

The effect of distracters on student performance on the force concept inventory

N. Sanjay Rebello^{a)} and Dean A. Zollman^{b)}

116 Cardwell Hall, Physics Department, Kansas State University, Manhattan, Kansas 66506-2601

(Received 27 February 2001; accepted 30 September 2003)

We have compared students' responses on four multiple-choice force concept inventory (FCI) questions with similar responses to equivalent open-ended questions. Our results indicate a good agreement between the percentages of correct responses in each of the two formats, indicating that distracters on the FCI do not adversely affect performance as measured by the number of correct answers. However, a significant percentage of the open-ended responses fall into categories that are not included in the FCI multiple choices. When these alternative categories were presented to the students as distracters in a revised multiple-choice format, a significant percentage of the students chose these alternative responses. © 2004 American Association of Physics Teachers.
[DOI: 10.1119/1.1629091]

I. INTRODUCTION

Teachers and researchers have often speculated that the presence of distracters in multiple-choice force concept inventory (FCI)^{1,2} questions could bias students toward the incorrect answer and inaccurately measure students' conceptual understanding. Steinberg and Sabella³ have shown that students performed better on open-ended examination questions than on FCI questions based on the same concept. However, the examination questions were not identical to any of the FCI questions; instead the open-ended examination question evaluated student knowledge on the same concept as the corresponding FCI question. Also, unlike the FCI questions, the open-ended questions were abstract with strong contextual clues to set up the physics associations.

Recently, Schecker and Gerdes⁴ analyzed the FCI as a tool for understanding the model that students applied in dynamics problems. They assumed that students would generally hold one of three models—Aristotelian, Impetus, or Newtonian. To determine the students' model they needed to look beyond the right answers and see which wrong answers the students selected. Then, they needed to determine if the students consistently selected the wrong answer associated with the same model. However, the FCI did not lend itself to such an analysis because all three models were not represented in each of the questions about forces. Thus, it was not possible to use an analysis of wrong answers to determine the students' preferred models.

Schecker and Gerdes also investigated briefly how the context of the question may affect the students' responses.⁴ One of the questions on the FCI asks students to select an answer to describe the forces on a golf ball after it has been hit and is traveling in the air toward a green. They modified the question slightly and asked the students to describe the forces on a soccer ball after it has been kicked and is traveling through the air toward a goal. For the golf ball problem 42 of 87 students included a force in the direction of motion. However, when faced with an identical problem involving a soccer ball, 23 of the 42 students selected either only gravity or gravity plus air resistance. Similar behavior was noted on another question. The authors concluded that the model that students apply to a situation depends on the context.

The lack of consistency was also evident in the models that students applied to problems that involved the same

physics, but were not simple variations of each other. The choice of model depended on the context and the situation presented. This lack of consistency led the authors to conclude that these students were in a mixed state (Mischzustand) when they applied dynamical models. Other research⁵ has shown that naive student beliefs may be too fragmented to characterize any kind of mental model. For instance, DiSessa⁶ prefers to describe student knowledge as a cluster of phenomenological primitives which can be either right or wrong depending on the context in which they are triggered.

The study of the mental models that students apply in various FCI questions is beyond the scope of this study. Rather this study aims to learn more about the effectiveness of the distracters that are currently used on the FCI, and whether alternative distracters would be more effective than the ones currently used. The results of the studies described above indicate that the role of the incorrect answers (distracters) may need further investigation. We are also motivated to look at these distracters in detail for two additional reasons: (1) Ten years have passed since the FCI was constructed. Changes in instructional procedures and student experiences, both in and out of the classroom, may have changed the value of the present distracters. (2) Hestenes² and co-workers designed the FCI from the Mechanics Diagnostic Test⁷ that they had originally developed based on research by others.⁸ The FCI questions were validated through interviews of students over a large range of physics backgrounds from ninth grade to graduate level. The target audience of the FCI may or may not have the same physics background as the population that was interviewed to create the FCI. Thus, it is worthwhile to investigate whether the distracters are effective for students with a particular background.

To investigate these questions, we completed a two-phase investigation. In Phase I we compared student performance on four FCI questions with the same questions that have been rephrased as open-ended questions. Then in Phase II we used the responses to these open-ended questions and created multiple-choice questions with new sets of distracters. Our goal was to determine whether students' scores on the FCI are affected by the multiple-choice format or by the content of the distracters of the questions. Specifically we sought to determine whether:

- (1) students' performances on a multiple-choice question differ significantly from those on an equivalent open-ended question;
- (2) responses to open-ended questions could be categorized into the same choices that are provided on the corresponding multiple-choice question, or whether different categories arise;
- (3) the presence of distracters, as choices for the FCI questions, affects students' selection of incorrect responses;
- (4) the selection of students' responses would change if alternative distracters that arise from our analysis of the open-ended responses are presented instead of or in addition to some of the other FCI distracters affecting student performance.

The FCI is primarily designed to test for a minimal understanding of Newtonian concepts. This goal is accomplished by asking students to select the Newtonian concept over other common alternatives that might be more appealing. In part, the FCI is very successful at meeting this objective because it has a very small percentage of false negatives (selection of a non-Newtonian choice by students who in fact understand Newtonian mechanics) or false positives (selection of a Newtonian choice by students who in fact do not understand Newtonian mechanics).

The FCI also is designed to call student misconceptions to the teacher's attention. The authors of the FCI² have cautioned that the FCI is most prone to misinterpretation in this area, because it is important not to read too much into the responses to a single or even a small subset of the FCI questions. Their data² suggest a threshold of about 60% correct on the entire FCI as a reasonable benchmark for understanding Newtonian concepts. Because our research focuses on an in-depth analysis of distracters on only four FCI questions, our results do not detract from the overall usefulness of the FCI.

We do not investigate how the FCI has met its goals. Instead, we use questions from the FCI to examine a broader issue—how students respond to multiple-choice and open-ended questions on the same topic and what we can learn from the differences in these responses.

II. PHASE I

We developed a set of instruments based on four questions from the most recent version of the FCI. We chose questions that, based on published data,² address the largest number of misconceptions. For each of these questions we created an equivalent open-ended question. With one exception, the open-ended question required only trivial changes and removal of the five choices. FCI Question #15 has multiple choices that needed more extensive rewriting as an open-ended question. With these eight questions—four multiple-choice and four equivalent open-ended—we created two questionnaires, each containing two questions of each type. Table I shows the contents of each questionnaire.

Each student received a questionnaire with two multiple-choice and two open-ended questions. Half of the students in each class answered the first version while the remaining answered the second version. The students were randomly selected for each questionnaire. In effect, students answering one questionnaire were the control group for those answering the other and vice versa.

We performed a pilot test of the questionnaires on the first day of class with 25 students in a second-semester algebra-

based introductory physics course. The questionnaire was presented as a diagnostic and students were told that it would not affect their grades. Students were given up to a maximum of 15 minutes to answer all four questions on the test. No special incentives (for example, extra credit) were used to induce the students to take the test. Based on the responses, we were able to keep the design unchanged.

Next, we administered the questionnaires to 238 students in an algebra-based introductory physics course. Again, the questionnaire was presented on the first day of class as a diagnostic with no implications for grades and no incentives were provided. For the multiple-choice questions we recorded the number of students who gave each choice as their answer. Using phenomenographical methods^{9,10} we categorized the open-ended responses. In this approach the categories are selected from those that naturally occur in the students' responses. We did not establish categories in advance of reading the responses. The categories were established, modified, and agreed upon by multiple readers. Then, each reader independently placed all responses in one or more of the agreed upon categories. Using this procedure three researchers placed each response in a category. The reliability of the three researchers for this method of categorizing the responses was more than 90%.

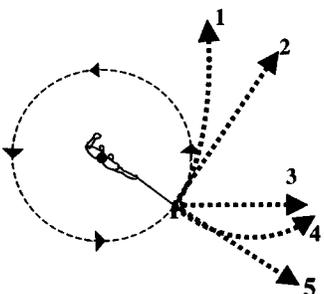
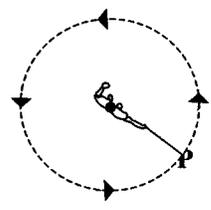
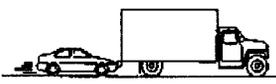
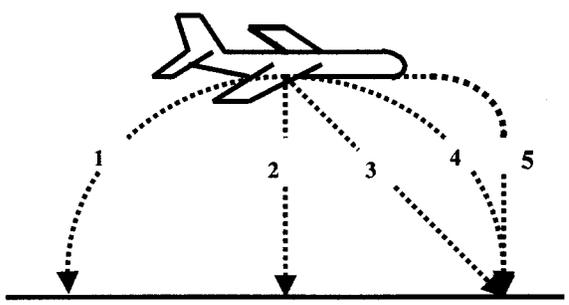
The students taking the algebra-based course are primarily non-physics science majors. Most of these students are pre-medical or pre-veterinary students. Typically, students who take this course have completed a year of high school physics before entering Kansas State University. The gender ratio is typically one. The university is located in rural northeastern Kansas and the level of racial and ethnic diversity in the student body is typically less than the national average. The vast majority of the students are traditional students who have entered the university directly after completing high school.

During the first phase we were primarily interested in how the open-ended responses compared to the concepts represented by the multiple-choice responses. In the following we will consider each question and then draw some general conclusions. We will discuss Question II at the end, because it was more significantly altered than the other questions when converted to the open-ended format. In all of the discussion that follows, the term "category" of responses refers to the categories that arose from the phenomenographical analysis of the open-ended responses. The term "choice" refers to the alternative that was selected by the students in the multiple-choice format.

Question 1. Responses to the multiple-choice and open-ended formats are shown in Figs. 1(a) and 1(b), respectively. Categories 1, 2, and 3 of the open-ended responses all appear to be tangential to the circle and have been combined. Categories 5 and 6 do not have equivalent multiple-choice responses. None of the categories for the open-ended responses are equivalent to choices 1 or 3.

The percentages of correct responses in the open-ended and multiple-choice formats agree within 5%. However, the most frequent incorrect open-ended response is Category 4 (22%), which differs from the most frequent incorrect multiple-choice response (choice 5, 11%). Also, about 9% (Categories 5 and 6) of the responses in the open-ended format do not correspond to any multiple choice responses and 22% (choices 1 and 3) of the multiple-choice responses do not correspond to any of the categories in the open-ended questions. These results indicate that although the percentage

Table I. The multiple-choice (FCI) and equivalent open-ended questions in each questionnaire. The number in parentheses in the left-hand column is the question number on the latest version of the FCI.

Q#	Questionnaire A	Questionnaire B
I (7)	<p>A steel ball is attached to a string and swung in a circular path in a horizontal plane as illustrated in the figure below. At point P the string suddenly breaks near the ball. If these events are observed from directly above... which of the paths 1–5 below would the ball most closely follow after the string breaks.</p> 	<p>indicate on the diagram below the path which the ball would most closely follow if the string breaks.</p> 
II (15)	<p>A large truck breaks down out on the road and receives a push back into town by a small compact car as shown in the figure below.</p>  <p>While the car still pushing the truck is speeding up to get up to cruising speed...</p> <ul style="list-style-type: none"> • Does the car exert a force on the truck? • Does the truck exert a force on the car? If so, how does it compare with the force exerted by the car on the truck? • Will your answers to the above question change if the engine of the truck were running? 	<ol style="list-style-type: none"> 1. The amount of force with which the car pushes on the truck is equal to that which the truck pushes back on the car. 2. The amount of force which the car pushes on the truck is smaller than that with which the truck pushes back on the car. 3. The amount of force which the car pushes on the truck is smaller than that with which the truck pushes back on the car. 4. The car's engine is running so the car pushes against the truck, but the truck's engine is not running so the truck cannot push back against the car. The truck is pushed back simply because it is in the way of the car. 5. Neither the car nor the truck exerts any force on the other. The truck is pushed forward simply because it is in the way of the car.
III (14)	<p>A bowling ball accidentally falls out of the cargo bay of an airliner as it flies along in a horizontal direction. As observed by a person standing on the ground and viewing the plane as in the figure below... which of the paths 1–5 would the bowling ball most closely follow after leaving the plane?</p> 	<p>Indicate on the diagram the path that the ball would most closely follow after leaving the plane?</p> 
IV (27)	<p>A woman exerts a constant horizontal force on a large box. As a result the box moves across the floor at a constant speed v_0. If the woman suddenly stops applying the horizontal force to the block... Describe the motion of the block.</p>	<p>Circle the correct statement.</p> <ol style="list-style-type: none"> 1. Immediately comes to a stop. 2. Continues moving at a constant speed for a while and then slows to a stop. 3. Immediately starts slowing to a stop. 4. Continues at a constant speed. 5. Increases its speed for a while and then starts slowing to a stop.

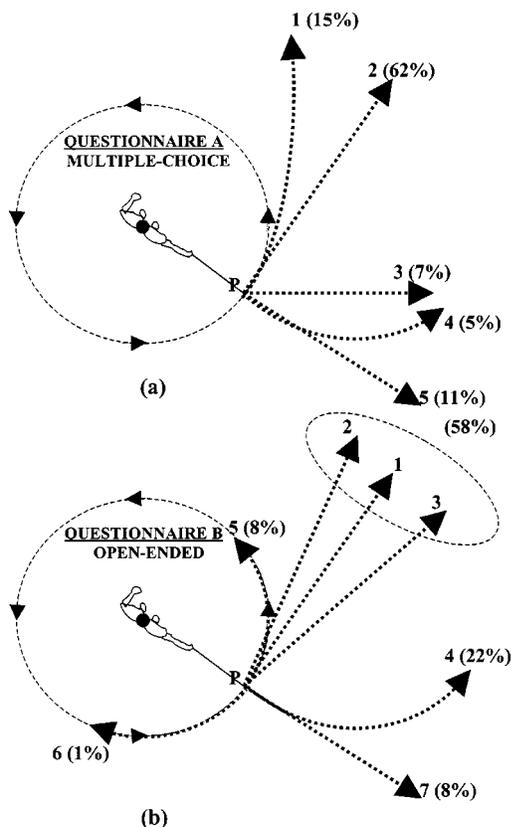


Fig. 1. Responses to multiple-choice and open-ended versions of Question I in Phase I. The open-ended responses were categorized. The percentages of each response are shown.

of correct responses may not be affected by the format, some of the incorrect responses that students give will change with the format.

Question III. Responses to the multiple-choice and open-ended formats are shown in Figs. 2(a) and 2(b), respectively. There are two significant differences between the two formats. First, Category 3 (28%) in the open-ended responses is not one of the available multiple choices. Second, none of the students selected choice 5 in the multiple-choice format. Similar to Question I, the percentages of correct responses in the open-ended and multiple-choice formats agree within 7%. The most frequent incorrect response was choice 2 (37%) in the multiple-choice format and Category 5 (28%) in the open-ended format, which had no equivalent multiple-choice response. Also similar to Question I, these results indicate that although the percentage of correct responses may not be affected by the format, some of the incorrect responses that students give depend on the format.

Question IV. Responses to the multiple-choice and open-ended formats are shown in Table II. Except for two of the categories, the rest were significantly different from the FCI choices. We categorized responses that said the box would “stop” (Category 4), separately from those that said it would “stop suddenly/immediately” (Category 5), because in the latter case we are more certain of the nature of the student misconceptions than in the former. Category 2 was created for responses that the box would “stop if the floor was frictional and continue if it was frictionless.” These students were unable to identify the frictional interaction between the floor and box from the information in the problem.

Similar to Questions I and III the percentages of correct

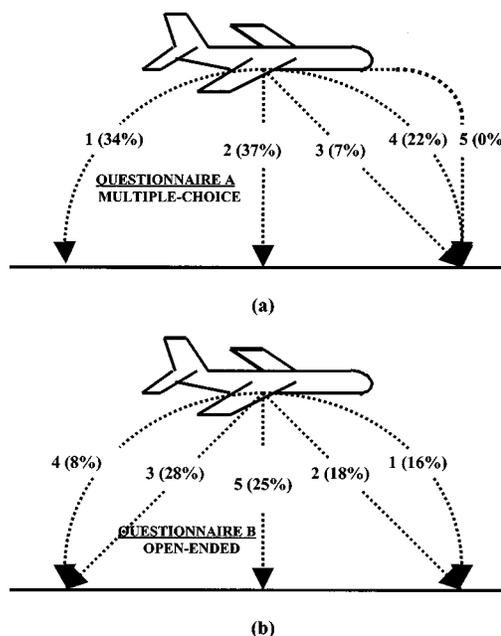


Fig. 2. Responses to multiple-choice and open-ended versions of Question III in Phase I. The open-ended responses were categorized. The percentages of each response are shown.

responses in the open-ended and multiple-choice formats agree within 7%. The most frequent incorrect response was choice 1 (“stops immediately” 51%) in the multiple-choice format and Category 4 (“stops” 43%) in the open-ended format. Only 5% of the open-ended responses mentioned that the motion of the box would depend on friction (Category 2).

Question II. This question was rewritten in the open-ended format with significant changes compared to the other questions and hence the data had to be analyzed differently. We divided the question into three subquestions each of which was categorized separately. Responses to the multiple-choice and open-ended formats are shown in Table III.

Subquestion: “Does the car exert a force on the truck?” Almost all (98%) of the students answered yes to this question. Hence, it appears that this question had an obvious answer and need not have been asked.

Subquestion: “Does the truck exert a force on the car?” Again, almost all (98%) of the students answered yes to this question. A second part of this subquestion asked the students to compare the forces of the car and the truck. This key subquestion addressed the primary misconception of the original FCI question. Forty-two percent of the open-ended responses and 22% of the multiple-choice responses were correct. Forty-nine percent of the open-ended responses indicated that “the truck would exert more force than the car,” while 60% of the students selected the corresponding choice 3 in the multiple-choice format. Thus, the distracter (choice 3) in the multiple-choice format did have a significant impact on student performance.

Subquestion: “Will your answers to the above questions change if the engine of the truck were running?” This subquestion was included to account for choice 4 on the original FCI question. Sixty-one percent of the students responded no to this question and the remaining (39%) responded yes. We then proceeded to categorize the reasons that students gave for their responses. The most common reason given by those who responded yes was that “the truck was moving under its

Table II. Responses to multiple-choice and open-ended versions of Question IV in Phase I. The open-ended responses were categorized. The percentages of each response are shown.

Multiple-choice	Open-ended
1 (51%): immediately comes to a stop.	1 (3%): continues moving at a constant speed.
2 (3%): continues moving at a constant speed for a while and then comes to a stop.	2 (9%): if ground is frictional it slowly stops, if not frictional it continues at same speed.
3 (39%): immediately starts slowing to a stop.	3 (32%): slows to a stop.
4 (2%): continues at a constant speed.	4 (43%): stops.
5 (4%): increases its speed for a while and then starts slowing to a stop.	5 (10%): stops suddenly.

own power” or “the truck would exert less force.” About 13% of the students stated that their answer would depend upon “the gear of the truck/car.” Among the students that responded no to the above question, about a third of the students mentioned Newton’s third law or related reasons. Sixteen percent of the non-calculus-based students said that there would be no difference as long as the truck/car were not accelerating.

In general, a significant number of students (over 35%) who had correctly answered the first two subquestions failed to answer the third subquestion correctly. When we compare these results with the multiple-choice format, we find that only 9% of the students selected choice 4, which is the only choice that mentions the running engine. Thus, in this question, the FCI distracter (choice 4) was not effective in misleading the students when they were asked to select from the five available FCI choices. However, when students were

explicitly asked whether the running engine of the truck would make a difference to their answer, they responded yes. Hence, we conclude that subquestion 3 of the open-ended format was effective in uncovering a conceptual difficulty that does not arise when students see the same idea expressed in only one of five choices.

Based on these results we can draw some general conclusions. Overall, we notice that there is no notable difference between student performances in terms of the percentage of correct responses on the two formats. If the FCI is used for determining how many students can answer the FCI questions correctly, the multiple-choice and open-ended formats give equivalent results. The most frequent incorrect responses for each question varied significantly between the open-ended and multiple-choice. For Question I and Question III, the category of the most frequent response had no equivalent choice on the multiple-choice format. Conversely,

Table III. Responses to multiple-choice and open-ended versions of Question II in Phase I. The open-ended responses were categorized. The percentages of each response are shown.

Multiple-choice	Open-ended
1 (22%): The amount of force with which the car pushes on the truck is equal to the force with which the truck pushes back on the car.	Does the Car exert a force on the Truck? 98% Yes 2% No
2(9%): The amount of force with which the car pushes on the truck is smaller than the force with which the truck pushes back on the car.	Does the Truck exert a force on the Car? 98% Yes 2% No
3(60%): The amount of force with which the car pushes on the truck is greater than the force with which the truck pushes back on the car.	If so, how does it compare with the force exerted by the car on the truck? 42% Equal forces. 49% Truck exerts less force than Car. 9% Truck exerts more force than Car.
4(9%): The car’s engine is running so the car pushes against the truck, but the truck’s engine is not running, so the truck cannot push back on the car. The truck is pushed forward simply because it is in the way of the car.	Will your answers to the above question change if the engine of the truck were running? 61% No Reasons: 50% Truck exerts more force 50% Truck under own power. 39% Yes Reasons: 37% Truck under own power. 37% Truck exerts less force. 14% Depends upon gear of car.
5(0%): Neither the car nor the truck exerts any force on the other. The truck is pushed forward simply because it is in the way of the car.	5% Friction against car is less. 2% Truck is accelerating.

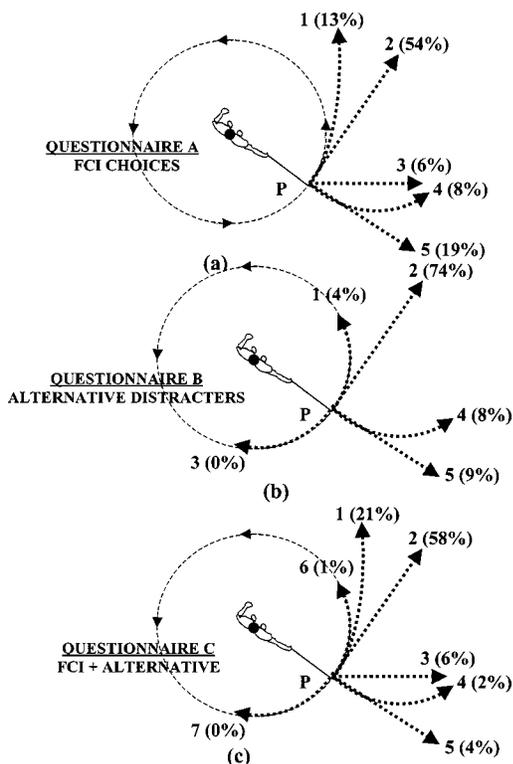


Fig. 3. Responses to the three multiple-choice versions of Question I in Phase II. The percentages of each response are shown.

at least one choice on the multiple-choice format for Questions I and III did not have any corresponding open-ended category, and was selected by only a few ($<15\%$) on the multiple-choice formats. For our students, the more effective distracters derived from the category of the most frequently incorrect open-ended response could replace these choices. Hence, if the FCI is being used to determine the students' misconceptions, it is less effective than the equivalent open-ended questions. For the questions that we used, at least one of the distracters does not yield any "hits" and one notable category does not correspond to any multiple choices. Thus, teachers trying to determine students' misconceptions will lose information by using the multiple-choice format.

From our results for Question II, which was significantly modified in the open-ended format, we found that students who gave the correct response on the first two subquestions were misled by the third subquestion. This subquestion was introduced to reflect choice 4 on the multiple-choice format. Although almost no students selected choice 4 on the multiple-choice format, they responded incorrectly to this subquestion. Thus, a misconception stated in one of the multiple choices is not selected by any of the students, but it does appear when students are asked about it specifically.

In general, the multiple-choice format of the FCI seems to be useful in determining which students choose the right answer, but is of limited value in determining the alternative conceptions for students who do not respond correctly.

III. PHASE II

Based on the categorization of the open-ended responses to the questions asked in Phase I, we observe that notable categories do not have equivalents in the present FCI choices. To determine whether these categories could be ef-

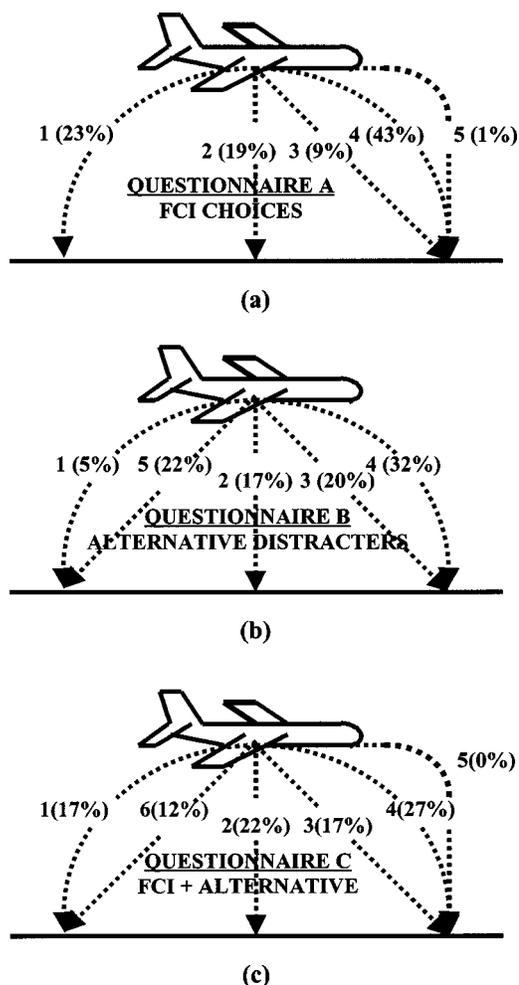


Fig. 4. Responses to the three multiple-choice versions of Question III in Phase II. The percentages of each response are shown.

fective distracters, we constructed three questionnaires. All of the questionnaires used the original FCI questions and had multiple-choice answers. They differed in the content of the distracters. Questionnaire A contained the original FCI distracters. In questionnaire B we removed those original distracters that were chosen by very few students and replaced them with distracters constructed from categories mentioned frequently in open-ended responses from Phase I. Questionnaire C contains all of the distracters from questionnaires A and B.

We administered the questionnaires to 234 students in an algebra-based introductory physics course. Each student completed one randomly chosen version of the questionnaire. Again, the questionnaire was presented as a diagnostic on the first day of class, with no implications on student grades.

Question I. Responses to Question I are shown in Fig. 3. Choices 1 and 3 in FCI questionnaire A were replaced with other alternatives in questionnaire B. This change caused the percentage of correct responses to increase by about 20%, which is approximately the percentage of students that were distracted toward choices 1 and 3 in the FCI questionnaire. Choices 1 and 3 in questionnaire B (which are choices 6 and 7 in questionnaire C) were extracted from the categories of the open-ended responses in Phase I where together they were about 10% of the responses. When presented as alter-

Table IV. Responses to the three multiple-choice versions of Question IV in Phase II. The percentages of each response are shown.

FCI choices (questionnaire A)	Alternative distracters (questionnaire B)	FCI+ alternative (questionnaire C)
1 (25%): immediately comes to a stop.	1 (21%): immediately comes to a stop.	1 (16%): immediately comes to a stop.
2 (5%): continues moving at a constant speed for a while and then comes to a stop.	2 (13%): immediately starts slowing to a stop.	2 (9%): continues moving at a constant speed for a while and the comes to a stop.
3 (64%): immediately starts slowing to a stop.	3 (1%): continues at a constant speed.	3 (21%): immediately starts slowing to a stop.
4 (1%): continues at a constant speed.	4 (60%): continues at the same speed if the ground is non-frictional. If the ground is frictional it slows to a stop.	4 (0%): continues at a constant speed.
5 (0%): increases its speed for a while and then starts slowing to a stop.		5 (0%): increases its speed for a while and then starts slowing to a stop.
		6 (49%): continues at the same speed if the ground is non-frictional. If the ground is frictional it slows to a stop.

natives on a multiple-choice instrument (questionnaires B and C) however, they were less than 5% of the overall response.

These data indicate that choices 1 and 3 on FCI questionnaire A serve as effective distracters and will significantly alter the percentage of correct responses if they are omitted as in questionnaire B. On the open-ended response no students drew the curved path represented by choice 1 on questionnaires A and C. Those students who drew paths in the general direction of somewhere between a tangent to the circle and the circle itself, always drew straight lines. However, when presented with this alternative a rather sizable fraction of the students chose it. These results indicate that the percentage of correct responses depends on the distracters used in a multiple-choice format, although some of these distracters may not correspond to responses to an open-ended version of the same question.

Question III. Responses to Question III are shown in Fig. 4. The percentage of correct responses (choice 4) decreases by at least 10% when choice 5 on the original FCI questionnaire A is replaced by a new choice, a backward diagonal path (choice 5 in B, choice 6 in C). Over a fifth of the respondents selected the backward diagonal path when it was presented as a distracter in questionnaire B. Conversely, over a fifth of the respondents selected the backward parabolic path (choice 1) in FCI questionnaire A, while only 5% selected this choice when the backward diagonal path was also provided as a choice. Almost no respondents selected choice 5 in questionnaires A and C.

These data indicate that choice 5 of FCI questionnaire A is not as effective a distracter as the backward diagonal path (choice 5 in B, choice 6 in C). It also appears that students, who may have selected the backward diagonal path, instead selected the backward parabolic path (choice 1) in the original FCI, where the backward diagonal path was not provided. These results indicate that the backward diagonal path serves as an effective distracter and should be introduced as a possible choice on the FCI. Alternatively, choice 5 on the FCI could be removed because almost nobody selected it in

any of the questionnaires. The present choices on the FCI seem to be steering students toward a correct response even though they may prefer an alternative.

Question IV. Responses to Question IV are shown in Table IV. Over 60% of the respondents on FCI questionnaire A selected the correct answer (choice 3). When the distracter mentioning friction (choice 4 in A, choice 6 in C) is introduced, however, the results change dramatically. Over 60% of the respondents selected this distracter in questionnaire A and nearly half in questionnaire C, where the original FCI distracters are also present. About 13% of the respondents of questionnaire B and about 21% of the respondents in questionnaire C selected the correct answer (“immediately starts slowing to a stop”). The FCI distracter (choice 2 in A and C) “continues moving at a constant speed for a while and then slows to a stop” was chosen by fewer than 10% of the students in either of the questionnaires. Similarly, hardly any respondents selected the FCI distracter “increases its speed for a while and then starts slowing to a stop” (choice 5 in B and C) or FCI distracter “continues at a constant speed” (choice 4 in B and C).

These data indicate that choices 2, 4, and 5 on the original FCI question (questionnaire B) are selected by virtually no students. Conversely the distracter that points students toward friction appears to be extremely effective in that it changes the percentage of correct responses from 60% to less than 25% when it is introduced. This distracter was also selected by 60% of the respondents. These results indicate that the choice about friction serves as an effective distracter and should be introduced as a possible choice on the FCI. Alternatively, choices 4 and 5 on the FCI could be removed because almost nobody selected it in any of the questionnaires. Again, the presence of a new distracter can significantly alter the percentage of correct responses. This new distracter concerning friction uncovers a previously hidden possible student misconception about friction.

Given an answer that includes a lack of friction, students may choose it to be safe. They may have become accustomed to textbook situations in which frictionless surfaces

Table V. Responses to the three multiple-choice versions of Question II in Phase II. This question was subdivided into two subquestions based on the open-ended categories in Phase I. The percentages of each response are shown.

FCI Choices (questionnaire A)	Alternative distracters (questionnaire B)	FCI+ alternative (questionnaire C)
1(22%): The amount of force with which the car pushes on the truck is equal to the force with which the truck pushes back on the car.	Subquestion 1: How does the force exerted on the truck compare with the force exerted on the car? 1(23%): Force with which the car pushes on the truck is equal to that which the truck pushes back on the car.	Subquestion 1: How does the force exerted on the truck compare with the force exerted on the car? 1(21%): Force with which the car pushes on the truck is equal to that which the truck pushes back on the car.
2(9%): The amount of force with which the car pushes on the truck is smaller than the force with which the truck pushes back on the car.	2(14%): Force with which the car pushes on the truck is smaller than that which the truck pushes back on the car.	2(12%): Force with which the car pushes on the truck is smaller than that which the truck pushes back on the car.
3(60%): The amount of force with which the car pushes on the truck is greater than the force with which the truck pushes back on the car.	3(63%): Force with which the car pushes on the truck is greater than that which the truck pushes back on the car.	3(60%): Force with which the car pushes on the truck is greater than which the truck pushes back on the car.
4(9%): The car's engine is running so the car pushes against the truck, but the truck's engine is not running, so the truck cannot push back on the car. The truck is pushed forward simply because it is in the way of the car.	Subquestion 2: If the engine of the truck were running, the answer to the above question... (circle the correct statement) 1(58%): would not change.	4(2%): The car's engine is running so the car pushes against the truck, but the truck's engine is not running, so the truck does not push against the car
5(0%): Neither the car nor the truck exerts any force on the other. The truck is pushed forward simply because it is in the way of the car.	2(30%): would change depending upon the gear in which the truck's engine is running.	5(0%): Neither the car, nor the truck exert any force on each other.
	3(8%): would change, and the force exerted by the truck would be greater than that of the car.	Subquestion 2: If the engine of the truck were running, the answer to the above question... (circle the correct statement) 1(46%): would not change.
	4(5%): would change, and the force exerted by the car would be greater than that of the truck.	2(35%): would change depending upon the gear in which the truck's engine is running.
		3(7%): would change, and the force exerted by the truck would be greater than that of the car.
		4(5%): would change, and the force exerted by the car would be greater than that of the truck.

are present and thus choose an answer that covers both friction and non-friction. If we allow both the answer “Immediately starts slowing” and the one that explicitly mentions friction as correct, the number of correct responses for this question increases by 9% for version B and by 6% for version C. These correct answers are consistent with answers from students who would choose “immediately comes to a stop” or “continues to move at a constant speed then comes to a stop.” The latter of these answers did not appear in the open-ended responses. Thus, in this case we seem to be seeing a complex interaction in which the students’ selections of answers depend not only on the answer they choose but on the others that they have read.

The authors of the original FCI avoided the use of the word “friction” in the choices so that students would not be deliberately confused with unfamiliar scientific terminology. Although this reason may be appropriate for excluding a distracter for students who have not had prior exposure to physics, we believe that it is particularly important to include it for students who may have learned about friction. The distracter tests whether or not these students have understood how to apply the concept of friction in this problem.

Question II. Responses to Question II are shown in Table V. In each of the three questionnaires, about 60% of the respondents stated that the force of the car is greater than that of the truck and about 15% stated that the force of the truck is greater than that of the car. In each of the three questionnaires about 20% of the respondents selected the correct response (equal forces). Very few (<10%) of the students selected the other FCI distracters (choices 4 and 5 in questionnaire A).

The revised format consisted of two subquestions to ac-

commodate the categories of open-ended responses from Phase I. In subquestion 2 over half of the respondents in questionnaires B and C indicated that their response would not change if the engine of the truck were running. About a third of the respondents indicated that their response would change depending upon the gear in which the truck is operating.

These data indicate that choices 4 and 5 on the original FCI (questionnaire A) are not effective distracters because they are selected by less than 10% of the respondents. There is good agreement (within 10%) between the responses that compare the forces of the truck and the car, with most of the students incorrectly stating that the force of the car is greater than that of the truck. However, nearly one-third of the students incorrectly indicated that their response would change depending upon the gear of the truck.

These results indicate that the choice specifically asking them whether their response would change depending upon the gear of the truck serves as an effective analysis of their understanding. Choices 4 and 5 on the FCI could be removed because fewer than 10% of the respondents selected them in any of the questionnaires. Here, the presence of a new distracter (“answer depends upon gear of truck”) when asked as a specific question evoked incorrect responses and may possibly uncover a previously hidden misconception regarding Newton’s third law.

It should also be pointed out that in the FCI, this question is followed by a companion question (FCI Question #16). Students are presented with identical choices in which the car has reached a “constant cruising speed” as it pushes the truck. It is likely that when students encounter this question in the original FCI, they begin to reflect on their choice to the

previous question (our Question II) in which the car is “speeding up to a cruising speed.” Indeed Bao and co-workers¹¹ have found that the acceleration is a relevant physical feature in determining the mental model that students apply in certain contexts of Newton’s third law problems. It is possible that student responses to this question in the original FCI were affected by the question that followed it. The dependence of student responses on the context of the questions asked before and after it merits further study, but is beyond the scope of this paper.

Based on the results of these four questions we note that in most cases the incorrect responses to the open-ended questions in Phase I can serve as effective distracters when introduced as choices in the multiple-choice format. Some of these distracters (Questions II and IV) may uncover misconceptions that may not have been addressed in the existing FCI choices. These revised distracters could possibly replace some of the existing FCI distracters. In versions where both the FCI distracters as well as the revised distracters were presented, the latter tended to dominate.

IV. SUMMARY AND CONCLUSIONS

We selected four FCI questions that addressed the greatest number of misconceptions. In Phase I we presented these questions in two questionnaires, each containing two open-ended and two multiple-choice questions. The open-ended and multiple-choice responses to each question were compared. The open-ended responses were categorized and compared with the multiple-choice responses.

In Phase II we created revised multiple-choice distracters based on the categories of the open-ended responses in Phase I. We compared the student performance on three versions of each question: the original FCI with the revised distracters, with a combination of the revised distracters, and the original FCI choices.

Based on our results for these four questions we conclude the following.

(1) The percentage of correct responses to an open-ended version of the FCI questions does not differ significantly with the percentage of correct responses to the multiple-choice (original) FCI question. In fact, the percentage of correct responses in both formats is quite high and may be because most of these students have taken physics in high school.

(2) The categories of the open-ended responses do not exactly match the choices provided on the original FCI question. Often a significant percentage of incorrect open-ended responses will not have equivalent multiple-choice distracters.

(3) The distracters on the original FCI question alter the distribution of the incorrect responses, although they may not significantly affect the percentage of correct responses.

(4) When the categories of the open-ended responses are presented as alternative distracters in a multiple-choice format, they may significantly alter the percentage of correct responses. Often the categories that were taken from the incorrect open-ended responses serve as more effective distracters than the original FCI distracters.

Based on these conclusions we believe that the FCI in its present form is as effective for determining the percentage of students who can provide the correct answers as the open-ended questions. However, a significant percentage of open-ended responses do not correspond to any of the distracters

on the present FCI questions. Thus, an analysis of the incorrect responses to FCI questions may not be an effective way to determine which parts of the students’ conceptual understanding are deficient. This conclusion is similar to that discussed in Ref. 4 where the FCI was considered as a possible way to determine the students’ underlying model for describing motion.

It may be possible to create a revised version of the FCI questions with revised distracters extracted from open-ended responses such as the ones that our students gave. Then, the percentage of correct responses on this revised FCI could be quite different from the original FCI. These revised FCI distracters would be more closely linked with some of the student misconceptions than the original FCI distracters and could serve as a better tool for determining students’ alternative conceptions.

The FCI was originally created using responses supplied by students to open-ended questions. Why then do we find that several of the open-ended responses do not correspond to any of the FCI choices? Also, why do we find that when these open-ended responses are presented as alternative distracters, they can significantly affect the percentage of correct responses? Although the original FCI design was validated by interviews with students ranging from ninth graders to graduate students, the participants in our study were just beginning their introductory undergraduate course and could have been exposed to physics at a level different from the pool of students that were used to validate the FCI responses. Further, the focus on change in physics instruction brought about in part by results¹² on the FCI over the past 10 years could have influenced what the students have learned and thus their understanding of the laws of motion.

A broader impact of the study is the implication for all multiple-choice instruments. Many such instruments are used in pre/post-instruction analysis. The effect of distracters could change during the course of instruction. The distracters that are effective before students have completed instruction may be ineffective or more effective after instruction. Further, students may develop a new set of alternative conceptions that are not addressed in the instrument or language¹³ used in the questionnaire, which could lead to student responses that do not accurately reflect the nature of the students’ conceptual understanding. This phenomenon could possibly lead to pre/post-comparisons that do not accurately reflect the level of student understanding that they have acquired.

ACKNOWLEDGMENTS

The authors would like to thank David Hestenes and an anonymous referee for their detailed comments on a draft of this paper.

^{a)}Electronic mail: srebello@phys.ksu.edu

^{b)}Electronic mail: dzollman@phys.ksu.edu

¹E. Mazur, *Peer Instruction* (Prentice-Hall, Englewood Cliffs, NJ, 1997).

²D. Hestenes, M. Wells, and G. Swackhamer, “Force concept inventory,” *Philos. Trans. R. Soc. London* **30**, 141–157 (1992).

³R. N. Steinberg and M. S. Sabella, “Performance on multiple-choice diagnostics and complementary exam problems,” *Phys. Teach.* **35**, 150–155 (1997).

⁴H. Schecker and J. Gerdes, “Messung von Konzeptualisierungsfähigkeit in der Mechanik-Zur Aussagekraft des Force Concept Inventory,” *Zeitschrift für Didaktik der Naturwissenschaften* **5** (1), 75–89 (1999).

⁵T. A. Halloun and D. Hestenes, “Common sense concepts about motion,” *Am. J. Phys.* **53** (11), 1056–1065 (1985).

- ⁶A. A. DiSessa, "Towards an epistemology of physics," *Cogn. Instruction* **10** (2-3), 105–225 (1993).
- ⁷I. A. Halloun and D. Hestenes, "The initial knowledge state of college physics students," *Am. J. Phys.* **53** (11), 1043–1055 (1985).
- ⁸See Refs. 1–5 cited in Ref. 7 above.
- ⁹F. Marton, "Phenomenography—describing conceptions of the world around us," *Instructional Sci.* **10**, 177–200 (1981).
- ¹⁰F. Marton, "Phenomenography—a research approach to investigating different understanding of reality," *J. Thought* **21**, 29–39 (1986).
- ¹¹L. Bao, D. Zollman, and K. Hogg, "Model analysis of fine structures of student models: An example with Newton's third law," *Physics Education Research: A Supplement to the American Journal of Physics* (to be published).
- ¹²R. P. Hake, "Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," *Am. J. Phys.* **66**, 64–74 (1998).
- ¹³D. Clark and M. Rutherford, "Language as a confounding variable in the diagnosis of misconceptions," *Int. J. Sci. Educ.* **22**, 703–717 (2000).
-



Center-of-Mass Demonstration. The otherwise-uniform wooden wheel is loaded toward the edge with a circular slug of lead. Two holes are provided in the wheel for the pivot pin: one at the geometrical center and one at the center of mass. This apparatus was purchased by the Denison University physics department in 1905 from the Central Scientific Company of Chicago. (Photograph and notes by Thomas B. Greenslade, Jr., Kenyon College)