Telling new stories by reanalyzing old data: FMCE edition

Trevor I. Smith
Department of Physics & Astronomy, and Department of STEAM Education, Rowan University, Glassboro, NJ 08028

I present a reanalysis of data from the Force and Motion Conceptual Evaluation. The current analysis focuses on comparing student responses to force questions based on whether the answer choices are presented as verbal descriptions of forces or as graphs of force vs. time. I use contingency tables and the χ^2 test of independence to measure the correlations between responses and consistency plots to represent how student responses change after a full term of algebra-based introductory mechanics. Consistency plots display the wide variety of students' learning trajectories. Preliminary analyses reveal that student responses to graph questions improve more than their responses to verbal/description questions.

PACS numbers: 01.40.Fk, 01.30.Cc

I. INTRODUCTION

The Force and Motion Conceptual Evaluation (FMCE) includes many questions that ask students to choose a force that would produce a described motion [1]. Recent analyses group questions into several clusters based on the type of motions presented and the form of the possible answer choices [2]. The Force Sled (FS) cluster focuses on a sled on a (frictionless) icy surface and "a person wearing spiked shoes . . . [who] can apply a force to the sled and push it along the ice" [1]. The FS cluster questions describe a motion, e.g. "1. Which force would keep the sled moving toward the right and speeding up at a steady rate (constant acceleration)?" The answer choices for the FS cluster are descriptions of forces, e.g. "**B.** The force is toward the **right** and is of **constant** strength (magnitude)."

The Force Graphs (FG) cluster focuses on an isomorphic situation in which a toy car moves to the right or left on a frictionless surface under the influence of a single applied force. The setup indicates that right is the positive direction. Like the FS cluster, each question presents a particular motion, e.g. "16. The car moves toward the right and is speeding up at a steady rate (constant acceleration)." For the FG questions, answers are provided as graphs of force vs. time (Fig. 1).

Four of the questions in the FS cluster have corresponding questions in the FG cluster that present the same motion, e.g. questions 1 and 16 presented above. Each of these questions has one correct answer choice (force is proportional to the rate of change of velocity, i.e. the acceleration) and one answer choice that is consistent with the common incorrect model (force is proportional to velocity) [2]. Table I summarizes the four question pairs.

TABLE I. Isomorphic question pairs from the Force Sled (FS) and Force Graphs (FG) question clusters of the FMCE, as well as the responses corresponding to both the correct model ($F \propto \Delta v/\Delta t$) and common incorrect model ($F \propto v$).

Case	Described Motion	Question		Correct		$F \propto v$	
		FS	FG	FS	FG	FS	FG
1	moving right, speeding up	1	16	B	A	A	C
2	moving right, steady speed	2	14	D	E	B	A
3	moving right, slowing down	3	18	F	B	C	H
4	moving left, speeding up	4	19	F	B	G	D

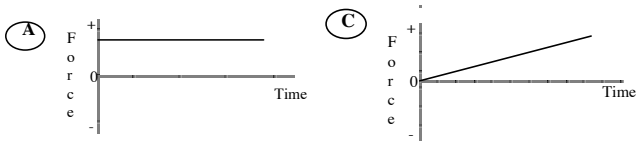

FIG. 1. Sample response choices for the Force Graphs question cluster on the FMCE.

Previous results suggest that student learning differs between the FS and FG clusters [3]. Studies have also revealed student difficulties understanding and interpreting kinematics graphs [4], and shown that students' responses to isomorphic force questions depend on representational cues [5]. In this study I use students' responses to the FMCE as a measure of their understanding of graphs and track individual students' responses to question pairs over an entire course.

I seek to answer several research questions. 1) Do students answer isomorphic questions with coherent responses? 2) How strongly are students' responses to FS and FG questions correlated? 3) Are students consistent with each other? 4) Do students' responses depend on the motion being considered? The answers to these could inform the type and variety of instructional methods needed to help students understand the relationship between forces and motion.

II. MEASURING STUDENT CONSISTENCY

Data for this study come from the algebra-based introductory mechanics course at a 2-year college in the midwestern U.S. Data were collected in eight sections over a span of seven years ($N = 194$) [6]. The FMCE was given at the beginning and end of each term. My analysis focuses solely on the eight questions of the FMCE listed in Table I. For the sake of brevity, I limit the following discussion to Cases 1 and 3.

I categorize each student's response to each question as either Correct (corresponding to the $F \propto \Delta v/\Delta t$ model), Common Incorrect (corresponding to the $F \propto v$ model), or Other Incorrect. There are multiple ways to represent these data. One of which is to use a 3×3 contingency table to show the

2015 PERC Proceedings, edited by Churukian, Jones, and Ding; Peer-reviewed, doi:10.1119/perc.2015.pr.074
Published by the American Association of Physics Teachers under a Creative Commons Attribution 3.0 license.
Further distribution must maintain attribution to the article's authors, title, proceedings citation, and DOI.

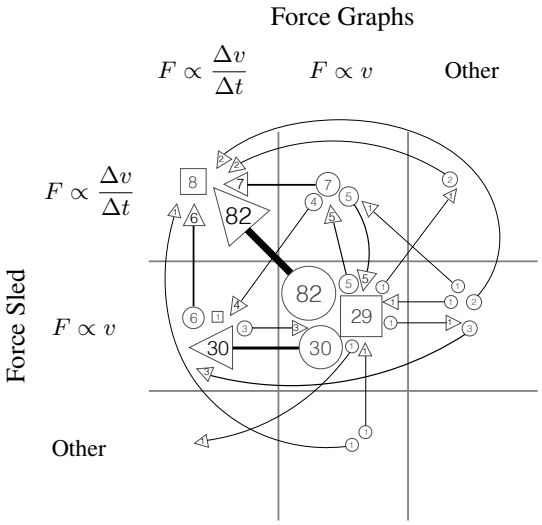

FIG. 2. Consistency plot for Case 1: moving right, speeding up ($N = 194$). Students start at circles and move to triangles; the size of the shape scales with the number of students making the same transition. The circle-line-triangle "arrow" shows a particular learning trajectory. Squares represent students who were unaffected by instruction.

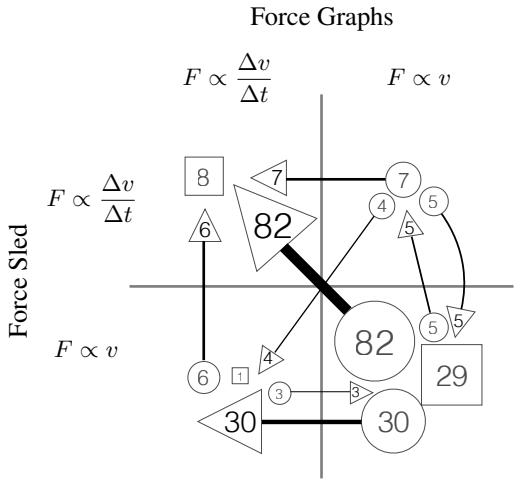

FIG. 3. Simplified consistency plot for Case 1 ($N = 180$)

number of students who gave each response pair either before or after instruction, e.g. Tables II and III. I use a χ^2 test of independence (with $\alpha = 0.05$) and report Cramér's coefficient, ϕ_c, as a measure of the strength of the correlation between the responses to the questions in each pair to determine whether on not performance on one question predicts performance on the other ($\phi_c \geq 0.35$ is a large correlation) [7–9].

There are three different forms of "consistency" that may be discussed with regard to these data: external, internal, and temporal. External consistency, which I refer to as just consistency, is a measure of how similar a student is to his/her classmates: *consistent* responses are represented by a table cell containing a large percentage of the class. Consistent students may be expected to respond the same way to particular instructional intervention, whereas inconsistent students may require a more varied approach. Internal consistency, which I call coherence, is a measure of whether or not a student answers the same way on FS and FG questions: *coherent* responses are represented by the diagonal table cells (upper left to lower right). Temporal consistency, which I call stubbornness, indicates whether or not a student changes his/her response(s) after instruction; *stubborn* responses cannot be represented on the pre- or posttest contingency tables.

All three forms of consistency may be simultaneously represented by using a Wittmann-Black consistency plot [10]. In this representation, a circle indicates a student's response pair on the pretest, and a triangle represents the student's response pair on the posttest; an arrow from the pretest response to the posttest response is formed by connecting the circle to the triangle by a line segment (or arc) [10]. The size of the shapes, and the thickness of the connecting line are scaled to represent the number of students who made the same move from pre- to postinstruction. Students who give the exact same response pair before and after instruction are represented by a square. In this representation, consistent responses are indicated by large shapes and thick lines, coherent responses are represented by shapes along the diagonal, and stubborn responses are represented by squares.

Contingency tables allow us to perform statistical tests to measure the association between the FS and FG responses, and consistency plots allow us to see the rich variety of student learning trajectories.

III. CASE 1: MOVING RIGHT, SPEEDING UP

Tables II and III respectively show the pre- and postinstruction contingency tables for Case 1: moving to the right

TABLE II. Pretest contingency table for Case 1 (moving right, speeding up): number of students who gave each response pair. A moderate correlation exists between the FS and FG questions: $\chi^2(4,\ N = 194) = 17.83,\ p = 0.001,\ \phi_c = 0.21$.

		Force Graphs		
		$F \propto \Delta v/\Delta t$	$F \propto v$	Other
Force Sled	$F \propto \Delta v/\Delta t$	8	16	2
	$F \propto v$	10	149	7
	Other	0	2	0

TABLE III. Posttest contingency table for Case 1 (moving right, speeding up): number of students who gave each response pair. A strong correlation exists between the FS and FG questions: $\chi^2(4,\ N = 194) = 53.22,\ p < 0.001,\ \phi_c = 0.37$.

		Force Graphs		
		$F \propto \Delta v/\Delta t$	$F \propto v$	Other
Force Sled	$F \propto \Delta v/\Delta t$	108	6	1
	$F \propto v$	38	39	1
	Other	1	0	0

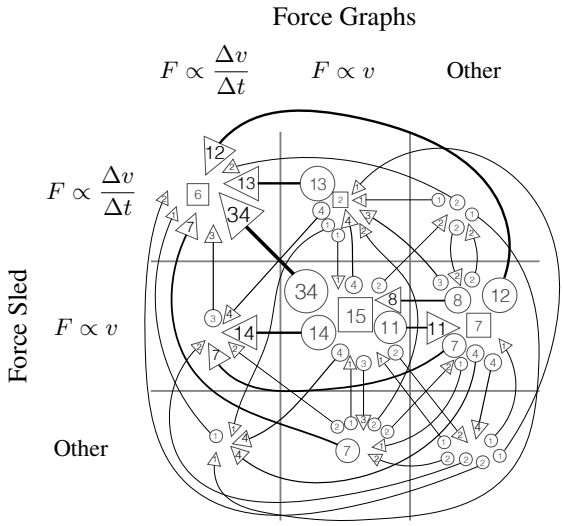

FIG. 4. Consistency plot for Case 3: moving right, slowing down ($N = 194$)

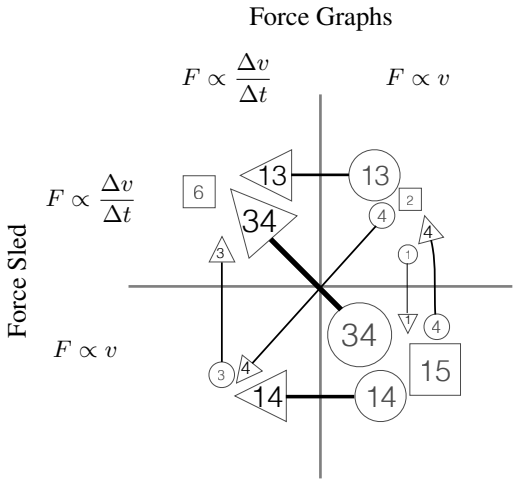

FIG. 5. Simplified consistency plot for Case 3 ($N = 94$)

and speeding up at a steady rate. These tables show that most students use the common incorrect model ($F \propto v$) before instruction, that most students use the correct model after instruction, and that moderate to strong correlations exist between the responses to the FS and FG questions ($\phi_c = 0.21$ on the pretest and 0.37 on the posttest); these results are consistent with previous analyses of these data [3]. However, these contingency tables also show that students become less consistent and less coherent after instruction: no cells in Table III are as large as the center cell in Table II, and 38 students (20%) answer the FG question correctly but use the common incorrect model to answer the FS question.

The consistency plot for Case 1 (Fig. 2) supports the contingency tables by showing that the most popular transition is from coherently using the common incorrect model before instruction to coherently using the correct model after instruction (82 students from center cell to upper left). Figure 2 shows the rich variety of students' learning trajectories, but this variety is also visually overwhelming. Figure 3 shows a simplified version of the Case 1 consistency plot that only includes students who used either the correct or common incorrect models ($N = 180$, over 90% of the students). This presentation provides greater clarity and allows more careful examination of the most consistent transitions. The second-most-popular transition in Fig. 3 is from coherently using the common incorrect model before instruction to using the correct model only on the FG question after instruction (30 students move left), with about as many staying in the coherently incorrect cell (29 students). This is in stark contrast to the relatively few students who only increase on the FS question: more students move left than up, indicating greater learning on the graphs questions.

IV. CASE 3: MOVING RIGHT, SLOWING DOWN

Tables IV and V respectively show the pre- and post-instruction contingency tables for Case 3: moving to the right and slowing down at a steady rate. Like Case 1, these tables show that most students use the common incorrect model ($F \propto v$) before instruction, that most students use the correct model after instruction, and that moderate to strong correlations exist between the responses to the FS and FG questions ($\phi_c = 0.22$ on the pretest and 0.33 on the posttest). However, consistency is much lower on these tables than on those for Case 1, with the largest cell containing only 89 students (46%). Furthermore, many more students select Other Incorrect responses than in Case 1, particularly before instruction.

The consistency plot in Fig. 4 supports these findings and clearly shows that students are much less coherent and consistent than they were on Case 1: this plot has a greater variety of learning trajectories than Fig. 2, and more students are in

TABLE IV. Pretest contingency table for Case 3 (moving right, slowing down): number of students who gave each response pair. A moderate correlation exists between the FS and FG questions: $\chi^2(4,\ N = 194) = 18.09,\ p < 0.001,\ \phi_c = 0.22$.

		Force Graphs		
		$F \propto \Delta v/\Delta t$	$F \propto v$	Other
Force Sled	$F \propto \Delta v/\Delta t$	6	19	5
	$F \propto v$	3	89	48
	Other	1	14	9

TABLE V. Posttest contingency table for Case 3 (moving right, slowing down): number of students who gave each response pair. A strong correlation exists between the FS and FG questions: $\chi^2(4,\ N = 194) = 43.32,\ p < 0.001,\ \phi_c = 0.33$.

		Force Graphs		
		$F \propto \Delta v/\Delta t$	$F \propto v$	Other
Force Sled	$F \propto \Delta v/\Delta t$	80	13	4
	$F \propto v$	27	26	23
	Other	10	5	6

the off-diagonal cells. Figure 5 shows a simplified consistency plot that eliminates students who chose Other Incorrect responses. The basic shapes and patterns on Fig. 5 are similar to Fig. 3, but due to the lower consistency and coherence of Case 3, fewer students are represented on the simplified plot ($N = 94$, only 48%). Even with this limitation, one can still see that the most popular transition (on Figs. 4 and 5) is from a coherently common incorrect response pair to a coherently correct response pair (34 students from center cell to upper left). The second-most-popular transition is from coherently common incorrect to correct on the FG question only (14 students move left), with about as many stubbornly staying in the coherently incorrect cell (15 students). Again, relatively few students improve on the FS question without also improving on the FG question: like Case 1, more students move left than up, again indicating greater learning on the graphs questions.

The simplified consistency plot allows a clearer comparison between some of the most popular transitions, but in Case 3, several popular transitions start or end within an "Other" cell. Many students begin the term selecting the common incorrect response on the FS question but an Other response on the FG question (circles in the right middle cell): 12 students transition to being correct on both, 7 become correct on the FG question, 8 transition to using the common incorrect answer for both questions, and 7 stay in the same cell. Moreover, 11 students begin the term choosing the common incorrect answer for both questions but transition to choosing an other response on the FG question. There are certainly some common patterns between Cases 1 and 3, but the lower consistency on Case 3 makes it very difficult to generalize these results and claim that these patterns are robust.

V. INTERPRETATIONS AND IMPLICATIONS FOR FUTURE WORK

Data from the contingency tables and consistency plots show that students do not necessarily answer FS and FG questions coherently. Students are also not always consistent with their classmates. These effects are more prominent in Case 3: student responses to the more complicated motion (slowing down, acceleration opposite velocity) yield a much more complicated consistency plot due to lower coherence and consistency among students. However, medium correlations exist between the pretest responses ($\phi_c \approx 0.21$) for both Cases, and large correlations exist between the posttest responses ($\phi_c \approx 0.35$).

More students select incorrect responses to the FG questions than the FS questions before instruction: for Case 1, 26 are correct on FS and only 18 on FG, while for Case 3, 30 are correct on FS and only 10 on FG. This agrees with previous literature on students' difficulties with graphs [4]. However, a more interesting result is that for both Cases more students improved on the FG question than on the FS question. Students seem to understand the graphs better than the descriptions. Additional data from Cases 2 and 4 show similar trends.

Several questions remain. Why are coherence and consistency so much lower on Case 3? Additional analyses should look at which Other Incorrect responses the students are choosing. Perhaps visual cues (on the FG questions) could explain some of this variability.

Why do students improve more on the FG questions than the FS questions? One possible factor is that the FS questions are the first on the FMCE, and the FG questions are in the middle. A simple test of this hypothesis would consist of switching the order of the questions and looking for any differences in the results. Another possibility is that this particular course emphasized student learning of graphs. Data do not exist to support or refute this hypothesis, but other similar datasets with known emphasis (or lack thereof) on graphs could provide touchstones for comparative analysis. Another possibility is that the verbal descriptions of forces in the FS cluster cues students to think about their day-to-day lives, while the graphs in the FG cluster cue students into a mode of thinking about physics class. Previous research has shown that students often answer differently when asked what a scientist would say compared to what they actually believe [11]. Comparing student responses to the FMCE with data regarding their attitudes and beliefs about physics could provide evidence to support or refute this interpretation.

ACKNOWLEDGMENTS

I am indebted to the faculty in whose classes data were collected. I wish to thank Ashley Smith for the inspiration for this project, and Andrew Heckler and David Meltzer for useful conversations about this work.

[1] R. K. Thornton and D. R. Sokoloff, Am. J. Phys. **66**, 338 (1998).
[2] T. I. Smith and M. C. Wittmann, Phys. Rev. ST Phys. Educ. Res. **4**, 020101 (2008).
[3] T. I. Smith, M. C. Wittmann, and T. Carter, Phys. Rev. ST Phys. Educ. Res. **10**, 020102 (2014).
[4] For example, R. J. Beichner, Am. J. Phys. **62**, 750 (1994)
[5] P. Nieminen, A. Savinainen, and J. Viiri, Phys. Rev. ST Phys. Educ. Res. **6**, 020109 (2010).
[6] This is a more detailed analysis of data originally reported in Ref. [3]: matched pre/post students at School 2.
[7] R. Rosenblatt and A. F. Heckler, Phys. Rev. ST Phys. Educ. Res. **7**, 020112 (2011).
[8] T. Coladarci, C. D. Cobb, E. W. Minium, and R. C. Clarke, *Fundamentals of Statistical Reasoning in Education* (John Wiley & Sons, Inc., 2004).
[9] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Lawrence Erlbaum Associates, 1988).
[10] M. C. Wittmann and K. E. Black, Phys. Rev. ST Phys. Educ. Res. **10**, 010114 (2014).
[11] T. L. McCaskey, M. H. Dancy, and A. Elby, AIP Conf. Proc. **720**, 37 (2004).