PHYSICAL REVIEW SPECIAL TOPICS - PHYSICS EDUCATION RESEARCH **5**, 010105 (2009)

Comparing the force and motion conceptual evaluation and the force concept inventory

Ronald K Thornton
Center for Science and Mathematics Teaching, Tufts University, Medford, Massachusetts 02155, USA

Dennis Kuhl
Marietta College, Marietta, Ohio 45750, USA

Karen Cummings
Southern Connecticut State University, New Haven, Connecticut 06511, USA

Jeffrey Marx[*]
McDaniel College, Westminster, Maryland 21157, USA

(Received 12 April 2007; revised manuscript received 14 July 2008; published 20 March 2009)

In this paper we compare and contrast student's pretest/post-test performance on the Halloun-Hestenes force concept inventory (FCI) to the Thornton-Sokoloff force and motion conceptual evaluation (FMCE). Both tests are multiple-choice assessment instruments whose results are used to characterize how well a first term, introductory physics course promotes conceptual understanding. However, the two exams have slightly different content domains, as well as different representational formats; hence, one exam or the other might better fit the interests of a given instructor or researcher. To begin the comparison, we outline how to determine a single-number score for the FMCE and present ranges of normalized gains on this exam. We then compare scores on the FCI and the FMCE for approximately 2000 students enrolled in the Studio Physics course at Rensselaer Polytechnic Institute over a period of eight years (1998–2006) that encompassed significant evolution of the course and many different instructors. We found that the mean score on the FCI is significantly higher than the mean score on the FMCE, however there is a very strong relationship between scores on the two exams. The slope of a best fit line drawn through FCI versus FMCE data is approximately 0.54, and the correlation coefficient is approximately $r=0.78$, for preinstructional and postinstructional testings combined. In spite of this strong relationship, the assessments measure different normalized gains under identical circumstances. Additionally, students who scored well on one exam did not necessarily score well on the other. We use this discrepancy to uncover some subtle, but important, differences between the exams. We also present ranges of normalized gains for the FMCE in a variety of instructional settings.

DOI: 10.1103/PhysRevSTPER.5.010105 PACS number(s): 01.40.Fk

I. INTRODUCTION

The motivation for this paper is to compare and contrast several aspects of the force concept inventory (FCI) with the force and motion conceptual evaluation (FMCE). We hope this information will allow instructors and researchers to make an informed decision about which of these tools might be most appropriate for their specific assessment goals. We begin with a brief history of the two exams and summary of their respective domains and representational formats. We then report FMCE gains for college and university students in both traditionally structured courses and courses which make use of research-based instructional materials or interactive methods. The bulk of the paper is devoted to comparing performance on the FMCE and FCI for studio physics students at Rensselaer Polytechnic Institute. We do this to highlight interesting similarities and differences between these two exams. We conclude by arguing that both exams serve different, but important roles in the physics education community.

Since its publication in 1992, the FCI has been widely used to help demonstrate the need for improving students' conceptual understanding of mechanics and to evaluate the effectiveness of reforms intended to accomplish this.[1] Most notably, in 1999 the physics education community was presented the "Hake plot" that compared FCI normalized gains for over 6,000 students in a wide range of instructional settings.[2] In the late 1980s, Sokoloff and one of the authors (Thornton) developed the FMCE to specifically identify Newtonian and common student views of force and motion in one dimension by analyzing student responses (including "wrong" responses) within specific clusters of items on the exam. The FMCE was published in 1998.[3]

The domains of the two diagnostic exams are described in previous publications. See, for example, Refs. 1,3. For the purposes of this paper, we note that both exams cover one-dimensional kinematics and Newton's laws. Additionally, the FCI includes the following topics: two-dimensional motion with constant acceleration, which implies parabolic motion; impulsive forces; vector sums; cancellation of forces; and identification of forces. For this discussion, it is interesting to note that an examinee who had a poor understanding of the content covered by the FMCE could score above the historically cited 60% "Newtonian threshold" on the FCI by correctly answering 82% of the 22 FCI items that are outside the domain of the FMCE.[1] In addition to probing different topics to various depths, both exams utilize several representational formats. The FCI largely uses a combination of verbal and pictorial representations, while the FMCE relies on verbal and graphical representations.

II. PROCEDURE

During the 1990s we collected matched pretest and post-test data on FMCE scores for college and university students undergoing physics instruction employing various instructional methods and calculated the normalized gains on the class averages. Separately, starting in the Spring of 1998 and running through the Spring of 2006, we collected data by matching FCI to FMCE scores for students in the studio physics course at Rensselaer Polytechnic Institute. We did this both preinstruction and postinstruction.

Since 1993, Rensselaer has taught introductory physics in a "studio" environment in which classes of 30–45 students meet for integrated lecture-recitation-laboratory sessions.[4,5] During the spring 1998 semester, several sections of studio physics were additionally supplemented with curricular materials and methods based on the outcomes of physics education research,[6] namely, Interactive Lecture Demonstrations (ILDs)[7] and cooperative group problem solving (CGPS).[8] Since then, the leadership of the course and the curriculum has changed several times. These instructional changes provided a fruitful opportunity for comparing the FCI and the FMCE in a variety of similar yet distinct instructional settings.

The FCI and the FMCE were administered to all students taking studio physics as both a preinstructional and postinstructional evaluation. We labeled the two diagnostics "part A" and "part B" of a single exam packet. We administered the pretest during the first class session and the post-test about 15 weeks later near the end of the semester. The students were allowed up to 25 min for the FCI and 35 min for the FMCE. These times reflect the ratio of the number of items on each part (30 for the FCI and 47 for the FMCE) to the total number of items, multiplied by the 60 min of available time. These timeframes are less time than both exams' authors recommend, but most students finished in the time allowed. For the purposes of this study, students with more than four blanks at the end of either evaluation were considered to have failed to finish, so we dropped them from the analysis. This amounted to less than 1% of our total population.

III. RESULTS AND DISCUSSION

A. Creating single-number scores

The FMCE was not originally designed to have results analyzed with a single-number score, but to begin our comparison, we felt it necessary to create such a score for the exam. The FCI scores discussed here are simply the ratio of the number of correct responses to the maximum possible (30).

The FMCE scores we describe here are the ratio of points accumulated out of 33 possible points. The 33 points are a composite of the FMCE's first 43 items. This scoring rubric is advocated by the test's designers to best characterize a student's mastery of the concepts of motion and force in one dimension. Note that the final four items of the exam, not considered here, deal with mechanical energy and were not included in the published version of the FMCE.

TABLE I. FMCE normalized gain for more than 3000 students at ten colleges and universities divided into two categories: traditional instruction and research-based methods. Three institutions are included in both categories. Uncertainties shown are standard errors in the mean with N equal to the number of classes.

	N		Normalized Gain		
	Students	Institutions	Low %	Average %	High %
Traditional	926	4	8	15 ± 3	22
Research-based	2494	9	33	63 ± 6	93

As discussed elsewhere FMCE items 5, 15, 33, 35, 37, and 39 are frequently answered "expertly" by students even before they are consistently Newtonian thinkers.[3,9] For example, most non-Newtonian students believe that if an object is standing still, there is no (net) force on it (item 15) and that two identical objects that move at the same speed and collide will exert equal and opposite forces (item 33). Third law items 35, 37, and 39 are useful for determining student views if connected with other second and third law questions, but are often answered "expertly" for the wrong reasons before students understand the third law. Although these questions are helpful in some contexts, they are *not included* in these analyses. We also omitted from our study item 6, which is sometimes answered incorrectly, even by experts. The authors of the FMCE argue that all of these items are useful, however, only when one relates them to "right" and "wrong" responses on other items.

Items on the FMCE dealing with the acceleration of a tossed coin (27–29), the force on a tossed coin (11–13), and the force on a cart moving up and down a ramp (8–10) form three distinct clusters. These items are some of the most difficult for students to correctly answer. The test designers believe that students only understand the concepts involved if they answer all three correctly. Essentially all students will get at least the direction of the acceleration or force correct for one of the three segments of the motion of the object (on the way up, at the top, on the way down) even with the most naive view. For example, if they decide the acceleration or force is constant in each segment and not changing, they will also get one of the three cases correct, despite the fact that they still may not have a Newtonian view. Hence, when calculating a single-number score for the entire evaluation, we awarded two points if all three items in a given set were correct, zero otherwise. With the weighting we described, students who miss at least one item in each of three clusters can only achieve a maximum of 82% on the FMCE.

Approximately one-third of the points for the FMCE single-score evaluate knowledge of kinematics while the remaining two-thirds evaluate knowledge of Newton's laws. Each of the three laws receives roughly equal weighting on the exam.

B. Normalized gains on the FMCE

Table I shows the range and average of normalized gains

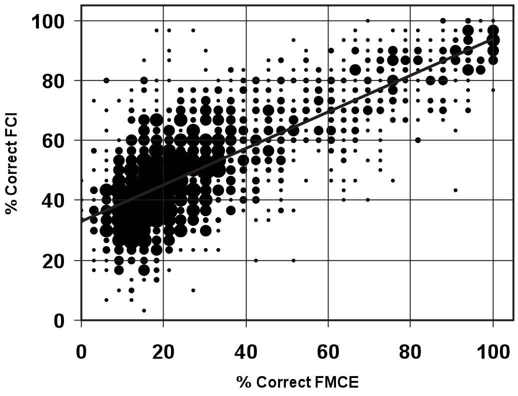

FIG. 1. A “bubble plot” of *Preinstruction* % correct on the FCI vs % correct on the FMCE. The size of the point indicates how many students had that pair of scores. There are 1617 students in this group. The slope of the best fit line through the data is 0.58 and the *y* intercept is 35. The correlation coefficient is 0.74. One student had a score of 20% on the FCI and 3% on the FMCE. Five students had a score of 97% on the FCI and a score of 100% on the FMCE.

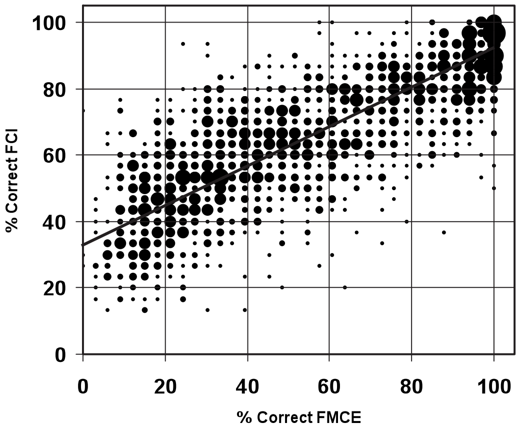

FIG. 2. A bubble plot of *Postinstruction* % correct on the FCI vs % correct on the FMCE. There are 1702 students in this group. The slope of the best fit line through the data is 0.52 and the *y* intercept is 36. The correlation coefficient is 0.78. One student had a score of 20% on the FCI and 3% on the FMCE. Twelve students had a score of 97% on the FCI and a score of 100% on the FMCE.

on the FMCE for more than 3000 students attending a variety of colleges and universities. We divided the classes into the usual two categories: classes that used more traditional instruction and classes that used research-based methods or materials. Our intention is to allow instructors and researchers some reference to which they might compare their single-score results on the FMCE. There are a total of ten institutions included in this data set. Three institutions are included in both categories. The uncertainty noted in the mean is the standard error in the mean, where *N* is the number of courses. As one can see from the standard errors, there was less variation among traditional instruction than there was among the research-based methods. This is no surprise given the diversity of research-based methods used in these courses, which included cooperative-group problem solving,[8] Interactive Lecture Demonstrations,[7] peer instruction,[10] tutorials in introductory physics,[11] RealTime Physics,[12] studio physics,[4–6] and workshop physics.[13]

C. Relationships

We now turn our attention to comparing the FMCE to the more widely used FCI. This comparison is done using only data on students from the studio physics courses at Rensselaer Polytechnic Institute. Figures 1 and 2 show “bubble” plots of an individual student’s FCI score versus the corresponding FMCE score (both as a percentage of the maximum possible score). The size of the bubble in these plots corresponds to the number of occurrences for each pair of scores. In each of the preinstruction and postinstruction plots, there is a strong relationship between the scores on the two exams. Specifically, the slope of the best fit line through the pretest data is 0.58 (%FCI/%FMCE) and the correlation coefficient is 0.74. The slope of the best fit line through the post-test data is 0.52 (%FCI/%FMCE) and here the correlation coefficient is 0.78. The scores generally fall toward the upper left region of the graphs, indicating that an individual’s FCI score tends to be higher than the FMCE score. The FMCE scores also show a greater difference between the lowest scores and the highest. There are 1617 students in the preinstruction group and 1702 students in the postinstruction group. The typical student is included in both of these groups.

The authors carefully evaluated these data in regard to the FCI/FMCE score comparisons, as well as in regard to the issues discussed in Secs. III D–III F below. Except where specifically noted in the discussion of Fig. 7 in Sec. III D, we found no pertinent differences between pre-test and post-test data. This includes comparisons between preinstructional data and postinstructional data and comparisons between data collected for groups of students who were engaged in different instructional techniques (e.g., CGPS, ILDs, modified RealTime Physics laboratories) within the studio physics environment. Hence, all the data are combined into a set of 3319 score pairs in Fig. 3 and the discussions that follow. Again, since some of these score pairs are pre-test scores and

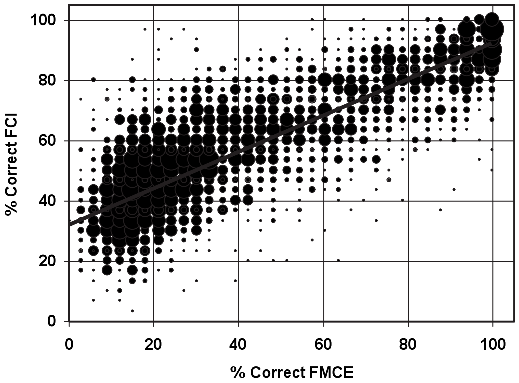

FIG. 3. A bubble plot of % correct on the FCI vs % correct on the FMCE for all scores. There are 3319 pairs of scores. The slope of the best fit line through the data is 0.54 and the *y* intercept is 36. The correlation coefficient is 0.78. Two students had a score of 20% on the FCI and 3% on the FMCE. Seventeen students had a score of 97% on the FCI and a score of 100% on the FMCE.

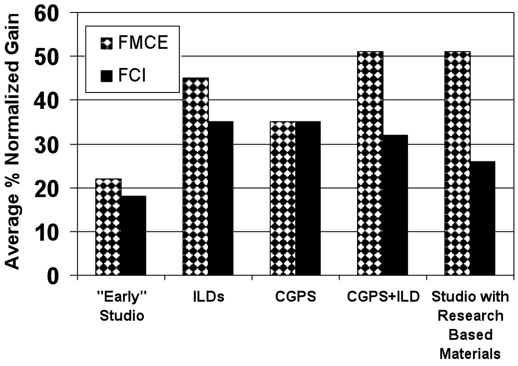

FIG. 4. FMCE and FCI normalized gains for groups of students having had various instructional experiences within the studio physics environment at Rensselaer. $N=182$ students for the "early studio" group, $N=101$ for the Interactive Lecture Demonstration group, $N=32$ for the cooperative group problem solving group, $N=30$ for the ILD+CGPS group and $N=450$ for studio with research-based materials.

some are post-test scores, most students are represented twice in this combined set. The slope of the best fit line through the combined data set is 0.54 (%FCI/%FMCE) and the correlation coefficient is 0.78.

All students who completed both the FCI and FMCE either pre-test, post-test, or both are plotted, over eight years of data collection and evolution of the studio physics course. In earlier years, two of the authors (Cummings and Marx) worked on the course and gains on both exams were in the "research-based methods" range, as shown in Table I. In later years, neither author was involved in the course and gains tended to be in the "traditional" range. More than 15 different instructors taught the students this data set represents. Despite these differences, we found that the relationship between the patterns of responses on the two exams was consistent as were other characteristics of the data (i.e., slope and correlation of %FCI vs %FMCE) that are important for this comparison. This indicates that for the studio physics courses at Rensselaer the relation between the exams is independent of the instructor or specific instructional materials or approaches to which the students were exposed.

On the other hand, the normalized gains are dependent on instructor and the specific instruction experienced by the students. Figure 4 shows normalized gains on the two exams for several instructional approaches used within the studio physics environment. We see, for example, students who experienced Interactive Lecture Demonstrations achieved relatively large normalized gains on the FMCE as compared to students in the "early" implementation studio physics course (e.g., fall, 1998). This is understandable since the FMCE focuses almost exclusively on one-dimensional motion and forces, which is the focus of the most of the ILDs performed. On the other hand, the FCI has relatively few items directly related to these topics. So, although the gains on the FCI were significant when compared to students who did not have the ILDs, they were not as large as the gains on the FMCE. We will explore these and other differences between the exams in more detail after examining student responses to specific questions.

D. Students with low FMCE but high FCI scores

Given the strong relationship between student scores on the FCI and FMCE, we expected that low FMCE scores are often paired with low-FCI scores, likewise high-FMCE scores pair with high-FCI scores. This is often the case; however, there are many cases where students score inconsistently on the two exams. We believe that these inconsistent-scoring students lend insight into differences between the exams. Since almost all Rensselaer students scored lower on the FMCE than on the FCI, we start by examining the FCI items that were frequently answered correctly when scores on the FMCE are low.

Evidence that student responses on the FMCE match their conceptual ideas is demonstrated by written conceptual descriptions, interviews, and responses to additional free-response conceptual questions as discussed in Refs. 3,9. In addition, Ramlo has recently formally established that the FMCE is a reliable and valid assessment instrument.[14] We assert that students who score below 40% on the FMCE are reliably non-Newtonian, but some may understand kinematics.

The combined pool of 3319 FCI/FMCE score pairs includes both pre- and postinstruction scores and 1810 of these have FMCE scores below 40%. Since this pool includes both pre-and postinstruction scores, some students may be represented in the 1810 score pairs twice. Nonetheless, we refer to these as "low-FMCE students." The average FMCE score for these 1810 low-FMCE students is 22%. The average FCI score for this group is 47%. Hestenes *et al.*,[1] claim that "there exists a kind of conceptual threshold near 60% on the FCI; below this threshold, a student's grasp of Newtonian concepts is insufficient for effective problem solving." We used these two metrics as a guide to gather our first subgroup of examinees: those with FMCE scores below 40%, but FCI scores above 60%. We refer to this group as "low-FMCE/high-FCI" students. A substantial number of students scored at or above 60% on the FCI, but scored below 40% on the FMCE. On the preinstruction evaluation there were 232 students who scored in the $\text{FMCE} < 0.4$ and $\text{FCI} > 0.6$ range, and there were 174 students in this range on the postinstruction evaluation. (Again, some students are included in both of these groups.) Since the item-by-item results for the two groups were consistent, we treated them as one group of 406 score pairs.

As already stated, interviews and free-response questioning indicate that students scoring below 40% on the FMCE do not understand Newton's laws as physicists. This is supported by Figs. 5(a) and 5(b), which show the results for each item on the FMCE for the 406 low-FMCE/high-FCI score pairs. We grouped items by their conceptual domain and, as one can see, the highest percentage of correct responses occurs for velocity questions. Students also do relatively well on acceleration questions. On the other hand, they do quite poorly on the questions related to Newton's first and second laws and the forces acting on the coin or cart on the ramp. This indicates that these students have an understanding of kinematics, but not a Newtonian understanding of dynamics as measured by this exam. A few items that would seem to evaluate knowledge of Newton's laws are answered

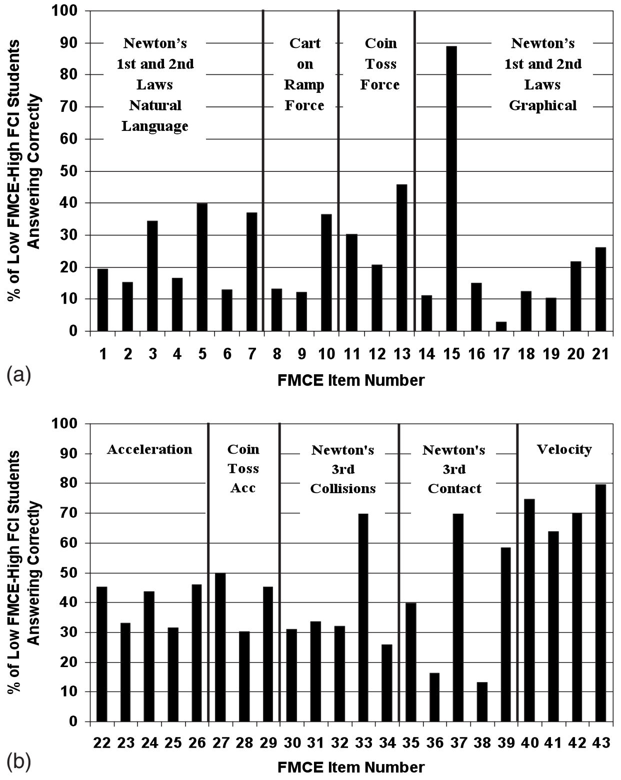

FIG. 5. [(a) and (b)] Results on individual FMCE items for students with 406 low on the FMCE (<0.4) and high on the FCI (≥0.6) score pairs. Students are taken from both the pre- and postevaluations.

correctly by most of these students. However, the items with anomalously high results—15, 33, 37, and 39—are not included in the single-number score for the FMCE because, as discussed above, students often answer these questions correctly regardless of their conceptual model.

Figure 6 is a plot of the students' performance grouped into conceptual categories as suggested by the FMCE's authors.[3] The percent correct values are calculated with items 5, 15, 33, 35, 37, and 39 dropped as discussed above in Sec. III A on creating a single-number score for the FMCE. Similarly, the items on the coin toss and cart on the ramp are treated as all-or-nothing units. The results shown in Fig. 6 are an alternative illustration of the idea that students scoring below 40% on the FMCE have some understanding of kinematics, but items on Newton's laws were more difficult for them.

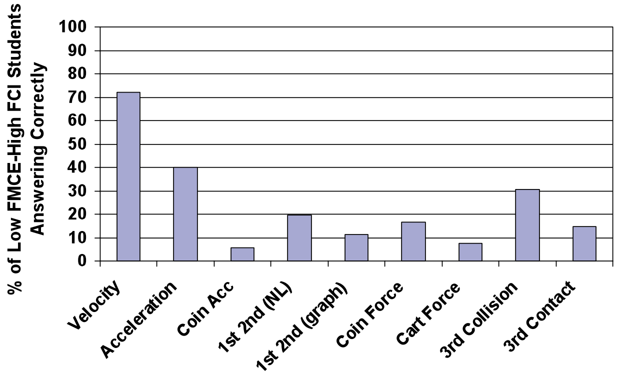

FIG. 6. (Color online) FMCE results grouped by conceptual category for students with 406 low on the FMCE (<0.4) and high on the FCI (≥0.6) score pairs. Students are taken from both the pre- and postevaluations. NL stands for "natural language," see Ref. 3 for more details.

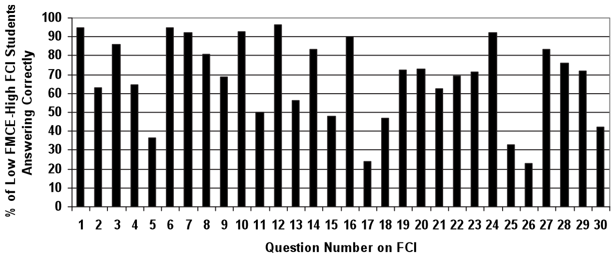

FIG. 7. Results on individual FCI items for students with 406 low on the FMCE (<0.4) and high on the FCI (≥0.6) score pairs. Students are taken from both the pre- and postevaluations.

Figure 7 shows how students with these 406 low-FMCE/high-FCI score pairs responded to items on the FCI. In contrast to all the other findings we discuss in this paper, the precise percentage of correct answers on a given FCI item did vary noticeably for some of these items if we separate out various subgroups (e.g., post-tests for students who experienced ILDs). However, that discussion is beyond the scope of this paper. Furthermore, we found that we could arrive at a consistent group of items that students with low-FMCE/high-FCI score pairs did well on, regardless of subgroup, if we set a threshold of 80% or higher correct answers. This group contains FCI items 1, 3, 6, 7, 8, 10, 12, 14, 16, 24, and 27. Table II lists these items with a short description of their content.

Since the low-FMCE/high-FCI students performed well on FMCE items pertaining to velocity (items 40–43), it is no surprise that they also scored well on FCI items 9, 10, and 24, which deal with the speed of an object. The two exams do differ in their approach to probing students' ideas about velocity. On the FCI, there is an explicit reference to the

TABLE II. 54 low-FMCE/high-FCI students scored > 80% on these FCI items.

FCI Item Number	Description
10, 24	Speed of an object after all forces have ceased. All motion is two dimensional.
1, 3	Gravity acting on a falling object. In item 1, students must identify that two metal balls of different weights fall to earth at approximately the same time. In item 2, they must answer that a stone speeds up because of the constant force of gravity.
6, 7, 8, 12, and 14	Trajectory of an object moving in two dimensions.
16	Newton's third law: car and truck in contact, moving at constant speed.
27	Motion of a block after the force pushing it is removed while friction slows it to a stop.

forces acting on the object having ceased prior to the time under consideration; on the FMCE, items pertaining to velocity are graphical. Interestingly, this does not seem to produce a difference in student performance.

There are many items on the FCI that simply probe a conceptual domain that is not covered by the FMCE, so students with low-FMCE and high-FCI score pairs could answer them correctly on the FCI, yet still have a low-FMCE score. Items 1 and 3 on the FCI check student knowledge regarding objects in free fall. Items 6, 7, 8, 12, and 14 all require the student to identify the trajectory of an object moving in two dimensions; the FMCE is restricted to one-dimensional motion. Item 19 requires students to make a qualitative observation about the relative speed of two accelerated objects. Finally, item 27 on the FCI requires students to identify friction as the force that slows an object with no other motive force. The FMCE does not ask students to identify forces. More importantly, though, we argue that this item can be answered correctly without a sophisticated Newtonian view.

Item 16 on the FCI poses a question about third law forces in a situation where a car and a truck are in contact and moving at a constant speed. Item 37 on the FMCE is very similar, and students who answer item 16 on the FCI tend to correctly answer item 37 on the FMCE. Based on interviews conducted during the development of the FMCE, we know students generally answer this item on the FMCE correctly by way of ignoring Newton's third law and applying an incorrect second law model ($a=0$, so $F^{\text{net}}=0$, so $F^{\text{on car}}=F^{\text{on truck}}$). Consequently, as discussed above, we omit item 37 from a single-number FMCE score.

E. Students with high FMCE scores

As one can see from Fig. 3, there are relatively few students who have a low-FCI score and a high-FMCE score. (We refer to this group as "low-FCI/high-FMCE.") If we chose to define the low-FCI/high-FMCE group in a manner symmetric to the low-FMCE/high-FCI group, we would look for score pairs with FMCE scores above 60% and FCI scores below 40%. There are only six score pairs (out of 3319) which fall into this category. Recall that there were over 400 in the symmetrically defined low-FMCE/high-FCI group. It is interesting to see so few students in this subgroup, and we infer from this that students who score high on the FMCE are likely to do well on FCI, which includes correctly answering items on the FCI that are outside the domain of the FMCE. It is important to note though, that in contrast to the FCI, the research community has yet to formally established a "high" score threshold for the FMCE. So, our symmetric definition above is arbitrary.

We can also look at all "high" scores on the FMCE, regardless of FCI scores, and see which items on the FCI were most difficult for these students to answer. If we arbitrarily consider 90% correct to be a "high" score on the FMCE then just about 10% of the 3319 score pairs discussed here fall into this range (specifically, 376 score pairs in this sample). If we look for FCI items on which this subgroup scored less than 80% (a previously chosen, arbitrary threshold), we identify six items.

Students with high-FMCE scores identified the forces as equal and opposite for a car pushing a truck and speeding up to cruising speed (FCI item 15) only 74% of the time, but in general the same students missed a very similar question on the FMCE (item 36, 72% correct). Both evaluations are uniform in this respect, and we know from our research that this is one of the last items on the FMCE to be answered correctly.

In the high-FMCE group, students were able to identify all the forces on an elevator being pulled at constant speed by a cable (FCI item 17) only 54% of the time. The fact that students who score well on the FMCE miss this item on average more than other items is consistent with the fact that the FMCE does not measure students' ability to identify multiple forces acting on an object. Along those same lines, high-FMCE students correctly identified the forces on an object moving in a circular channel (FCI item 5) 75% of the time and only 71% could identify the forces acting on a chair at rest on the floor (FCI item 29).

The last two items on the FCI that the high-FMCE group had trouble with are item 25 (74% correct) and item 26 (74% correct). These two items probe student understanding of the nature and relationship between force and velocity for the case of a woman pushing a box across a floor in the presence of a frictional force. As Fig. 7 indicates, these items are among the most challenging on the FCI.

F. Comparison of gains on the FMCE and the FCI

We saw earlier from Fig. 4 that the gains on the FCI and the FMCE were not identical across various instructional experiences. Our examination of student responses to specific items on both evaluations helped us understand the differences. Since the FMCE primarily measures the understanding of kinematics and Newton's laws in one dimension we should expect large gains on the FMCE for students who go from non-Newtonian to Newtonian thinking during instruction. The FCI has relatively few items evaluating the relationship between force and motion so we might expect the gains to be smaller, if having students attain a Newtonian worldview is the primary focus of the course. For example, many students who experience the mechanics ILD sequence make relatively large gains in understanding Newton's laws. Hence, we might expect the larger gains on the FMCE for these students that are shown in Fig. 4.

In order to generalize this comparison between gains on the two exams, we will consider two specific sets of students. First, consider students who begin with a weak understanding of force and motion in one dimension, but after instruction are very strong in this domain. Identifying these students as ones whose pretest scores on the FMCE are below 40% and whose post-test score is over 80% automatically creates a high-normalized gain group. We found 46 students in this group; they had an averaged normalized gain of 86% (raw gain 65%) on the FMCE. These same students have a smaller averaged normalized gain of 51% (raw gain 23%) on the FCI. Interestingly, half of these students began with a FCI score over 60%, and, this subset of students (with an average normalized gain of 45% on the FCI, and raw gain of 16%)

only answered an average of about four additional FCI items correctly (or 13% more items). In contrast, these students correctly answered about 20 additional items on the FMCE postinstruction (or 43% more items).

Next, consider students who begin low on both evaluations but do not become Newtonian during instruction. As discussed above, students who are non-Newtonian will score low on the FMCE (below 40%) both preinstruction and postinstruction even if they gain a significant understanding of kinematics. The average normalized gain is 5% on the FMCE and 12% for the FCI for the 409 students in this category. While the majority of these students have low gains on both tests, more than 20% of the students have normalized gains above 30% (going as high as 90%) on the FCI. These students increase their FCI scores by demonstrating their knowledge of kinematics. Low-scoring FCI students can realize strong normalized gains on that exam via two avenues: by learning kinematics and/or Newton's laws. There is no harm in either case, unless one automatically assumes that the FCI gain is only a measure of student's Newtonian understanding.

IV. FINAL THOUGHTS AND CONCLUSIONS

The FCI's long history and wide administration has allowed researchers to compile a vast array of data for the exam. The FMCE is not as widely utilized, so we gathered data from a host of institutes to provide researchers with the opportunity to compare their normalized gains to results from the broader community. In particular we found that normalized gains on the FMCE for traditional instruction are about 15%, while for research-based instructional environments the normalized gains are about 60%.

Scores on the FCI and the FMCE are strongly related—the line of best fit between the two data sets has a correlation coefficient of about 0.78 and a slope of approximately 0.54—for the population we examined, namely, studio physics students at Rensselaer Polytechnic Institute. This relationship exists across a range of instructional approaches used within the studio physics environment and in both preinstruction and postinstruction testings. The students in the introductory physics sequence at Rensselaer already have a relatively strong formal background in math and science, so it is conceivable that some of the relationships we found may not hold for preinstructional testing in populations who have a weaker background in math and science. Further investigation with less prepared preinstructional populations would reveal any differences. In addition, a comparison between the exams at other institutions would help to generalize these results.

Despite this strong relationship between the scores on these two exams, it is clear that they do not sample identical domains. The FMCE is only designed to measure student understanding of one-dimensional forces and motion. The FCI has a broader domain that includes the aforementioned topics, as well as two-dimensional motion and a wider application of forces in more diverse settings. As a result, a typical student's percentage score on the FMCE is almost always lower than on the FCI. Seemingly because of its broader focus, a significant number of students can do well on the FCI, but not score very high on the FMCE.

The data presented in this paper highlight how risky it can be to over rely on single-number scores and normalized gain calculations for any single exam. For example, careful examination of Fig. 4 reveals that the FCI and FMCE would yield different answers to the questions, "did this intervention have a significant impact on my students' learning?" or "which instructional technique results in the largest conceptual learning gains?" (Figure 4 indicates that CGPS resulted in the greatest learning gains as measured by the FCI; but, in the same instructional setting, ILDS resulted in the largest learning gains as measured by the FMCE.)

We believe that both exams are carefully crafted and reasonably robust instruments. We argue that the FMCE provides a more detailed measure of student understanding by virtue of a greater number of items covering a narrower range of topics. On the other hand, the authors of the FCI have argued that a strong performance on the FCI is a good indication of student's ability to solve problems dealing with Newtonian mechanics: a claim we have not tested here. Because of the FMCE's sharper focus, we assert that students who register a large normalized gain on this exam made significant motion toward a Newtonian viewpoint. In contrast, the additional breath of and higher starting scores on the FCI result in smaller normalized gains for these same students. In light of this we argue that the FMCE may be a better exam for instructors and researchers wishing to assess students' understanding of Newton's laws. However, if one is looking at an introductory physics course more generally, then the FCI's wider range of topics may make it the more fitting evaluative instrument.

---

*Corresponding author; jmarx@mcdaniel.edu

[1] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, Phys. Teach. **30**, 141 (1992).

[2] R. R. Hake, Interactive Engagement vs. Traditional Methods: A Six-thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses, Am. J. Phys. **66**, 64 (1998).

[3] R. K. Thornton and D. R. Sokoloff, Assessing Student Learning of Newton's Laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula, Am. J. Phys. **66**, 338 (1998).

[4] M. A. Cooper, An Evaluation of the Implementation of an Integrated Learning System for Introductory College Physics, PhD thesis, Rutgers, The State University of New Jersey, 1993.

[5] J. Wilson, Phys. Teach. **32**, 518 (1994).

[6] K. Cummings, J. Marx, R. K. Thornton, and D. E. Kuhl, Innovations in Studio Physics at Rensselaer, Am. J. Phys. **67**, S38 (1999).

[7] D. R. Sokoloff and R. K. Thornton, Using Interactive Lecture Demonstrations to Create an Active Learning Environment, Phys. Teach. **35**, 340 (1997).

[8] P. Heller, R. Keith, and S. Anderson, Teaching Problem Solving through Cooperative Grouping-Parts 1 and 2, Am. J. Phys. **60**, 627 (1992).

[9] R. K. Thornton, *Conceptual Dynamics: Following Changing Student Views of Force and Motion*, Proceedings of the International Conference on Undergraduate Physics Education, (American Institute of Physics, New York, NY, 1997).

[10] E. Mazur, *Peer Instruction: A User's Manual*, (Prentice Hall, Upper Saddle River, 1997), pp. 45–59.

[11] L. C. McDermott, P. S. Schaffer, and the Physics Education Group, *Tutorials in Introductory Physics* (Prentice Hall, Upper Saddle River, 1998).

[12] D. S. Sokoloff, R. K. Thornton, and P. W. Laws, *RealTime Physics: Active Learning Laboratories* (Wiley, New York, 1998).

[13] P. W. Laws, *Workshop Physics Activity Guide* (Wiley, New York, 1997).

[14] S. Ramlo, Validity and Reliably of the Force and Motion Conceptual Evaluation, Am. J. Phys. **76**, 882 (2008).