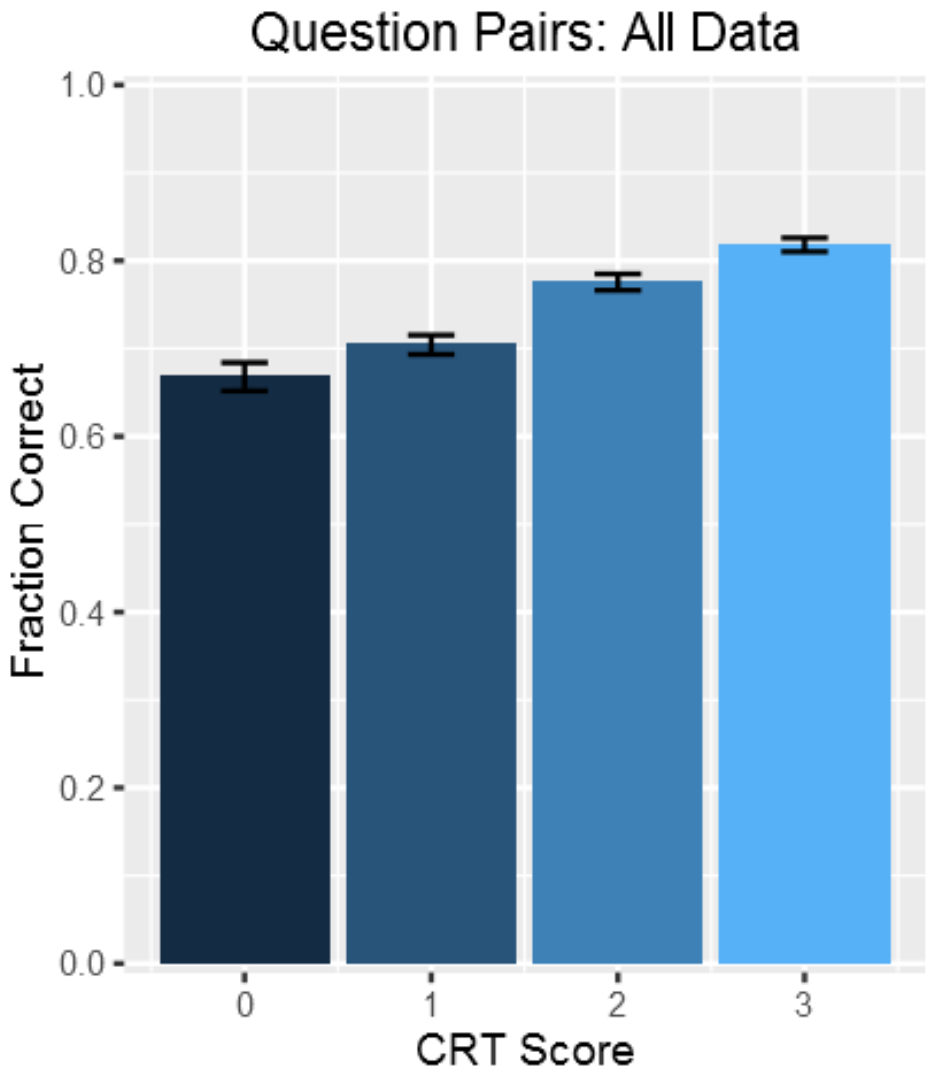


Results: A statistically significant improvement in question performance when pairs asked together ("Treatment")



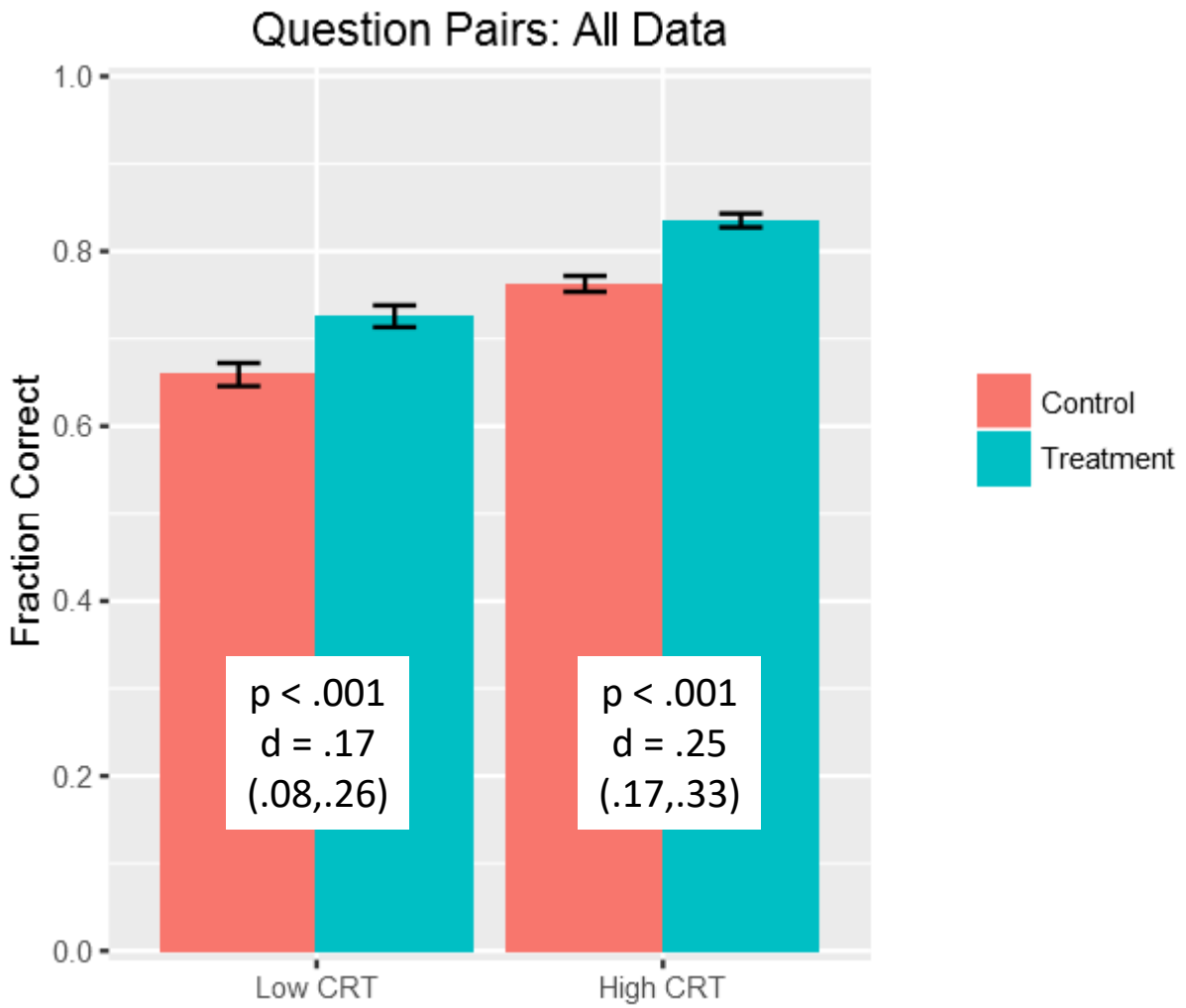
	p value	Odds ratio (95% CI)	Effect size (Cohen's d)
Fisher's Exact Test	<.001	1.43 (1.28, 1.60)	0.19 (0.14, 0.26)
Logistic Regression	<.001	1.53 (1.36, 1.72)	0.23 (0.17, 0.30)

Results: Performance correlates with CRT score



(CRT and FCI are correlated: Wood et al. *Phys. Rev. Phys. Educ. Res.* 12.2 (2016): 023101)

Results: The "Treatment" does not appear to impact students with different CRT scores differently.



Comparisons made using Fisher's Exact Test; Cohen's d reported for effect size with 95% Confidence intervals

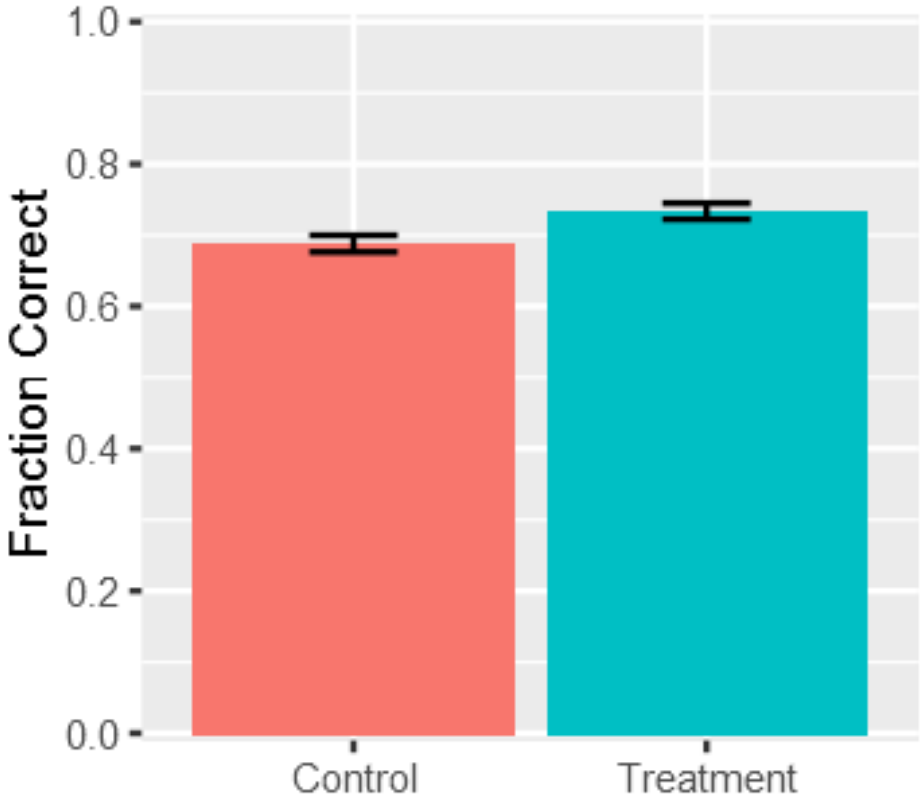
Experiment 2: “Explain your answer”

- Idea: Slow students down by including an “Explain your answer” box with short-answer questions (“Treatment”):

[0.5 points] Explain your reasoning in 1-2 sentences:

- Used on two questions on each of three exams
- First-year physics, N = 650 students

Results: A statistically significant improvement in question performance when asked to explain ("Treatment")



	p value	Odds ratio (95% CI)	Effect size (Cohen's d)
Fisher's Exact Test	.0061	1.25 (1.06, 1.46)	0.12 (0.03, 0.21)
Logistic Regression	.0031	1.28 (1.09, 1.51)	0.14 (0.05, 0.23)

Summary: Extraneous factors impact student performance

- Both **Question pairs** and **Explain your answer** impacted student performance, *without changing the question*
- Competing relevant and irrelevant dimensions may be mediated by automatic processes
  (but CRT did not interact with Treatment)

Broader questions: What does developing expertise look like, and how do we assess it?

jared@phas.ubc.ca