What can normalized gain reveal about individual learning on the FCI?

Andrew Pawl

Engineering Physics Department, University of Wisconsin-Platteville, Platteville, WI 53818

This work investigates whether the class-average normalized gain score is a useful measure of individual student learning on the Force Concept Inventory. Average normalized gain emerges as a poor description of the learning of students who enter the author's mechanics classes with pretest scores less than 20, but a reasonable description of the learning of those with pretest scores of 20 or more. This pretest threshold prompted a study of the impact of misconceptions on gain. Among the author's students, it appears that those exhibiting certain key misconceptions or failing to exhibit certain core skills on the pretest will have average normalized gains lower than the average among students who do not exhibit the misconceptions or do exhibit the skills. A future study will investigate whether early intervention explicitly addressing key misconceptions and core skills makes the class-average normalized gain score a better description of individual gains.

PACS numbers: 01.40.Fk, 01.40.G-

Keywords: force concept inventory, introductory course reform

I. BACKGROUND AND PRELIMINARIES

The Force Concept Inventory (FCI) [1] is a widely-used assessment for mechanics courses. One reason for its widespread use is the study by Hake [2] which introduced the concept of average normalized gain $\langle g \rangle$, defined as

$$\langle g \rangle = \frac{\langle \text{posttest score} \rangle - \langle \text{pretest score} \rangle}{\text{maximum possible score} - \langle \text{pretest score} \rangle}, \quad (1)$$

where $\langle \rangle$ denotes a class average. Hake showed that physics courses which featured interactive engagement consistently outperform lecture by this measure. This study popularized the use of average normalized gain on the FCI as a metric for evaluating the impact of instructional methods on physics courses. This raises the question of whether the concept of average normalized gain is equally useful in describing the impact of a course on the individual students. One way to investigate that question is by plotting the score shift (posttest score - pretest score) versus pretest score for each individual in a course. Figure 1 shows such a plot for 213 students who took a one-semester calculus-based mechanics course from the author at some point during the period from Fall 2011-Fall 2013. If the average normalized gain for the course is a good indicator of the learning of individual students, then this plot should resemble a line with the absolute value of the slope equal to the normalized gain and an x-intercept of 30 (the maximum possible score). Taking the data of Fig. 1 as a whole, however, it is difficult to believe that such a line is a good description of the distribution.

To illustrate this point, consider two separate divisions of the data set. Looking only at the 29 students who scored 20 or better on the pretest, the data would resemble Fig. 2. For this subset, a linear fit provides a reasonable description of the data, explaining 57% of the variance. The x-intercept is 29 ± 1 , which is consistent with 30. The slope of the fit is -0.67 ± 0.11 , which is statistically consistent with the average normalized gain of 0.56 ± 0.06 that results when Hake's formula is applied to this subgroup. Thus, the line which describes individ-

FIG. 1: Score shift (posttest - pretest) on the FCI vs. pretest score for 213 students. Scores are randomly adjusted by < 0.2 to avoid stacking repeated points. The dashed line is the model implied by the class-average normalized gain. Shaded regions lie outside the range of possible FCI scores.

ual learning is statistically consistent with the concept of an average normalized gain, and it would be reasonable to say that the students who enter the course with FCI pretest scores of 20 or more learn approximately 60% of the material they did not know on the pretest.

Figure 3, by contrast, shows only the subgroup consisting of the 184 students that scored less than 20 on the pretest. Here the linear fit is a poor description of the data, explaining about 3% of the variance. The slope of the best fit is -0.17 ± 0.07 and the x-intercept is 47 ± 8 . The x-intercept is not consistent with 30 and the slope is not very consistent with the computed average normalized gain of 0.32 ± 0.02 . It would be misleading to say that the students who enter the course with FCI pretest scores less than 20 learn approximately 30% of the material they did not know on the pretest, because the scatter in the observed individual learning rate is too large to justify that claim as representative of most individuals.

FIG. 2: Score shift vs. pretest score for the 29 students with pretest scores of 20 or more. Scores are randomly adjusted by < 0.2 to avoid stacking repeated points. The solid line is the best fit to the data. The dashed line is the model implied by the average normalized gain for these students.

FIG. 3: Score shift vs. pretest score for the 184 students with pretest scores less than 20 (the “core sample”). Scores are randomly adjusted by < 0.2 to avoid stacking repeated points. The solid line is the best fit. The dashed line is the model implied by the average normalized gain.

Comparing Figs. 2 and 3 leads to the conclusion that, in the author’s course, the concept of an average normalized gain is a reasonable way to describe the learning of individuals with pretest scores of 20 or more, but a poor way to characterize the learning of individuals with pretest scores under 20. One possible explanation for this split that bears investigating is whether there are certain key misconceptions that are not well addressed in the author’s course. (For a discussion of the alternate possibility that the effect is due to a correlation of high pretest scores with stronger science reasoning ability see [3].) The presence of key misconceptions could result in the split observed if students who come in without these misconceptions are able to learn the remaining concepts at a standard rate dictated by the structure of the

course (and thus achieve consistent gains), but students who enter the course holding one or more key misconceptions are prevented from learning on a significant group of questions at the same rate as those who do not hold the misconception (and thus achieve reduced gains). Students achieving high pretest scores would be unlikely to harbor the key misconceptions, while students achieving low pretest scores would likely hold one or more key misconceptions and therefore show varying levels of learning (depending on exactly how many were held or the impact of those that are held). The remainder of the paper will investigate this hypothesis.

II. SAMPLE AND METHODOLOGY

A. The Core Sample

The core sample for this study consists of the 184 students from five separate sections of calculus-based mechanics taught by the author at the University of Wisconsin-Platteville (a small state school with an engineering program) over four semesters (Fall 2011-Spring 2013) who took the FCI both pre- and post-instruction and who scored less than 20 on the pretest. The core sample is a subgroup of the full sample of 213 students who took the FCI both pre- and post-instruction. Most of the students were pursuing a major in engineering.

B. Relating Pretest Responses to Gain

In an attempt to determine if any key misconceptions are influencing gain, the average normalized gain was computed for subgroups of the core sample defined by pretest responses to each question on the FCI. For each question there are 5 possible subgroups, corresponding to responses A through E.

The core sample offers only marginal statistics when subdivided by pretest responses. For this exploratory investigation, discrepancies were deemed significant enough to investigate if they corresponded to a statistical confidence level $p < 0.2$. To avoid extremely small number statistics, responses that were given by fewer than 30 students (below the guessing rate for 184 students) were not analyzed. The data are shown in Table I. The interesting pretest responses fall into three categories, which are discussed below.

III. RESULTS AND ANALYSIS

A. Incorrect Responses Related to Below Average Gain: Two Basic Concepts Neglected

Three questions contain choices that appear related to below-average normalized gain, denoted by bold values less than 0.32 in Table I. Question 10 asks about the speed of a puck that has just received an impulsive kick. The students who responded that the puck’s speed will increase “for a while” and decrease thereafter (choice D) had an average normalized gain that was less than the class-average normalized gain. Question 12 illustrates

Question #	9	10	12	14	19	20	26	29
Gain for Subgroup A	-	0.34(3)*	-	0.28(4)	-	0.29(4)	0.28(4)	-
Gain for Subgroup B	0.28(4)	-	0.33(3)*	-	-	-	0.36(3)	0.30(3)*
Gain for Subgroup C	0.28(5)	-	0.26(4)	-	-	0.31(3)	-	-
Gain for Subgroup D	-	0.24(4)	-	0.36(3)*	0.27(3)	0.38(5)*	-	0.37(4)
Gain for Subgroup E	0.38(4)*	-	-	-	0.36(3)*	-	-*	-

TABLE I: Normalized gain computed for subgroups defined by pretest responses. Columns correspond to question number and rows to responses A-E. Only questions that yielded disparities significant at the $p < 0.2$ level from the core sample overall average of 0.32 ± 0.02 are reported, with the specific gains that differ significantly indicated in boldface. Numbers in parentheses denote uncertainty in the last digit. Dashes denote subgroups of less than 30 students. Asterisks designate the correct answer.

five possible trajectories for a horizontally-launched cannon ball. Students who believed that the action of the cannon causes the ball to coast in a straight line for a brief time before gravity takes over (choice C) exhibited reduced gain on average. Question 19 asks students to compare the speed of two objects whose motions are represented in a motion diagram. Students who believed that the objects would have the same *speed* when they were at the same *position* (choice D) exhibited reduced gain on average.

A total of 79 students from the core sample gave at least one of these three low-gain-related responses on the pretest. This represents almost half the core sample, or 37% of the full sample of 213 students who had matched FCI responses. (Of the 29 students who scored 20 or more on the pretest, only one student gave any of the three responses mentioned here.) Thirty students in the core sample gave two out of three of these responses. No student in the sample gave all three.

These questions that appear related to below-normal gains deal with concepts that are so fundamental that the author has neglected to explicitly teach them as separate topics. The first concept is the immediate cessation of impulsive forces after contact is lost, which is probed by questions 10 and 12. (The contrary notion that force is retained by a moving object after contact is lost is termed the “impetus” misconception by Hestenes *et al.* [1].) The other is the relationship of velocity to position as probed by question 19. The data suggest that over one-third of the students entering the author’s mechanics courses are uncomfortable with one or both of these concepts, and that these students are at a disadvantage when it comes to learning from the course. Neglecting to explicitly teach motion diagrams as a representation of motion appears to be particularly dangerous, since about 60% of the students who responded with option D for question 19 on the pretest chose D again on the posttest. The questions dealing with impulsive forces do show learning, however, as 60% of those starting with answer D on question 10 and 73% of those starting with response C on question 12 change to the correct answer on the posttest.

B. Correct Responses Related to Above Average Gain: Strong Math/Physics Connection

There are four questions for which choosing the correct response on the pretest appears related to above-average

normalized gain, denoted by boldface and asterisked values above 0.32 in Table I. Question 9 concerns a puck that has just received an impulsive kick (same situation as question 10). The question tests knowledge of vector addition by asking how the speed of the puck after the kick relates to the speed before the kick. Since the kick is at right angles to the initial velocity, the Pythagorean theorem must be employed to obtain the correct answer. Question 14 asks about the trajectory of a bowling ball that falls from a moving airplane. Arriving at the correct answer requires that students be able to visualize the motion from the stationary reference frame of the ground rather than the moving reference frame of the airplane. Questions 19 and 20 ask students to compare the speed and the acceleration (respectively) of two objects whose motions are represented in a motion diagram. Obtaining the correct answers require students to estimate the first and second time derivative (respectively) of the position based on reading a motion diagram.

Of the 184 students in the core sample, only 15 (8%) answered all four of these questions correctly on the pretest. 53 (29%) answered three of the four correct, 108 (59%) had two correct and 161 (88%) had at least one right. By contrast, of the 29 students with pretest scores over 20, 19 students (66%) had all four of these questions right, 28 (97%) had at least three right and all the students had at least two of the four correct.

These questions for which correct answers appear related to above-average gains require the ability to see connections between mathematical concepts and physical situations. Students are asked to realize how the Pythagorean theorem can be applied in a novel context (question 9), to be able to shift coordinate reference frames (question 14) or to understand the application of differential calculus to a motion diagram (questions 19 and 20). The data reported here suggest that less than half of the students entering the author’s mechanics classes are comfortable with these skills. The author has already made early instruction in vector math a priority in his classes by using a forces-first instructional approach that begins the course with a week on vector equilibrium of forces. As discussed above, however, motion diagrams have not been covered in detail. Transformation of reference frames is usually not formally covered, but is discussed conceptually in the unit on circular motion and centripetal force. The lack of formal instruction on reference frames shows up in the fact that only 29% of the

students in the core sample who gave incorrect responses to question 14 on the pretest change to the correct answer on the posttest. For the other three questions, between 43% and 45% of the students who initially gave incorrect responses shifted to the correct response on the posttest.

C. Incorrect Responses Related to Above Average Gain: A Subtle Difference

There are two questions for which choosing a particular *incorrect* response appears related to an average normalized gain that is *larger* than the core sample average normalized gain, denoted by boldface values with no asterisk that are larger than 0.32 in Table I. One of these, question 29, will not be discussed in detail here. Question 29 concerns the forces on a chair at rest, but confuses the issue by asking about the effect of the atmosphere. This makes the question difficult to analyze in the context of the rest of the FCI, as has been remarked by other researchers [4]. Question 26, on the other hand, probes student understanding of Newton's 1st and 2nd laws in a manner that is central to the mechanics curriculum and related to several other questions on the FCI.

Question 26 deals with a box that is being pushed along the ground. The box is initially being pushed at constant speed. The question asks what happens to the speed of the box if the pushing force is doubled. The vast majority of the students (74%) entering the author's classes believe that the box will shift to a faster constant speed. Only seven students of the core sample of 184 students and only 20 of the entire sample of 213 students correctly answered that the box would move with continuously increasing speed on the pretest. Interestingly, however, the students are offered two possible choices for the new faster constant speed. Choice A says the box's speed will precisely double when the force is doubled. Choice B says that the speed will exceed the original speed, but will not be exactly twice as great. The choices are initially approximately equally popular among the core sample, with 69 choosing A on the pretest and 79 choosing B. Students who choose B on the pretest have larger normalized gains on average than the core sample overall average.

Choice A seems to correspond to a very naive conception of motion, which only acknowledges the influence of applied force (termed the "active force" misconception by Hestenes *et al.* [1]). Choice B, however, appears to recognize that friction must also be considered, and therefore

acknowledges the fact that overall motion is the result of the combined action of all forces acting on an object. Even though neither choice indicates understanding of Newton's 1st or 2nd laws, choice B is a step closer to the Newtonian viewpoint than choice A. This step is apparently a significant one from the perspective of preparing students to learn in the course. Neither choice, however, is associated with strong learning on question 26 itself, with more than 60% of the students who entered answering A or B on the pretest giving one of those two answers on the posttest.

IV. CONCLUSIONS AND FUTURE WORK

The technique of examining the relationship of gain to specific pretest responses has uncovered a surprisingly narrow set of misconceptions and skills that appear related to suppressed or enhanced student learning in the author's mechanics courses. At this stage, however, it is unclear if the results indicate that these misconceptions and skills are actually preventing or enabling learning (and hence *causing* the adjusted gains) or that they are simply correlated with gain in some way.

Given that only one of the skills of interest (vector addition) is currently a central topic in the author's curriculum, it should be straightforward to design experiments to test for causal relationships. An early intervention on impulsive forces could take the form of a video analysis activity using high-speed video footage filmed in class or recovered from an internet repository (see, *e.g.*, [5]). Tutorials on multiple representations have long been a recommendation of physics education researchers (*e.g.* [6]) and could be adapted from various curricula (*e.g.* [7, 8]). To push students to question the most naive view of the relationship between force and motion, question 26 could be explicitly adapted to a laboratory activity asking students to find the force needed to cause a motion cart to travel with constant speed, then double that force and observe the effects with a motion sensor.

If one or more of these interventions result in a better correspondence between individual gains and the class-average gain, then the approach of looking for relationships between gain and pretest responses will have proven useful as a diagnostic which helps the instructor enable all students to learn effectively in the course.

ACKNOWLEDGMENTS

This work was supported by a Scholarly Activity Improvement Fund grant from UW-Platteville.

-
- [1] D. Hestenes, M. Wells and G. Swackhamer, *Phys. Teach.* **30**, 141-158 (1992).
 - [2] R. Hake, *Am. J. Phys.* **66**, 64-74 (1998).
 - [3] V.P. Coletta and J.A. Phillips, *Am. J. Phys.* **73**, 1172-1182 (2005).
 - [4] M. Planinic, L. Ivanjek and A. Susac, *Phys. Rev. ST Phys. Educ. Res.* **6**, 010103 (2010).
 - [5] <http://serc.carleton.edu/sp/library/dmvideos/>
 - [6] J.I. Heller and F. Reif, *Cognition Instruct.* **1**, 177-216 (1984).
 - [7] <http://modelinginstruction.org>
 - [8] A. Van Heuvelen and E. Etkina, *The Physics Active Learning Guide* (Addison Wesley, San Francisco, CA, 2006).