Determining a Hierarchy of Correctness Through Student Transitions on the FMCE

Kyle J. Louis [1,2], Bartholomew J. Ricci [1,3], and Dr. Trevor I. Smith [1,2]
Department of [1] Physics & Astronomy, [2] STEAM Education, and [3] Mathematics, Rowan University, Glassboro, NJ, 08028

Background

Purpose: To determine a hierarchy of "correctness" on a commonly used multiple-choice assessment in introductory physics courses.

- Data were collected from over 7,000 students
- We omit *J* as an answer choice because it provides no understanding of a student's understanding.

Item Response Theory

- **Assumption:** Students who choose correct responses on most questions are more likely to choose more sophisticated incorrect answers than students who choose few correct responses [2]
- 2-parameter nested logit model for a multiple choice test [3,4]

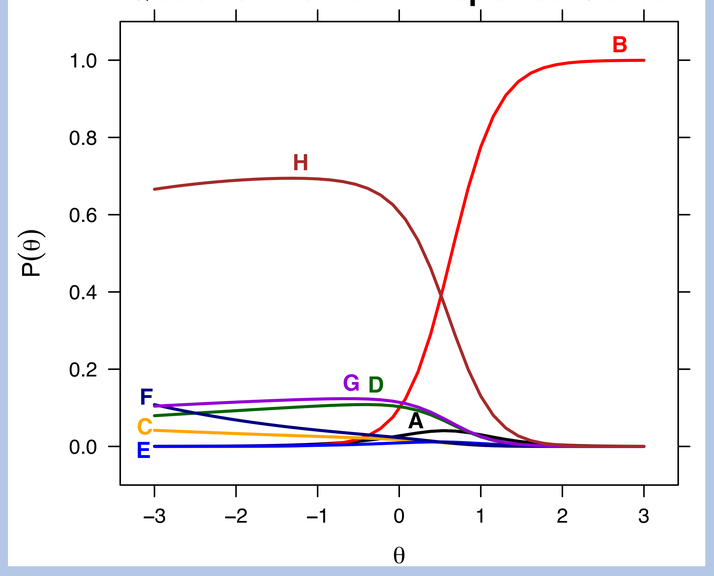
$$\textbf{Correct: } P(\theta) = \frac{1}{1+e^{-a(\theta-b)}}$$

- *a*: related to slope of the item response curve (IRC)
  - Discriminates student understanding
- *b*: Difficulty of the question
  - Right shift means harder question
  - Left shift means easier question

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

- Correctness of answer choices determined by higher a_k value

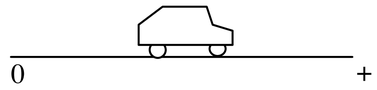
Question 18 Item Response Curve



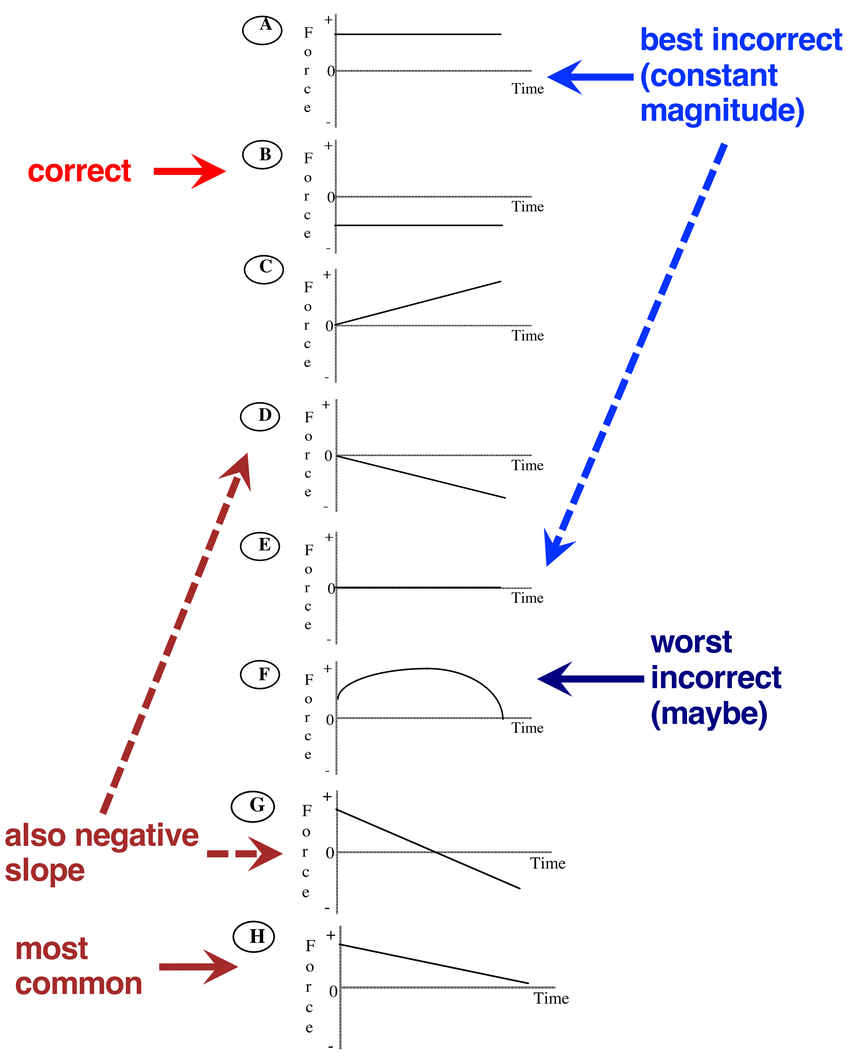
$$B > A > D = G = H > C > F$$
$$a_k: \; 0 > -1.43 = -1.47 = -1.52 > -1.87 > -2.11$$

Force and Motion Concept Evaluation [1]

Assume that friction is so small that it can be ignored. A force is applied to the car. Choose the one force graph (**A** through **H**) that for each statement below which could allow the described motion of the car.



__18. The car moves toward the right and is slowing down at a steady rate (constant acceleration).



Hypotheses

H_0: The number of transitions from one answer choice to another is the same in both directions

H_a: More students transition in one direction between two answer choices than in the other

McNemar-Bowker Chi-Square Test for Asymmetry

- **Assumption:** Students are more likely to choose more sophisticated responses after instruction than before instruction.
- Using the False Discovery Rate (FDR) correction, the adjusted *p*-value determines whether or not a transition is statistically significant [5,6]

Pretest \ Posttest	Q18	A	B	C	D	E	F	G	H
	A	4	28	2	9	0	2	4	26
	B	14	717	6	33	5	4	12	87
	C	7	37	6	10	1	4	12	46
	D	16	208	13	87	10	14	63	236
	E	1	11	1	4	2	1	0	8
	F	10	45	3	16	2	16	22	72
	G	22	250	13	59	7	21	93	281
	H	92	1420	52	227	26	72	281	1904

Statistically Significant Transitions		
Response Comparison	Adjusted *p*-value	Percent of Population
B > H	< 0.001	22.3%
B > G	< 0.001	3.9%
B > D	< 0.001	3.6%
A > H	< 0.001	1.7%
B > F	< 0.001	0.7%
B > C	< 0.001	0.6%
E > H	0.01	0.5%
A > G	0.002	0.4%

Statistically Insignificant Transitions		
Response Comparison	Adjusted *p*-value	Percent of Population
G = H	1	8.3%
D = H	1	6.9%
D = G	1	1.8%

$$B > \{H, G, D, F, C\}; \quad D = G = H$$
$$A > \{H, G\}; \quad E > H$$

Unified Ranking

$$B > A > D = G = H$$

Future Research

- Determining the models for each answer choice via interviews
- Synthesize results into a unified ranking system
- Use additional analyses to rank responses with different assumptions

References

1. R. K. Thornton and D. R. Sokoloff, *Am. J. Phys.* **66**, 338 (1998).
2. T. I. Smith, K. A. Gray, K. J. Louis, B. J. Ricci, and N. J. Wright, *PERC Proceedings*, p. 380 (2017).
3. Y. Suh and D. M. Bolt, *Psychometrika* **75**, 454 (2010).
4. R. D. Bock, *Psychometrika* **37**, 29 (1972).
5. Q. McNemar, *Psychometrika* **12**, 153 (1947).
6. A. H. Bowker, *J. Am. Stat. Assoc.* **48**, 572 (1948).

Acknowledgements

We thank Sam McKagan and Ellie Sayre for providing access to data from PhysPort's Data Explorer. We also thank Kerry Gray, Nicholas Wright, Ian Griffin, and Ryan Moyer for their previous contributions as members of the research team.