# Quantitative Analyses for Valuing Students' Incorrect Responses to Common Assessments

Trevor I. Smith[1,2]

Kyle J. Louis[1,2], Bartholomew J. Ricci IV[1,3]

[1]Departments of Physics & Astronomy, [2]STEAM Education, [3]Mathematics

Rowan University™

Physics Teacher Education Coalition

NSF

# Acknowledgements


- Contributors
  - Michael C. Wittmann
  - Tom Carter
  - Priscilla W. Laws
  - Ellie Sayre
  - Sam McKagan
  - Joe Redish

- Funding
  - PhysTEC (w/ support from NSF)
  - Rowan University College of Science and Mathematics

- Productive Conversations
  - Andrew Heckler
  - David Meltzer

- Additional Student Researchers
  - Ian T. Griffin
  - Nicholas J. Wright
  - Ryan A. Moyer
  - Kerry A. Gray
  - Paul J. Kelly
  - Mitchell A. Nussenbaum
  - Louis J. Remy

# Force and Motion Conceptual Evaluation

- 47-item multiple-choice survey[1]
- 40 questions are scored (total of 37 points)

Why should I listen to ANOTHER talk about the FMCE?

A sled on ice moves in the ways described in questions 1-7 below. *Friction is so small that it can be ignored*. A person wearing spiked shoes standing on the ice can apply a force to the sled and push it along the ice. Choose the <u>one</u> force (**A** through **G**) which would **keep the sled moving** as described in each statement below.

**A.** The f…

Direction of Force

Direction of Force

…left and is

…strength (magnitude).

___1. Which force would keep the sle… toward the right and speeding up at a steady rate (constant acceleration)?

___2. Which force would keep the sled moving toward the right at a steady (constant) velocity?

___3. The sled is moving toward the right. Which force would slow it down at a steady rate (constant acceleration)?

Questions 22-26 ref… along a…

…t or left on a horizontal surface …ion is to the right.

+

…ter (**A** to **G**) of the **acceleration-** …in each statement.

…at none is correct, answer

Time

**F**

Acc… Time

**C**

Acc… Time

**G**

Acc… Time

**D**

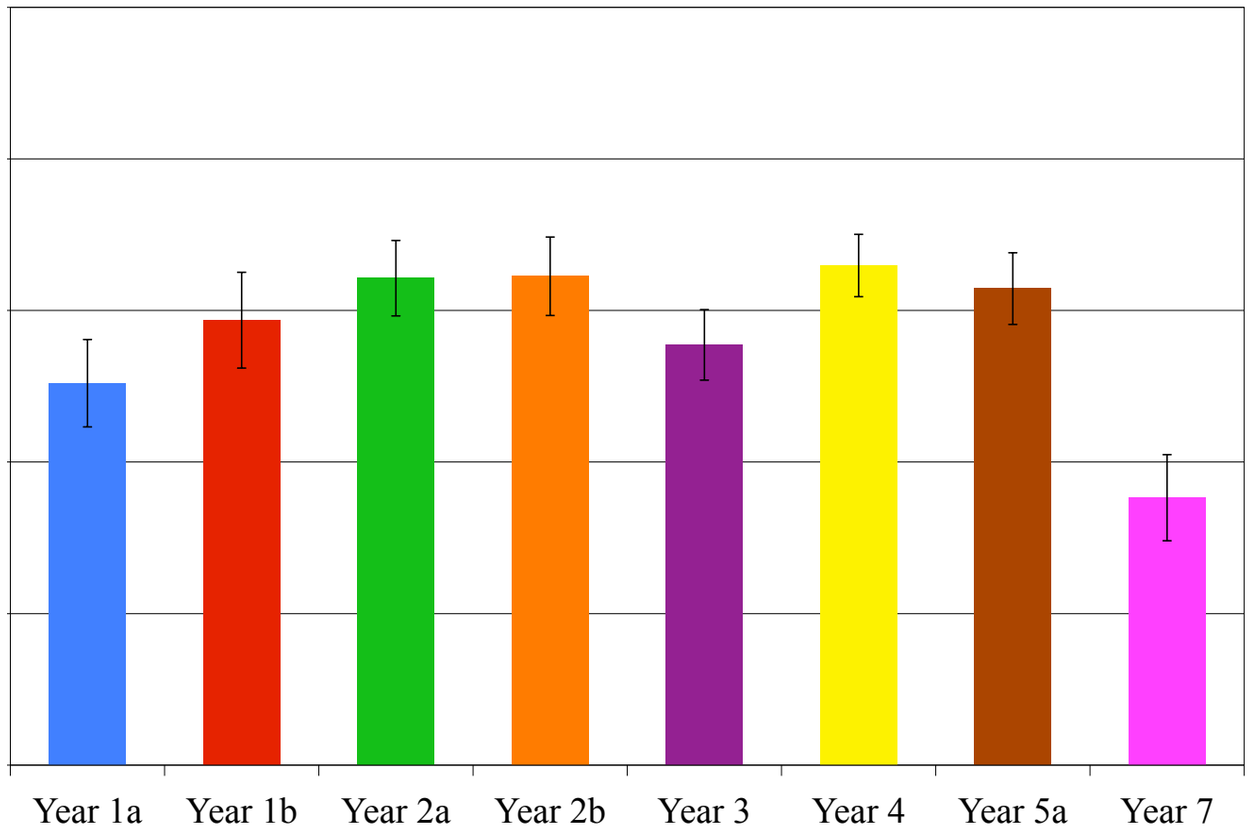
Acc… Time

**J** None of these graphs is correct.

1. Thorton & Sokoloff, *Am. J. Phys.* (1998)

# Typical Analyses

- Give FMCE at beginning and end of course
- **Score each response as correct or incorrect**
- Count the number of questions answered correctly
- Calculate average normalized gain <*g*>
- Because *g* is biased[2], maybe calculate Cohen's *d* or another measure of effect size
- Compare data sets




error bar = pooled standard error of pre/post-test scores

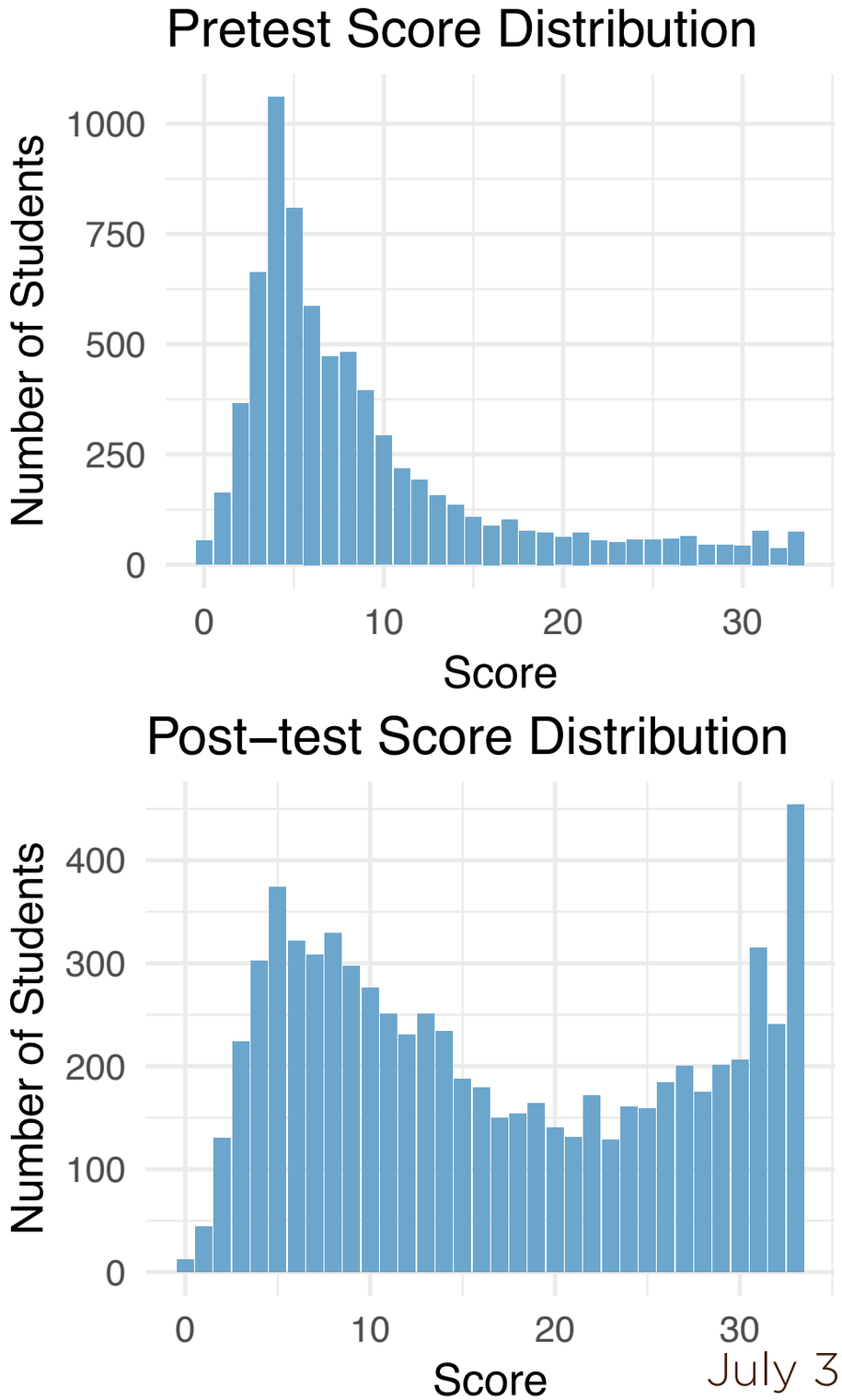
2. Nissen, Talbot, Thompson, and Van Dusen, *Phys Rev Phys Educ Res*, (2018)

# What's wrong with typical analyses?

- Many students answer most questions incorrectly (either before or after instruction)
- We focus on what they DON'T know rather than valuing what they DO know.
- Many students change their answer from one incorrect to another

**Are students learning even if they don't choose the correct answer?**

Data from 7,288 students



# Ranking incorrect responses

- **Are some incorrect responses better than others?**

- **What makes one response better than another?**

- **What productive ideas are students expressing by choosing a particular incorrect answer?**

# Assumption 1

Students who have a higher understanding of physics (as measured by the FMCE) are more likely to choose better responses than students who have a lower understanding of physics.

# Item Response Theory (IRT)

- Two-parameter-logistic (2PL) nested-logit model[3]
- Estimates students' overall knowledge[4]
- Probability of answering correctly:

$$P(\theta) = \frac{1}{1 + e^{-a(\theta - b)}}$$

- Probability of choosing the $k^{\text{th}}$ incorrect answer is the product of being incorrect and Bock's nominal response model[5]:

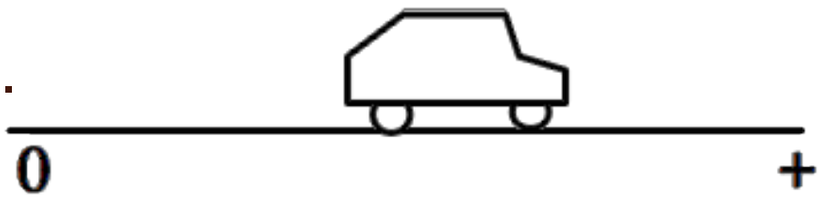
$$P_k(\theta) = \left(1 - \frac{1}{1 + e^{-a(\theta - b)}}\right) \frac{e^{a_k(\theta - b_k)}}{\sum_i e^{a_i(\theta - b_i)}}$$

- Incorrect responses ranked by value of $a_k$

3. Suh and Bolt, *Psychometrika* (2010)
4. Baker, The Basics of Item Response Theory (2001)
5. Bock, *Psychometrika* (1972)

# Question 14

A toy car ... can move to the right or left along a horizontal line (the positive part of the distance axis).



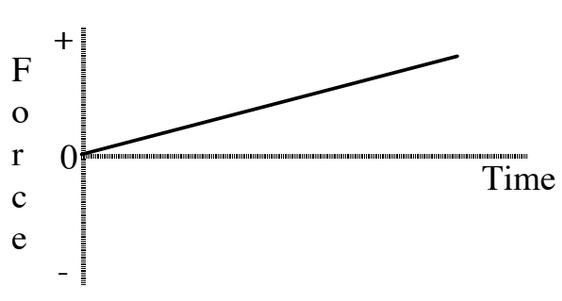

Assume that friction is so small that it can be ignored. **Choose the one force graph ... which could allow the described motion of the car to continue.** If you think that none is correct, answer choice J.

14. The car moves toward the right (away from the origin) with a steady (constant) velocity.

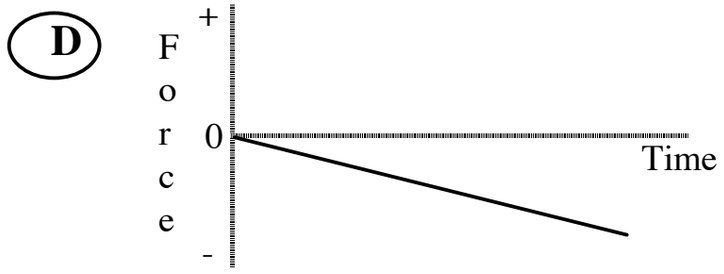
**A**


**B**

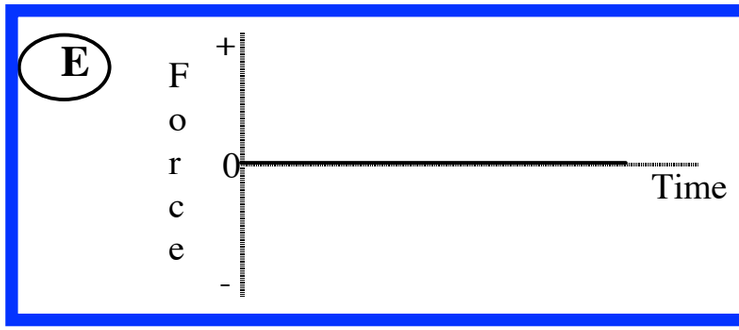

**C**



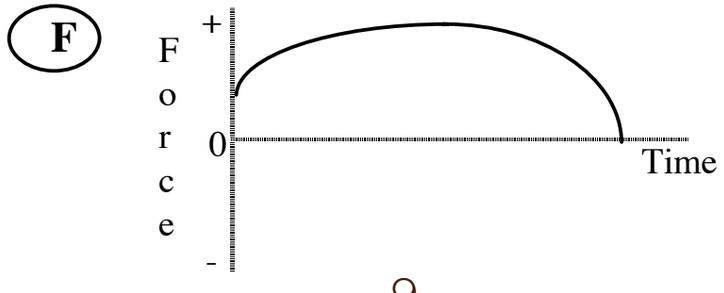

**D**



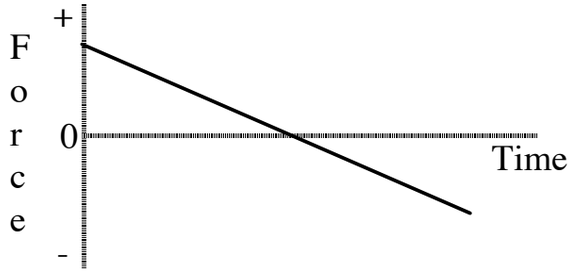

**E**



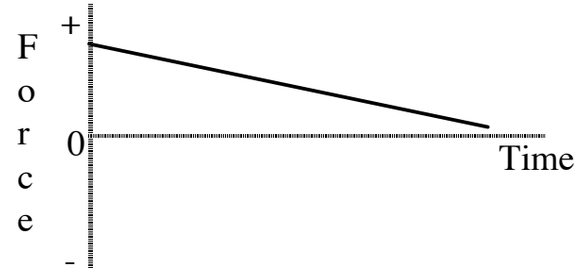

**F**




**G**

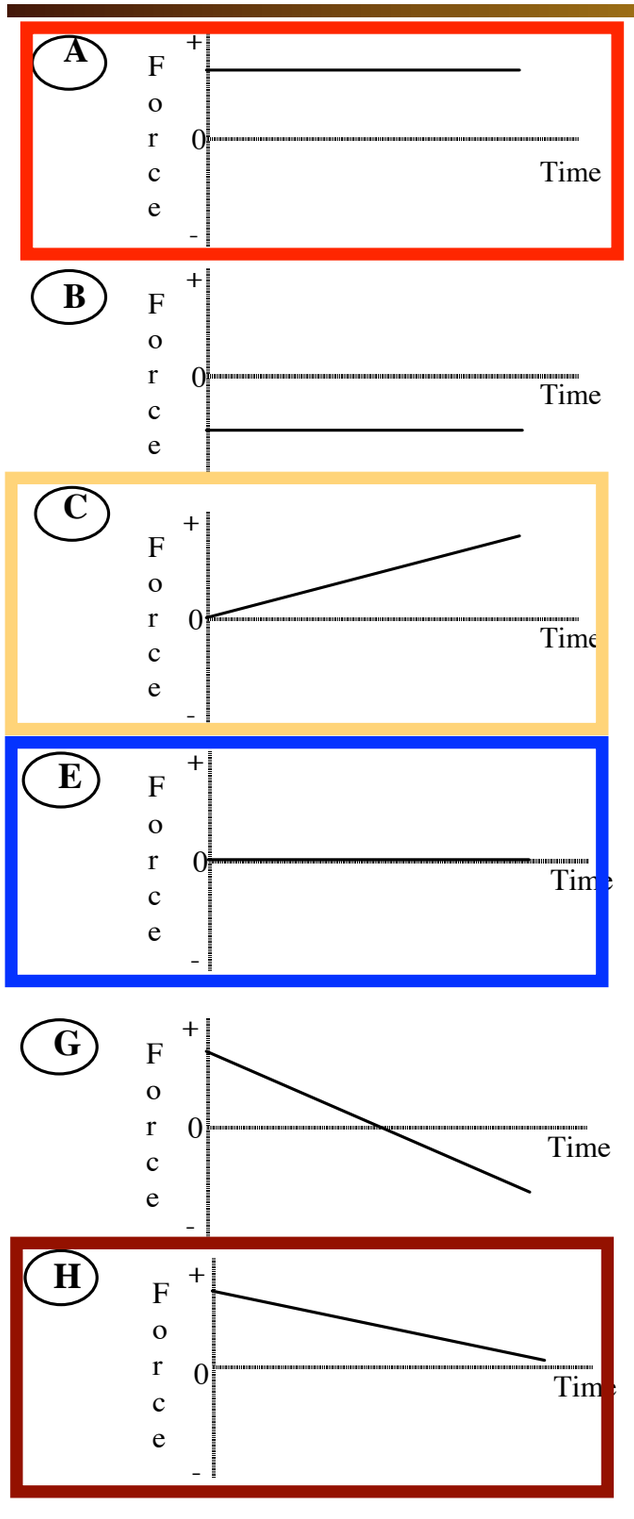


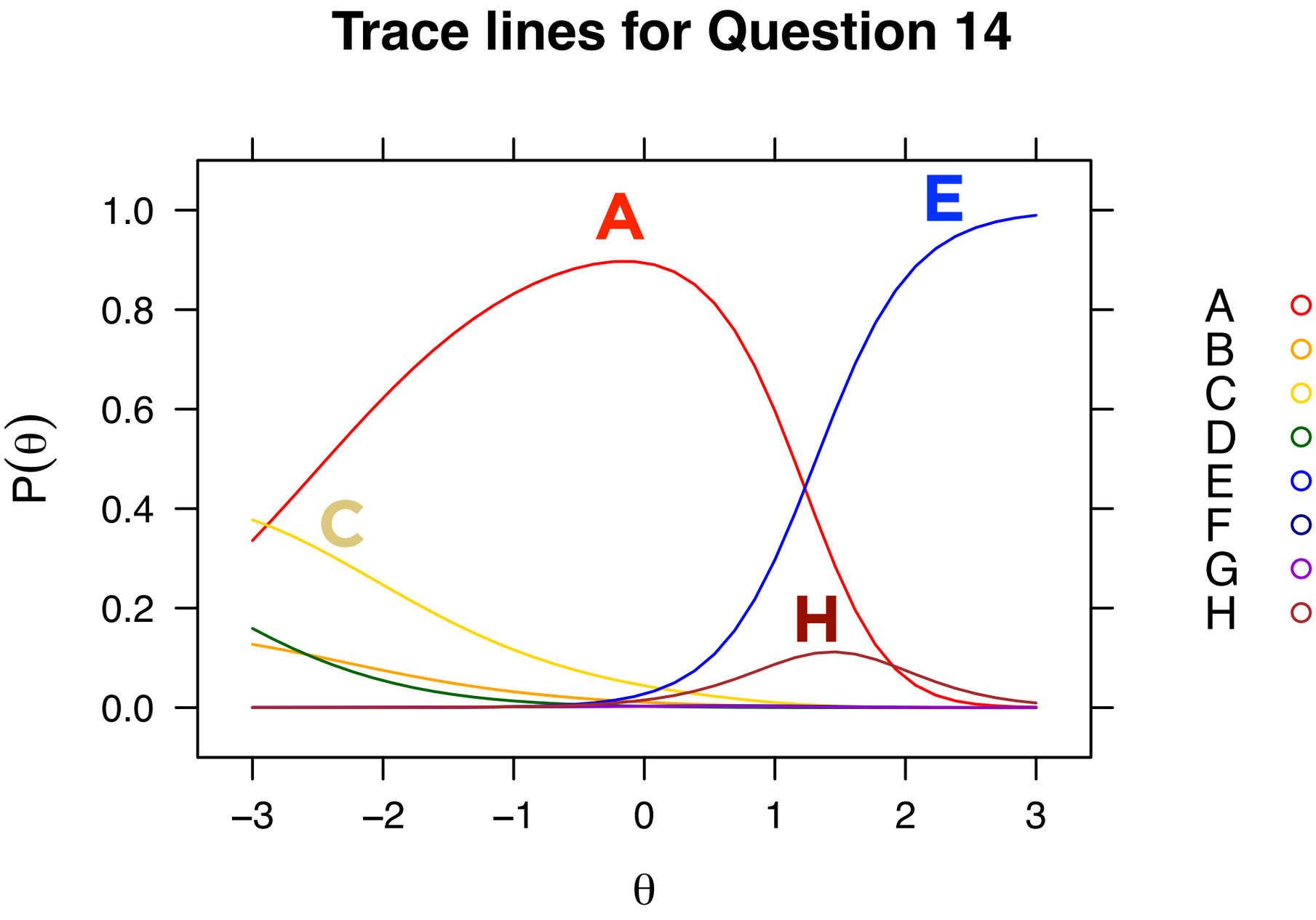

**H**



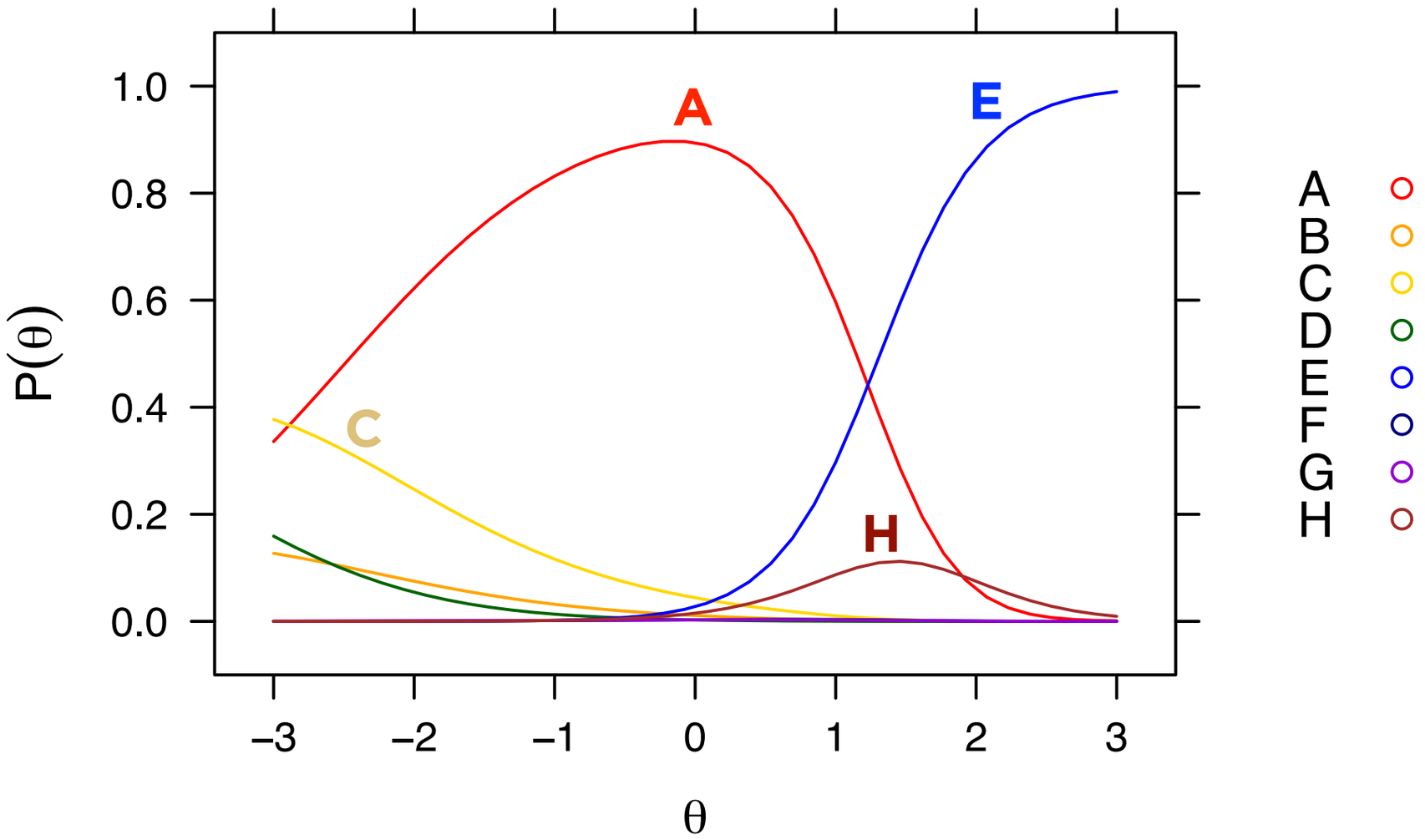

**J** None of these graphs is correct.

# IRT Plot


A
Force
+
0
-
Time
B
Force
+
0
Time
C
Force
+
0
Time
E
Force
+
0
Time
G
Force
+
0
Time
H
Force
+
0
Time


Trace lines for Question 14
P(θ)
1.0
0.8
0.6
0.4
0.2
0.0
−3
−2
−1
0
1
2
3
θ
A
E
C
H
A
B
C
D
E
F
G
H

# IRT Results




$$\begin{array}{ccccccc} \mathrm{E} & > & \mathrm{H} & > & \mathrm{A} & > & \mathrm{C} \\ a_k: & & 2.94 & > & 0 & > & -0.26 \end{array}$$

# Assumption 2

Students are more likely to choose better responses after instruction than before instruction.

# McNemar-Bowker (MB) chi-square test

- Considers the number of students who give each pre/post response pair (each transition)[6,7]
- Uses chi-square to look for asymmetries

| | | Post-test | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** |
| Pretest | **A** | 2881 | 73 | **309** | 42 | **1918** | **22** | 10 | **118** |
| | **B** | 55 | 8 | 13 | 1 | 35 | 1 | 1 | 2 |
| | **C** | **236** | 2 | 76 | 6 | 104 | 2 | 2 | 7 |
| | **D** | 24 | 1 | 9 | 2 | 12 | 0 | 0 | 0 |
| | **E** | **101** | 9 | 14 | 1 | 753 | 8 | 0 | 10 |
| | **F** | **8** | 0 | 1 | 0 | 7 | 0 | 1 | 0 |
| | **G** | 6 | 0 | 1 | 1 | 7 | 0 | 0 | 0 |
| | **H** | **32** | 0 | 3 | 0 | 114 | 0 | 1 | 11 |

6. McNemar, *Psychometrika* (1947)

7. Bowker, *Journal of the American Statistical Association* (1948)

# MB Results

- Pairwise comparisons show asymmetries
- p-values adjusted using false discovery rate (FDR) method

| Comparison | Adjusted p-value | Percent of Population |
|---|---|---|
| E > A | 0 | 28.3% |
| C > A | 0.00861 | 7.7% |
| H > A | 6.63E-12 | 2.1% |
| E > H | 2.52E-22 | 1.7% |
| E > C | 3.62E-17 | 1.7% |
| E > B | 0.00053 | 0.6% |
| F > A | 0.0402 | 0.5% |
| C > B | 0.0222 | 0.4% |
| E > D | 0.0114 | 0.3% |
| E > G | 0.0402 | 0.2% |

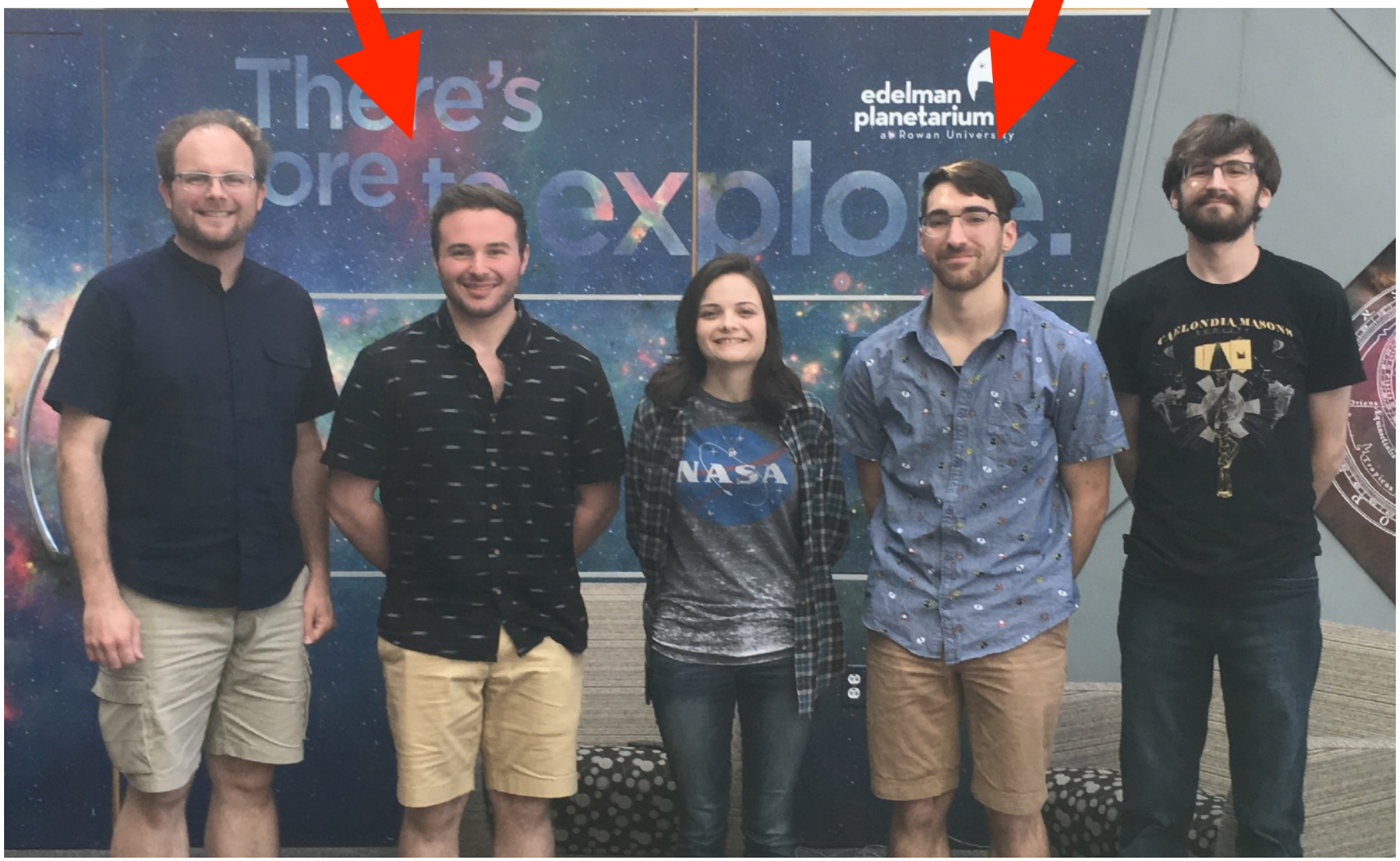
# Unified Ranking

$$E > H > A$$

- Supported by both methods
- The relationship with C is unclear
  - IRT says C < A
  - MB says C > A
- More analysis is needed (and maybe more data)
- Response A overwhelms pretest data and may skew MB results
- Future work will involve student interviews
- Ultimate goal: New metric for measuring student understanding and learning. (Down with <*g*>!)

# Rowan University Physics Education Research Team: Summer 2017/2018

RUPERt




## PERC Poster A51, Session 1, Wed. 8/1, 5:00pm

# Future work: Developing a new metric for student learning