Relation between students' problem-solving performance and representational format

David E. Meltzer[a)]
Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011

(Received 15 August 2003; accepted 7 January 2005)

An analysis is presented of data on students' problem-solving performance on similar problems posed in diverse representations. Five years of classroom data on 400 students collected in a second-semester algebra-based general physics course are presented. Two very similar Newton's third-law questions, one posed in a verbal representation and one in a diagrammatic representation using vector diagrams, were given to students at the beginning of the course. The proportion of correct responses on the verbal question was consistently higher than on the diagrammatic question, and the pattern of incorrect responses on the two questions also differed consistently. Two additional four-question quizzes were given to students during the semester; each quiz had four very similar questions posed in the four representations: verbal, diagrammatic, mathematical/symbolic, and graphical. In general, the error rates for the four representations were very similar, but there was substantial evidence that females had a slightly higher error rate on the graphical questions relative to the other representations, whereas the evidence for male students was more ambiguous. There also was evidence that females had higher error rates on circuit-diagram problems in comparison with males, although both males and females had received identical instruction.
[DOI: 10.1119/1.1862636]

I. INTRODUCTION

This paper reports on the initial phase of an investigation into the role of diverse representations in the learning of physics concepts. The goal is to explore the relation between the form of representation of complex concepts, and students' ability to learn these concepts. Much previous research has shown that the use of multiple forms of representation in teaching concepts in physics has great potential benefit, and yet poses significant challenges to students and instructors.[1,2] Facility in the use of more than one representation deepens a student's understanding, but specific learning difficulties arise in the use of diverse representations.[3]

By representation I mean any of the widely diverse forms in which physical concepts may be understood and communicated. In Appendix A I show an example of the use of four representations for what is essentially the same problem. The representations are referred to here as verbal (V), diagrammatic (D), mathematical/symbolic (M), and graphical (G), corresponding to questions 1–4, respectively.[4] Although these questions are nearly identical and illustrate four different ways of representing the same concept, to an introductory student they might appear very different. It often is assumed by instructors that a representation which they find especially clear and comprehensible (for example, a graph) also will be especially clear for the average student. Research and experience shows that this assumption often is not correct,[3] but relatively little work has been devoted to testing it systematically. In this paper I will discuss a variety of methods of investigating how specific representations may be related to student thinking, and I will analyze classroom data to generate some preliminary hypotheses regarding this relation.

II. THE ROLE OF MULTIPLE REPRESENTATIONS IN STUDENT LEARNING OF PHYSICS

A. Outline of previous research

There is no purely abstract understanding of a physical concept—it is always expressed in some form of representation. Physical scientists employ a variety of representations as a means for understanding and working with physical systems and processes.[5–9] In many recently developed curricular materials in physics[1,2,10–16] and chemistry,[17] there has been much attention to presenting concepts with a diversity of representations. Van Heuvelen was one of the earliest to emphasize the potential benefits of this instructional strategy in physics.[1] Numerous physics educators have stressed the importance of students developing an ability to translate among different forms of representation of concepts,[1,3,18–22] and researchers in other fields have stressed similar themes.[23–27] Moreover, it has been pointed out that thorough understanding of a particular concept may *require* an ability to recognize and manipulate that concept in a variety of representations.[2,3]

It is well established that specific learning difficulties may arise with instructional use of diverse representations.[3] Student difficulties in mastering physics concepts using graphical representations have been studied in considerable detail and specificity for topics in kinematics.[18,28–30] These studies and other related work in mathematics education[31] have delineated several broad categories of conceptual difficulties with graphs. Conceptual difficulties related to diagrammatic representations of electric circuits and fields have been addressed,[32] as have those in optics.[33] Difficulties arising from linguistic ambiguities (verbal representation) also have been explored.[34] Specific representational difficulties in chemistry education, largely parallel to similar issues in physics education, also have been investigated.[35,36]

B. Research issues related to multiple representations

Beyond the investigations in the literature cited, there are few available research results that focus on problems that arise in the learning of physics concepts with multiple forms of representation. As McDermott has emphasized, there is a need to identify the specific difficulties students have with various representations.[3] I suggest that additional insight might result from investigations that explicitly compare learning in more than one form of representation. Although a number of recent investigations in science education and other fields have focused on broader issues involved in student learning with diverse representations,[37,38] there seems to have been relatively little effort to compare representations in terms of their pedagogical effectiveness in particular contexts.[39]

A closely related issue is that of students' relative performance on similar problems that make use of different representational forms.[21,26,29,40,41] In this regard, Kozma[42] and Kozma and Russell[26] have reported on the relative degree of difficulty encountered by novice students presented with a chemistry problem posed in various representations. Among physics and chemistry educators, there has been speculation regarding the role that students' individual learning styles might play,[43] and the possible relevance of gender differences[35,40] and spatial ability.[44]

The present investigation focuses on specific issues arising when multiple representations are utilized in undergraduate physics instruction. Ultimately, the issues we plan to investigate include the following:

(1) What subject-specific learning difficulties can be identified with various forms of representation of particular concepts in the introductory physics curriculum?
(2) What generalizations might be possible regarding the relative degree of difficulty of various representations in learning particular concepts? That is, given an average class engaging in a typical sequence of instructional activities, do some forms of commonly used representations engender a disproportionately large number of learning difficulties?
(3) Do individual students perform consistently well or poorly with particular forms of representation with widely varying types of subject matter?
(4) Are there any consistent correlations between students' relative performance on questions posed in different representations and parameters such as major, gender, age, and learning style?

Preliminary results regarding these issues will be presented in this paper. The analysis and discussion are based on five years of classroom data, generated during the initial stages of an investigation into these issues. Ultimately, our goal is to investigate the relative effectiveness of various representations in learning; however, the initial data discussed in this paper will focus on student performance. Although these objectives are presumably closely related, it must be kept in mind that they are not identical, and that the connection between the two in the context of multiple representations must be explicitly investigated.

III. COMPARISON OF STUDENT PERFORMANCE: VERBAL VERSUS DIAGRAMMATIC VERSION OF NEWTON'S THIRD-LAW QUESTION

A. Description of questions

Two very similar questions related to Newton's third law were used to probe possible differences in students' interpretation of and performance on questions posed in different representational formats. The two questions are shown in Fig. 1(a); they were part of an 11-item quiz on gravitation, and they are numbered here according to their position on the original quiz. Question 1 is posed in a verbal (V) representation. Question 8 is posed in a diagrammatic (D) representation, making use of vector diagrams.

The quiz containing these questions was administered on the second day of class in a second-semester, algebra-based general physics course at Iowa State University. This quiz was administered in courses offered during five consecutive years, 1998–2002, during the fall semester. All students had completed the equivalent of a one-semester course focusing on mechanics, and had previous instruction related to Newton's laws with vector representations. Most took a traditional first-semester course.

The quiz did not count for a grade; students were told that it was given to help assess their level of preparation on topics that would be needed in subsequent class discussions. I will refer to this quiz as the gravitation pretest, because a second version of the same quiz was administered to the students after instruction had taken place.

B. Results

The responses to the gravitation pretest are shown in Table I.[45] Responses varied from year to year, with the percentage of correct responses ranging from 10% to 23% on question 1 (overall average: 16% correct, $N=408$) and 6% to 12% on question 8 (overall average: 9% correct). This low proportion of correct responses to a Newton's third-law question is consistent with previous research on traditional courses regarding students' belief that unequal masses in an interacting pair exert forces of unequal magnitude. It is related to a general view referred to as the "dominance principle."[46] There are two interesting and consistent discrepancies between the responses to the two questions: the significantly lower correct-response rate on the diagrammatic question ($p=0.03$ according to a two-sample t-test), and the far greater popularity on this question of a response that could be interpreted as a "larger mass exerts a smaller force" conception (response A on question 8, responses D and E on question 1).

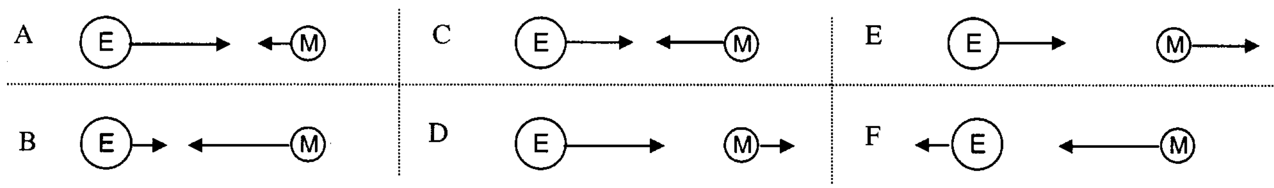
The first row of Table II shows the ratio of the number of correct responses on question 8 to that on question 1. It is particularly striking that although the proportion of correct responses (response C on both questions) varied substantially from year to year, the ratio of correct responses on one question relative to the other in a particular year is nearly constant. The range is 0.45–0.60 (the overall average is 0.53), a 33% variation that contrasts with the more than 200% year-to-year variation in the correct-response rate itself. These

(a)

#1. The mass of the sun is about 3×10^5 times the mass of the earth. How does the magnitude of the gravitational force exerted by the sun on the earth compare with the magnitude of the gravitational force exerted by the earth on the sun? The force exerted by the sun on the earth is:

A. about 9×10^{10} times larger
B. about 3×10^5 times larger
C. exactly the same
D. about 3×10^5 times smaller
E. about 9×10^{10} times smaller

#8. Which of these diagrams most closely represents the gravitational forces that the earth and moon exert ***on each other***? (Note: The mass of the earth is about 80 times larger than that of the moon.)



(b)

#1. A 5-kg lead sphere is hanging 12 m from a 500-kg lead sphere. How does the gravitational force exerted by the 5-kg sphere on the 500-kg sphere compare with the magnitude of the gravitational force exerted by the 500-kg sphere on the 5-kg sphere? The forced exerted by the 5-kg sphere on the 500-kg sphere is:

A. 100 times larger
B. 10 times larger
C. exactly the same
D. 10 times smaller
E. 100 times smaller

Fig. 1. Questions on the gravitation quiz: (a) gravitation pretest questions 1 (verbal representation) and 8 (diagrammatic representation); (b) gravitation posttest question 1. The posttest version of question 8 was unchanged from the pretest.

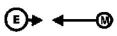
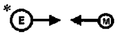
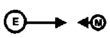
questions also were given once (in spring 2000) in the second-semester calculus-based general physics course. Although the correct-response rate was far higher on both questions in this course (62% on *V*, 38% on *D*), the ratio of the correct responses on *D* compared to *V* was consistent with the results from the algebra-based course (see the final column of Table II).

The proportion of students giving the response corresponding to "larger mass exerts a smaller force" (response A) on the *D* question also is consistently far higher than on the *V* question, as shown by the second row in Table II. Overall, this response accounted for only 5% of all responses to the *V* question, but 41% of those to the *D* question. On the gravitation pretest, those who correctly answered C on the *V* question were divided on their responses to the *D* question: 41% answered it correctly (response C), but nearly all others gave either response A (larger mass exerts a smaller force) or B (larger mass exerts a larger force), in almost equal numbers. This equally divided response pattern paralleled the behavior of the majority who had answered the *V* question incorrectly. Of all incorrect responses on the *D* question, 45% were A and 53% were B.

A posttest version of the gravitation quiz was administered approximately one week after the pretest. The posttest version of question 1 is shown in Fig. 1(b); question 8 was unchanged from the pretest. The posttest was a graded quiz. The instruction that occurred between the pre- and posttests was based on interactive-engagement methods[16] and was used to lead in to a discussion of electrical forces and fields.

The overall error rate on the posttest ($N=400$) dropped to 6% on *V* (range: 5%–8%), but only to 20% on *D* (range: 14%–25%). Even after substantial improvement in the overall correct-response rate, the significantly higher error rate on the *D* question persisted. Again, the errors on the *D* version of the question were split between the "larger mass exerts a smaller force" response A (25% of incorrect responses) and the more popular "larger mass exerts a larger force" response B (75% of incorrect responses). This preference for B

Table I. Responses to questions 1 and 8 on the gravitation pretest. For question 1, ''larger'' refers to responses A and B, ''the same'' refers to response C, and ''smaller'' refers to responses D and E. An asterisk (*) denotes the correct answer. The rate of correct responses fluctuates significantly from year to year, but the ratio of correct responses (on question 1 versus question 8) is nearly constant.

	1998	1999	2000	2001	2002
N	*78*	*96*	*83*	*77*	*74*
1. force by the sun is					
larger [A or B]	81%	83%	76%	70%	84%
*the same [C]	14%	10%	20%	23%	14%
smaller [D or E]	5%	6%	4%	6%	3%
8. earth/moon force					
 [B]	54%	45%	45%	55%	43%
*  [C]	6%	6%	12%	12%	7%
 [A]	38%	47%	41%	34%	46%
other	1%	2%	2%	0%	4%

contrasted with the much more even split observed on the pretest.[47] A large majority (81%) of the incorrect responses on the *V* posttest question were for response E, corresponding to the smaller mass exerting the smaller force. Therefore, among students who responded incorrectly, the preference for a response consistent with the dominance principle (larger mass exerts a larger force) was unchanged from the pretest.

In 2002, a pair of questions nearly identical to questions 1 and 8 in Fig. 1 was placed on the final exam of the course (see Fig. 2). These questions[48] changed the context to electrostatics, one of the major topics covered in the course. On the *D* question, students were required to explain their answer. The error rate on these questions was 9% on *V* and 14% on *D* ($N=70$). Again the errors on *D* were split almost evenly between responses A and B. Most of the written explanations for these incorrect responses were clearly consistent with a belief that the larger-magnitude charge exerts the greater-magnitude force, including 80% of the explanations given by those who had chosen response A for this question, that is, the diagram consistent with the smaller force being exerted by the larger charge. An example of an explanation given to justify choice A is that ''*Opposite charges attract. Since q_1 is the greater charge it will exert a greater force*.''

This explanation is consistent with the hypothesis that the large proportion of responses observed for the A option (smaller mass exerts a larger force) on question 8 of the gravitation quiz was due to students' confusion about whether the arrow in such diagrams represents the force exerted *on* or the force exerted *by* the object.

There also were several students who gave a correct response on the *V* question, but an incorrect response on the *D* question, and whose explanations were consistent with the dominance principle. This pattern is consistent with the observation that almost 60% of those who gave the correct response to the *V* question on the gravitation pretest from 1998 to 2002 did not correctly answer the *D* question, but instead gave an A or B response consistent either with the dominance principle or its opposite.

In 2002, 64% of the students who made errors on either the gravitation posttest or the final exam questions made representation-related errors on one or the other, but not on both tests. A representation-related error refers either to a correct answer on only one of the two (*D* and *V*) questions in the pair, or incorrect but inconsistent answers on both questions, such as B on 1 and A on 8. This observation is consistent with results regarding the consistency of students' responses, as will be discussed further in Sec. IV.

Table II. Comparison of responses on gravitation pretest: diagrammatic (*D*, question 8) versus verbal (*V*, question 1). First row: ratio of number of correct (C) responses on *D* to number of correct (C) responses on *V*; fluctuations are in a relatively narrow range. Second row: ratio of number of ''smaller than'' (A) responses on *D* to number of ''smaller than'' (D and E) responses on *V*; ratios are much greater than one, implying a consistent response discrepancy. Data for algebra-based second-semester general physics course (1998–2002) are shown. The final column shows data for a calculus-based second-semester general physics course (spring 2000), which are in good agreement with those for the algebra-based course.

Ratio of	1998	1999	2000	2001	2002	Calculus-based course (2000) $N=240$
correct on *D*/ correct on *V*	0.45	0.60	0.59	0.50	0.50	0.61
''smaller'' on *D*/ ''smaller'' on *V*	8	8	11	5	18	26

IV. MULTI-REPRESENTATIONAL QUIZZES: COMPARISON OF RESPONSES ON DIVERSE REPRESENTATIONS

A. Background

Two additional quizzes were designed to incorporate questions posed in the four representations described in the Introduction. (Note that in this context, ''graphical'' refers to bar charts and not to line graphs.)

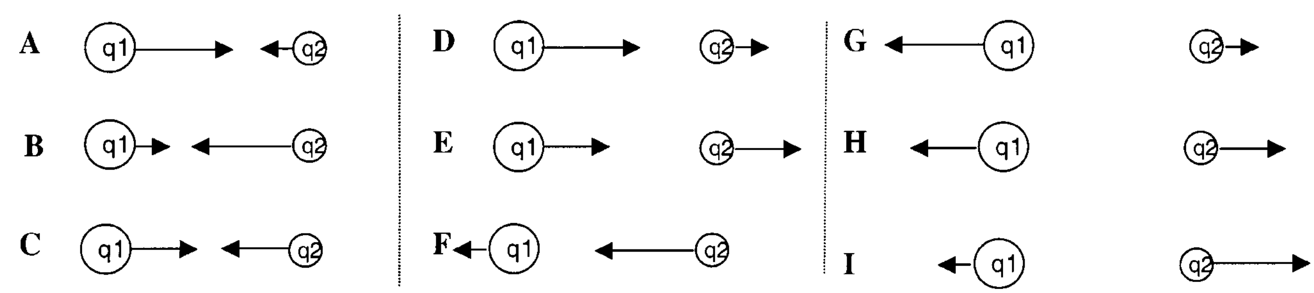
The first quiz (Appendix A, Coulomb quiz) required students to find the magnitude of the electrostatic force between two interacting charges, given the initial force and the initial and final separation distances. This quiz was administered midsemester and counted toward students' grades. The second quiz (Appendix B, circuits quiz) involved a comparison of two different two-resistor direct-current circuits, one series and one parallel. The two circuits utilize batteries of the same voltage, but the individual resistances are different. Students were required to determine whether the current through a specified resistor in the parallel circuit is greater than, equal to, or less than the current flowing through a specified resistor in the series circuit. This quiz also was administered midsemester, during 1998–2002.

The intention was to make the four questions on each quiz as nearly equal in difficulty to each other as possible. For example, the separation ratios in the Coulomb quiz (larger separation distance divided by smaller separation distance) are all small integers (2, 4, and 5), and all five answer options correspond to the same set of choices, that is, the force

1. Particle *A* has a charge that is ten times the magnitude of the charge on particle *B*. How does the magnitude of the electrical force exerted by charge *A* on the smaller charge *B* compare with the magnitude of the electrical force exerted by charge *B* on charge *A*? The force exerted by charge *A* on the smaller charge *B* is:

A. 100 times larger
B. 10 times larger
C. exactly the same
D. 10 times smaller
E. 100 times smaller

2. In the figure below, particle q1 has a charge of +10 C, and particle q2 has a charge of –2 C.



(A) *[3 points]* Which of these diagrams most closely represents the electrical forces that the two charges exert ***on each other?***

(B) *[2 points]* Explain your answer to part (A).

Fig. 2. Electrostatic version of Newton's third-law questions; administered as part of 2002 final exam.

increases or decreases by a factor equal to the separation ratio or the separation ratio squared, or no change. It is important to emphasize that by the time these quizzes were administered, the students had had extensive exposure to and practice with various questions and problems utilizing all four representations on many quizzes, exams, and homework assignments.

B. Common errors on Coulomb quiz and circuits quiz

On the Coulomb quiz, the most common error by far was the assumption that the electrical force was proportional to $1/r$, instead of $1/r^2$. This error corresponded to the response sequence B, B, D, D on questions 1–4, respectively. The proportion of all incorrect responses represented by this error was 74%, 62%, 51%, and 50%, respectively. Very few of the incorrect responses corresponded to the "no change" answer with the exception of question 2. On this question (the *D* version), the "no change" response C represented 16% of all incorrect responses. Interview data and informal discussions with students indicated that they sometimes overlooked the fact that in this question, the separation between the charges has been changed in the diagram on the right.

In 2001 non-multiple-choice variants of the *D* and *M* questions on the Coulomb quiz were given as part of a follow-up quiz (see Fig. 3). On this quiz, students were required to explain their answers to the *D* question. The nearly identical error rates on these questions (28% and 25% on *D* and *M*, respectively, disregarding explanations; $N=75$) were approximately double those on the earlier multiple-choice quiz (15% and 13%, respectively). The "$1/r$" error continued to represent the majority of incorrect responses, which was consistent with students' written explanations and algebraic work. The proportion of incorrect responses represented by this error on the follow-up quiz (76% for *D*, 58% for *M*) was comparable to that observed on the initial quiz in 2001 (64% for *D*, and 80% for *M*).

It appeared that many students who had not made the $1/r$ error on the original quiz did make this error on the follow-up quiz on one or another of the two questions. There was no clear pattern which would suggest that their error was due specifically to the form of representation. The number of

3. The diagram on the left shows two isolated particles with equal magnitude charges, along with the electrical forces acting on those particles due to their mutual interactions. The ***same*** charges are to be repositioned in the diagram on the right, this time separated from each other by ***half*** the distance that separated them in the diagram on the left. Complete the figure in the right diagram to represent the new positions and forces.

Explain how you decided on the lengths of the arrows:

4. Isolated particles with charges q_1 and q_2 ($q_1 = q_2$) are separated by distance r, and initially experience mutual interaction forces with magnitudes $F_1^{initial} = F_2^{initial} = 40$ N; $r^{initial} = 10$ m. The particles are repositioned so $r^{final} = 2\, r^{initial}$.

$F_1^{final} =$ ________

[no partial credit; answer must be within 15% of correct answer; –1 deduction for missing or incorrect units]

Fig. 3. Non-multiple-choice versions of diagrammatic and mathematical questions on the Coulomb quiz, administered as part of a follow-up quiz in 2001; numbered according to their position on the quiz.

students who switched from correct on D (on the initial quiz) to incorrect (on the follow-up quiz) was exactly the same as the number who switched from correct to incorrect on M, and the proportion who moved in the other direction—from incorrect to correct—was almost identical in the two representations. Of the students who made errors on the follow-up quiz, only 28% made consistent errors on both D and M questions (for example, making the $1/r$ error on both), while most (62%) made errors on only one of the two questions.

On the circuits quiz (Appendix B), the most common incorrect response corresponded to greater current flowing through the resistor in the series circuit (it has the smaller of the two resistances in three of the four questions), instead of the one in the parallel circuit. The proportion of all incorrect responses represented by this error was 88%, 89%, 79%, and 67%, respectively, on questions 1–4. The "equal currents" response (response B in all cases) represented 8%–15% of the incorrect responses on questions 1–3, but 30% on question 4. This difference might be due to the fact that in contrast to questions 1–3, the parallel and series resistors whose currents are being compared in question 4 are shown to be of equal resistance (instead of the parallel resistance being greater). This response pattern might imply the existence of a nonrepresentational artifact in the data.

The diagrams, algebraic work, and other notations written on students' papers were scrutinized carefully to ascertain why some students made an error on one or two questions, and yet did not do so on other questions on the same quiz. No pattern could be determined—the errors appear to occur almost randomly. This finding was consistent with observations made of students' work on all instruments employed in this study. In a further attempt to probe for any possible representation-related learning difficulties, students' responses to the quiz questions were subjected to considerable additional statistical analysis as will be described in the following.

C. Error rates

One question of interest is whether, on average, students find particular representations more difficult than others. The error rates for each question on the Coulomb and circuits quizzes are shown in Table III. There were no blank responses. "Any Error" refers to students who made errors on one or more of the questions on a given quiz, with the following exception: Students who gave four incorrect answers that were clearly consistent with each other were not counted in the "Any Error" statistic. Such a set of responses was, for instance, B, B, D, D on the Coulomb quiz, because each of these corresponded to an answer that assumed $F \propto 1/r$ (instead of $F \propto 1/r^2$). Such a set of consistent responses gives no evidence of any confusion related strictly to the representation.

The error rates are low; 31% is the highest rate observed on any of the quiz questions in any one year, and the year-to-year fluctuations are substantial. The error rates on the circuits quiz are much higher than those on the Coulomb quiz. However, the mean error rates of different representations on the same quiz differed only slightly. Moreover, the relative ranking of the four representations with respect to error rate varied from year to year, and varied between the two quizzes in the same year. No one representation yielded the highest error rate consistently for all five years on either quiz.

Statistical comparisons were made between representations using a paired two-sample t-test[49] in which the error rates on, for instance, the V question on the Coulomb quiz were compared to those for the D question on the same quiz, for the sample of five pairs of error rates, one pair for each year. Of the 12 possible comparisons, that is, V versus D, V versus M, V versus G, D versus M, D versus G, and M versus G (all six on each quiz), only one difference between

Table III. Error rates on multi-representational quizzes, in percent; the proportion of all students giving incorrect responses to each of four quiz questions. "Verbal" corresponds to question 1 on both Coulomb quiz and circuits quiz; "Diagrammatic" corresponds to question 2, Coulomb quiz and question 3, circuits quiz; "Mathematical" corresponds to question 3, Coulomb quiz and question 2, circuits quiz; "Graphical" corresponds to question 4 on both Coulomb quiz and circuits quiz. "Any Error" corresponds to students who made an error on one or more of the quiz items, not including students who gave four incorrect responses that were clearly consistent with each other (see text). Error rates in the "Average" row were calculated from cumulated total errors (1998–2002) divided by the 5-year total number of students.

All students		*N*	Verbal	Diagrammatic	Mathematical	Graphical	Any Error
Coulomb quiz	1998	*71*	4	7	10	14	24
	1999	*91*	11	15	18	21	30
	2000	*79*	14	11	10	11	24
	2001	*75*	12	15	13	23	35
	2002	*67*	15	16	24	19	33
	Average		**11**	**13**	**15**	**18**	**29**
Circuits quiz	1998	*68*	24	18	28	31	49
	1999	*88*	22	18	22	31	53
	2000	*68*	15	19	15	18	31
	2001	*75*	19	24	24	24	48
	2002	*63*	22	13	13	19	32
	Average		**20**	**19**	**20**	**25**	**43**

the means was statistically significant at the $p=0.05$ level according to a two-tailed test. This difference was on the Coulomb quiz, D versus G $(p=0.03)$.

The discrepancy that appears to be most consistent is that between the error rates on G and those on V, D, and M. The overall error rates on G, on both quizzes, are 5% higher than the combined V-D-M mean error rates on the respective quiz, while the differences among the mean error rates on V, D, and M are all $\leqslant 4\%$. This will be discussed further in Sec. V below.

D. Confidence levels

I attempted to assess students' confidence in their use of the various representations. Each question had an extra-credit option that allowed students with high confidence in the correctness of their response to gain additional points for a correct answer (see Appendices A and B). If this option is chosen, a correct answer is credited with 3.0 points instead of the 2.5 points it would be worth normally. However, there is a substantial penalty for an incorrect response. Instead of an incorrect answer being worth zero points, it is worth -1.0 points; that is, a deduction is taken from the student's total score. I analyzed students' responses on the extra-credit option to gauge their confidence with the various representations.

Students who gave a correct response but did not choose the extra-credit option are defined as giving a "low-confidence correct" response. This response suggests that although the student is able to find a correct answer, they lack full confidence in the correctness of their response. In Table IV, low-confidence correct responses are tabulated for each question on each quiz.

On both quizzes, the proportion of low-confidence correct responses on the V question is lower than that on the three other questions on the same quiz. The differences are not large, and so I tested the significance of the differences between low-confidence correct response rates on the V questions and those on the D, M, and G questions by employing a paired t-test. Each sample consisted of the five pairs (one for each year) of the error rates on the V question, and either the D, M, and G question, respectively, for a total of six comparisons (three for each quiz). The difference between the means was found significant at the $p\leqslant 0.01$ level (one-tailed test) for the V-D and V-G comparison on the Coulomb quiz, and $p\leqslant 0.05$ for the V-M and V-G comparison

Table IV. Correct but low-confidence responses: the proportion of students giving correct response but not choosing extra-credit option.

	1998–2002	Verbal	Diagrammatic	Mathematical	Graphical
Coulomb quiz	Number correct	340	333	326	315
	Low-confidence correct	17%	24%	22%	24%
Circuits quiz	Number correct	289	295	288	272
	Low-confidence correct	33%	37%	41%	45%

Table V. Consistency of responses: the students who took both quizzes and made one, two, or three errors on at least one quiz. A ''repeat'' error refers to an error on both quizzes for questions in a particular form of representation; ''≤50% repeat errors'' indicates that half or fewer of all incorrectly used representations (combined for both quizzes) were part of a repeat-error pair (see text). (Students who gave four incorrect but consistent responses on a single quiz as defined in the text were not counted as having made any errors on that quiz for the purposes of this tabulation.)

	N	Errors on one quiz only (*no* repeat errors)	Errors on both quizzes but *no* repeat errors	Errors on both quizzes, but ≤50% repeat errors	Errors on both quizzes, >50% repeat errors
2000	*23*	78%	9%	9%	4%
2001	*44*	73%	7%	14%	7%
2002	*26*	77%	12%	8%	4%

on the circuits quiz. Corresponding values for the remaining comparisons were $p = 0.10$ (V-M on the Coulomb quiz), and $p = 0.12$ (V-D on the circuits quiz). These results suggest that students had slightly greater confidence when responding correctly to questions posed in the V (''words only'') representation on these two quizzes. In comparison, among students responding *incorrectly*, lower-than-average confidence was associated with D and M responses on the circuits quiz.

E. Consistency of students' error

To explore whether a given student consistently made errors with the same form of representation, a subset of the data was examined in more detail. For the years 2000, 2001, and 2002, a tabulation was made of students who took both quizzes and made one, two, or three errors on at least one quiz. When students made four errors, there is no direct evidence as to whether they have—or have not—made a representation-related error (in contrast to a physics error).

Table VI. (a) Error rates on multi-representational quizzes, in percent; male students only. (b) Error rates on multi-representational quizzes, in percent; female students only.

		N	Verbal	Diagrammatic	Mathematical	Graphical	Any Error
			(a)				
Males							
Coulomb quiz	1998	*27*	7	7	7	11	26
	1999	*36*	6	11	11	11	14
	2000	*32*	13	16	9	13	22
	2001	*30*	10	10	10	10	31
	2002	*30*	17	10	30	20	30
	Average		**10**	**11**	**14**	**13**	**24**
Circuits quiz	1998	*27*	26	11	33	33	52
	1999	*35*	9	14	14	29	49
	2000	*29*	14	14	14	21	31
	2001	*28*	18	21	21	14	43
	2002	*28*	14	11	14	11	29
	Average		**16**	**14**	**19**	**22**	**41**
			(b)				
Females							
Coulomb quiz	1998	*44*	2	7	11	16	23
	1999	*55*	15	18	22	27	40
	2000	*47*	15	9	11	11	26
	2001	*45*	13	18	16	31	38
	2002	*37*	14	22	19	19	35
	Average		**12**	**14**	**16**	**21**	**32**
Circuits quiz	1998	*41*	22	22	24	29	46
	1999	*53*	30	21	26	32	57
	2000	*39*	15	23	15	15	31
	2001	*47*	19	26	26	30	51
	2002	*35*	29	14	11	26	34
	Average		**23**	**21**	**21**	**27**	**45**

Therefore, students who made four errors on either quiz (a very small proportion of students overall) are not counted in this tabulation. In contrast, students who gave four incorrect but consistent responses on a particular quiz were not counted as having made any errors on that quiz for the purposes of this analysis. These data are shown in Table V. A "repeat" error refers to an error on both quizzes for questions in a particular representation. If students made errors on V, D, and M on one quiz and D, M, and G on the other, 50% of their errors (two $[D,M]$ out of four $[V,D,M,G]$) are considered to be repeats. The statement "$\leqslant$50% repeat errors" in Table V indicates that half or fewer of all incorrectly used representations were part of a repeat-error pair.

The results of the three years are very consistent: most students made errors on one quiz only. Of those who made errors on both quizzes, most did not repeat the same error. That is, they did not make two errors using the same representation. If they did repeat an error, half or fewer of their representation errors were repeated. These data do not support the hypothesis that students tend to err consistently in one or another representation.

V. GENDER-RELATED DIFFERENCES

In Table VI, error rate data are shown for male, Table VI (a), and female, Table VI (b), students. This breakdown allows us to test for possible gender-related differences. We see that the mean error rates (average values, all years combined) for the female students are higher than those of the males, on all questions on both quizzes. In most cases, the male-female difference is relatively small. To gauge the statistical significance of the differences, a paired t-test was carried out separately for each question on each quiz, where each sample consisted of five pairs of values (male error rate, female error rate), one pair for each year.[49] This test also was done for the "Any Error" rate. Of these ten cases, the only difference in the mean error rate significant at the $p=0.05$ level with a two-tailed test was the D question on the circuits quiz (male: 14%, female: 21%, $p=0.008$). Due to the low statistical power of a test with a sample of only five pairs, and in view of the consistency of the observed male–female error rate difference, it may be more appropriate to use a $p \leqslant 0.10$ criterion and apply a one-tailed test. Two additional cases met that criterion: Coulomb quiz, G question (male: 13%, female: 21%, $p=0.08$), and Coulomb quiz, any error (male: 24%, female: 32%, $p=0.09$).

A noticeable contrast between the Table VI and Table III data is that the difference among the male students between the G error rate on the Coulomb quiz (13%) and the mean combined V-D-M error rate on the same quiz (12%) is much smaller than the corresponding difference in the "all students" sample (Table III). In contrast, a sizeable difference still exists for the female students (G: 21%; V-D-M: 14%). This observation suggests that the larger error rate on G (relative to V-D-M) in Table III is primarily due to the female students. It is not as clear whether this pattern may be true for the circuits quiz as well, for here a discrepancy is still present for males (G: 22%, V-D-M: 16%), as well as for females (G: 27%, V-D-M: 22%).

To examine this question more closely, I did three statistical tests. To probe the statistical significance of the observation that the G error rates are higher than V, D, or M error rates on the same quiz during the same year, I employed a Wilcoxon sign rank test.[50] This is a nonparametric test that does not depend on the shape of the distribution of sample values, and thus is less sensitive to deviations from normality in the data sample. In this test I considered all pairwise comparisons between the G error rate and the V, D, and M error rates, respectively, on a given quiz for a given year. This procedure yielded 15 comparisons on each quiz (three for each year), both for males and females. For instance, for male students on the Coulomb quiz, the G-V, G-D, and G-M pairs for 2000 were (0.13, 0.13), (0.13, 0.16), and (0.13, 0.09). For female students during the same year, the pairs were (0.11, 0.15), (0.11, 0.09), and (0.11, 0.11). The four samples and their resulting p values (for a two-tailed test) are Coulomb-male, $p>0.10$; Coulomb-female, $p<0.01$; Circuits-male, $p>0.10$; and Circuits-female, $p<0.02$; each sample consisted of 15 pairs of values. These results suggest that the error rates for females might be higher on G questions than on V-D-M questions.

A paired two-sample t-test was used to make a full set of 12 interrepresentation comparisons, separately for males and females. There were six on each quiz, that is, V versus D, V versus M, V versus G, D versus M, D versus G, and M versus G. Each sample consisted of five pairs of values, one for each year. No interrepresentation differences were found to be significant at the $p=0.05$ level using a two-tailed test. Several comparisons were significant at the $p \leqslant 0.10$ level using a one-tailed test; all p values corresponding to the one-tailed test are shown in Table VII.

Table VII. p values for statistical tests (one-tailed test) of the significance of differences between mean error rates on questions from the same quiz posed in different representations. The paired t-test and the test for correlated proportions are described in the text. These p values represent the probability that differences in mean error rates equal to or larger than those actually observed (but with the same sign) would occur in an ensemble of paired random samples of the same size, drawn from an infinitely large population in which the true difference in mean error rates is zero.

		Coulomb quiz		Circuits quiz	
		Paired t-test	Correlated proportions	Paired t-test	Correlated proportions
Females	G versus V	0.04	0.001	0.12	⋯
	G versus D	0.05	0.02	0.10	0.05
	G versus M	0.07	0.04	0.03	0.07
	V versus D	0.15	⋯	0.34	⋯
	V versus M	0.08	0.08	0.29	⋯
	D versus M	0.26	⋯	0.42	⋯
Males	G versus V	0.04	0.23	0.14	⋯
	G versus D	0.20	⋯	0.12	⋯
	G versus M	0.40	⋯	0.31	⋯
	V versus D	0.43	⋯	0.32	⋯
	V versus M	0.17	⋯	0.04	0.18
	D versus M	0.29	⋯	0.15	⋯

To examine these possibly significant comparisons more closely, a test for the difference between correlated proportions was applied.[51] With this method a test statistic z is calculated by comparing, for instance, the number of students (all five years) who were correct on the G question but incorrect on the V question (C_{GV}) to those who were incorrect on the G question but correct on the V question (C_{VG}). After applying a continuity correction,[52] we have $z=(|C_{GV}-C_{VG}|-1)/(C_{GV}+C_{VG})^{0.5}$. The calculated p values resulting from this statistic are shown in Table VII for those pairs that met the $p \leq 0.10$ criterion on the t-test.

Even with this wealth of statistical data, the conclusions remain ambiguous. However, the various results support the hypothesis that there is a discrepancy between the male and female students regarding the relative error rates on G questions in comparison to V-D-M questions, at least on the Coulomb quiz. On this quiz, the female students did more poorly on G questions in comparison to V-D-M questions, whereas the male students did not, or at least not as much. There also was support (noted above) for the hypothesis that female students perform more poorly on the diagrammatic question on the circuits quiz, in comparison to male students. Because the male and female students in this study received identical instruction, these results are potentially significant.

VI. DISCUSSION

A. Newton's third-law questions

The analysis of the gravitation quiz data leaves no doubt that there is a systematic discrepancy among students in this sample between their interpretation of the verbal and diagrammatic versions of the Newton's third-law question. Although the correct-response rate on the pretest version of the two questions varied substantially from year to year, the rate of correct responses on the diagrammatic version was never greater than 60% of that on the verbal version. A substantial majority (59%) of students who correctly answered the verbal version gave an incorrect response on the diagrammatic version. In the latter context they were influenced by the dominance principle that had not, apparently, determined their response to the verbal version. Written explanations on the electrostatic version of these questions on the 2002 final exam are consistent with this interpretation, although they do not directly support it.[53] (It is notable, however, that of the students who correctly answered the diagrammatic version of this question on the pretest, only 23% gave an incorrect response to the verbal version on the same test.)

Over the five years of this study, 59% of students who answered the Newton's third-law pretest question with a correct "equal-force" response on the verbal representation gave an "unequal-force" response on the diagrammatic representation. Yet the total number of such students is relatively small in comparison to the size of the full sample since only 16% of all students gave a correct response on the verbal pretest question. This discrepancy in response rates demonstrates how sharply divergent students' responses may be in different contexts[54]—even when the context is merely a different representation accompanied by slightly different wording.[55] However, this particular divergence is not representative of a large fraction of the student population. In contrast, the error corresponding to the "larger mass exerts a larger (smaller) force" response (described below) is one that characterizes a sizeable fraction—perhaps more than a third—of this population.

It was observed that response A on the diagrammatic question 8 of the gravitation quiz—what we call an "antidominance principle" response (larger mass exerts a smaller force)—represents more than 40% of responses to this question, while the corresponding D and E responses on the verbal question 1 represent only 5% of all responses to this question. The implication is that many students have an incorrect understanding of vector arrow conventions, that is, the arrow whose tail is attached to an object represents the force that is exerted on that object, not by it. This implication is strongly supported by the written explanations offered by students on the 2002 final exam questions.[56]

These observations are intriguing and important, and yet leave unanswered questions. What is still unclear is the precise nature of students' thinking that leads some to answer that the gravitational forces exerted by the sun and earth on each other are of equal magnitude, and yet moments later to select a vector diagram in which the interaction forces of earth and moon are clearly not the same. Similarly, the details of students' thinking regarding the representation of forces exerted on or by an object are not well understood. It is possible that confusion related to the specific words or phrases used in the gravitation questions has contributed to the differences observed in students' responses, independent of confusion introduced by the diagrammatic representation. Our experience suggests that extensive interviewing will be required to clarify these matters.

B. Multi-representational quizzes

The mean error rates on the Coulomb and circuits quizzes were consistently low (below 30% on each question), and year-to-year variations were high (up to 400%). These facts imply that statistical conclusions from this data set will have limited reliability. In particular, it would not be reasonable to generalize conclusions from these data to problem sets of significantly greater difficulty without further investigation. Most students in this data sample did not make errors on the test questions; therefore, one could argue that the interrepresentational competence of a substantial fraction of the population sample was not directly probed by these instruments. More difficult test questions (including non-multiple-choice items) that could probe a larger fraction of the population sample might yield conclusions that are different than, and even contradictory to, those discussed here.

Most students in this sample did not show a pattern of consistent representation-related errors on the multi-representational quizzes. The specific physics errors made by students were quite consistent; as discussed in Sec. IV, a large proportion of incorrect responses were concentrated on just one conceptual error on each quiz. However, the typical student made errors on only one or two questions (or none), and gave correct answers on the other questions. They typically did not make an error with the same representation on both quizzes, and this pattern of no repeat errors was consistent with results on the Newton's third-law questions discussed in Sec. III. The precise trigger that led a student to make a "standard" physics error when using one particular representation on a particular quiz—and not with any other representations, nor on a follow-up quiz—is unclear, and appeared to be almost random, both for individual students and for the students as a whole. On the Coulomb questions in 2001, for example, the number of students getting a *D* question incorrect later in the semester (after they had already answered it correctly earlier in the semester) was exactly matched by the number of students displaying the same pattern with the *M* questions. (See Sec. IV B).

There is evidence for slightly higher confidence rates on the verbal questions. This finding might surprise some, because many physics instructors would find the verbal version of the quiz questions to be awkward to interpret and analyze, in comparison to the *D*, *M*, and *G* versions based on very familiar and long practiced representations. This result suggests that the instructor's view of the ease or difficulty of a particular representation in a particular context might not match the views of a large proportion of students. The results of previous investigations regarding student understanding of kinematics diagrams[18,28–30] are consistent with this inference.

C. Gender differences

On the multi-representational quizzes, there is evidence that student performance on the *G* questions was slightly inferior to that on the *V*, *D*, and *M* questions. However, this evidence is strong only for female students on the Coulomb quiz. The poorer performance on *G* questions might be ascribed to less familiarity and practice with this representation. However, the instruction for both females and males was identical, and the relatively poorer performance by females on the *G* questions, at least on the Coulomb quiz, suggests a genuine performance discrepancy between the genders in the larger population. Whether this discrepancy may be due to different degrees of previous experience with *G* representations or some other cause is a matter for speculation. Similarly, the substantial evidence for poorer performance by females on the circuit-diagram question (*D* question; female error rate $=21\%$; male error rate $=14\%$) cannot be explained based on available information. The slightly higher error rates by females overall, in comparison to males, are not statistically significant for the most part.[57]

VII. CONCLUSION

We can summarize the results of this investigation as follows: (1) Some students did give inconsistent answers to the same question when it was asked using different representations; however, there was no clear evidence of a consistent pattern of representation-related errors among individual students. (2) Specific difficulties were noted when using vector representations in the context of Newton's third law. Many students apparently lacked an understanding of how to use vector arrows to distinguish forces acting *on* an object from forces exerted *by* that object. An apparently different difficulty was reflected by a smaller, though still substantial, number of students who gave a correct "equal-force" answer to a verbal question but an incorrect "unequal-force" answer to a very similar question using vector diagrams. (3) There was substantial evidence that females had a slightly higher error rate on graphical (bar chart) questions in comparison to verbal, diagrammatic, and mathematical questions, whereas the evidence for male students was more ambiguous. (4) Some evidence of possible gender-related differences was identified. Specifically, a possible difficulty related to electric circuit diagrams has been identified for females in comparison to males.

Although the observed error rate differences among the different representations were quite small or statistically insignificant in general, this result was in the context of a course that emphasized the use of multiple representations in all class activities. In addition, the overall error rates were quite low and suggest that the questions were too simple to probe possible representation-related difficulties among the majority of the students. What results might be found for students in a more traditional course which focuses on mathematical representations is an open question, as is the question of what results might be observed if significantly more challenging problems were posed.

However, this preliminary investigation has yielded at least one dramatic example of how student performance on very similar physics problems posed in different representations might yield strikingly different results (gravitation quiz, questions 1 and 8).[58] This "existence proof" serves as a caution that potential interrepresentational discrepancies in student performance must be carefully considered in the design and analysis of classroom exams and diagnostic test instruments. (This idea is already implicit in the work of many other authors cited in this paper.) For instance, if students are observed to make errors on Coulomb's law questions using a vector representation, representational confusion would be signaled by correct answers on closely related conceptual questions using other representations.

The evidence provided here for possible gender-related discrepancies in interrepresentational performance suggests that substantial additional investigation of this possibility is warranted, with a view toward possible implementation of appropriately modified instructional strategies. Many unanswered questions regarding the details of students' reasoning when using diverse representations must await more extensive data from interviews and analysis of students' written explanations.

ACKNOWLEDGMENTS

I am indebted to Leith Allen for many fruitful conversations and valuable insights regarding this work, and in particular for emphasizing the significance of the "larger mass exerts a smaller force" response discrepancy, and for designing the electrostatic version of the Newton's third-law problem discussed in Sec. III. She also carried out a series of interviews that added perspective to the analysis presented here, and carefully reviewed the manuscript. Larry Engelhardt carried out a series of interviews that shed additional light on the issues examined in this paper. Jack Dostal contributed to the analysis of the data from the gravitation quiz. This material is based in part on work supported by the National Science Foundation under Grant No. REC-0206683; this project is in collaboration with Thomas J. Greenbowe, co-principal investigator.

APPENDIX A

Coulomb quiz. Designations of representations, and correct answers: 1, Verbal, answer: A; 2, Diagrammatic, answer: A; 3, Mathematical, answer: E; 4, Graphical, answer: E.

Physics 112
Quiz #11
October 6, 2000

Name: ____________________________

IF YOU WANT A QUESTION GRADED OUT OF THREE POINTS (–1 [MINUS ONE] FOR WRONG ANSWER!!) WRITE "3" IN SPACE PROVIDED ON EACH QUESTION.

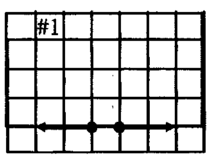
1. When two identical, isolated charges are separated by two centimeters, the magnitude of the force exerted by each charge on the other is eight newtons. If the charges are moved to a separation of eight centimeters, what will be the magnitude of that force now?

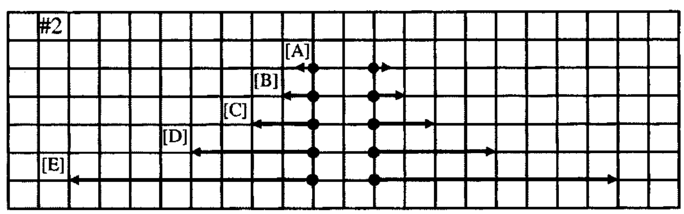
A. one-half of a newton
B. two newtons
C. eight newtons
D. thirty-two newtons
E. one hundred twenty-eight newtons

Grade out of three? Write "3" here: ______

2. Figure #1 shows two identical, isolated charges separated by a certain distance. The arrows indicate the forces exerted by each charge on the other. The same charges are shown in Figure #2. Which diagram in Figure #2 would be correct?

A.
B.
C.
D.
E.





Grade out of three? Write "3" here: ______

3. Isolated charges q_1 and q_2 are separated by distance r, and each exerts force F on the other. $q_1^{initial} = q_1^{final}$ and $q_2^{initial} = q_2^{final}$; $r^{initial} = 10m$; $r^{final} = 2m$. $F^{initial} = 25N$; $F^{final} = ?$

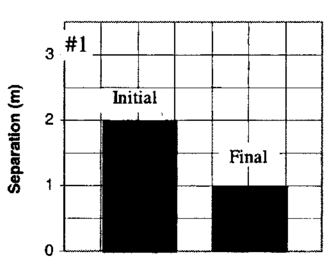
A. 1 N
B. 5 N
C. 25 N
D. 125 N
E. 625 N

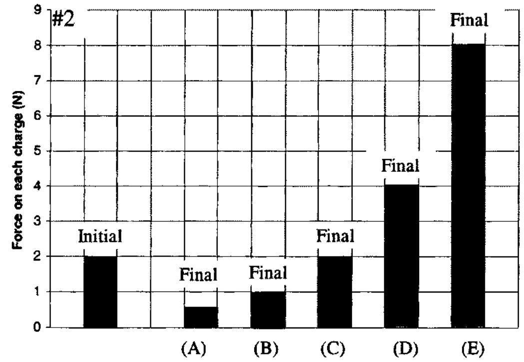
Grade out of three? Write "3" here: ______

4. Graph #1 refers to the initial and final separation between two identical, isolated charges. Graph #2 refers to the initial and final forces exerted by each charge on the other. Which bar is correct?

Grade out of three? Write "3" here: ______

A.
B.
C.
D.
E.





APPENDIX B

Circuits quiz. Designations of representations, and correct answers: 1, Verbal, answer: A; 2, Mathematical, answer: A; 3, Diagrammatic, answer: A; 4, Graphical, answer: C.

Physics 112
Quiz #16
October 27, 2000

Name: ____________________

IF YOU WANT A QUESTION GRADED OUT OF THREE POINTS (–1 [MINUS ONE] FOR WRONG ANSWER!!) WRITE "3" IN SPACE PROVIDED ON EACH QUESTION.

1. In a parallel circuit, a three-ohm resistor and a six-ohm resistor are connected to a battery. In a series circuit, a four-ohm and an eight-ohm resistor are connected to a battery that has the ***same*** voltage as the battery in the parallel circuit. What will be the ratio of the current through the six-ohm resistor to the current through the four-ohm resistor? Current through six-ohm resistor divided by current through four-ohm resistor is:

   A. greater than one
   B. equal to one
   C. less than one
   D. equal to negative one
   E. cannot determine without knowing the battery voltage

   Grade out of 3? Write "3" here: _____

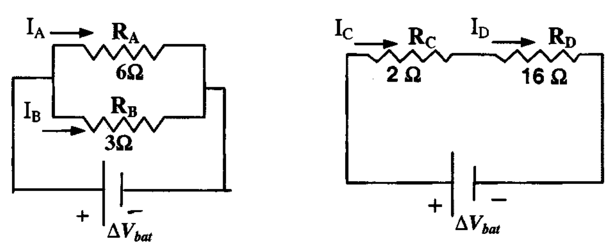
2. Parallel circuit: $R_A = 6\ \Omega$; $R_B = 9\ \Omega$.
   Series circuit: $R_C = 7\ \Omega$; $R_D = 3\ \Omega$.
   $\Delta V_{bat}(series) = \Delta V_{bat}(parallel)$

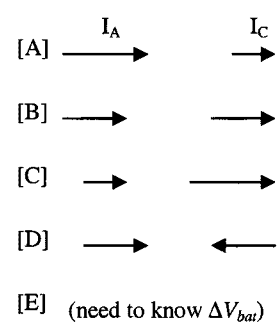
   A. $\frac{I_B}{I_C} > 1$ B. $\frac{I_B}{I_C} = 1$ C. $\frac{I_B}{I_C} < 1$ D. $\frac{I_B}{I_C} = -1$ E. *need* ΔV_{bat}

   Grade out of 3? Write "3" here: _____

3. The arrows represent the magnitude and direction of the current through resistors A and C. Choose the correct diagram.

   A.
   B.
   C.
   D.
   E. need to know ΔV_{bat}

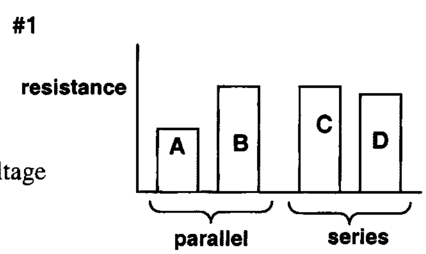


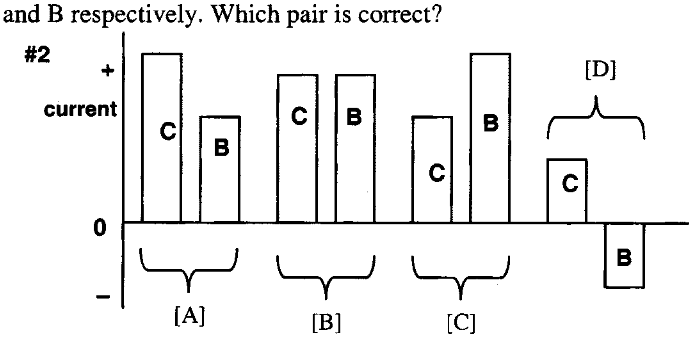


Grade out of 3? Write "3" here: _____

4. Graph #1 represents the relative resistances of resistors A, B, C, and D. Resistors A and B are connected in a parallel circuit. Resistors C and D are connected in a series circuit. The battery voltage in both circuits is the same. Graph #2 represents the currents in resistors C and B respectively. Which pair is correct?

A.
B.
C.
D.
E. need to know voltage





Grade out of 3? Write "3" here _____

[a]Electronic mail: dem@iastate.edu
[1]Alan Van Heuvelen, ''Learning to think like a physicist: A review of research-based instructional strategies,'' Am. J. Phys. **59**, 891–897 (1991); ''Overview, Case Study Physics,'' *ibid.* **59**, 898–907 (1991).
[2]David Hestenes, ''Modeling methodology for physics teachers,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 935–957 (1997)], pt. 2.
[3]Lillian C. McDermott, ''A view from physics,'' in *Toward a Scientific Practice of Science Education*, edited by M. Gardner, J. G. Greeno, F. Reif, A. H. Schoenfeld, A. diSessa, and E. Stage (L. Erlbaum, Hillsdale, NJ, 1990), pp. 3–30.
[4]In this investigation we will concentrate on the representations V, D, M, and G. In this paper G refers to bar charts and not to line graphs. ''Graphical'' is used in a broad sense to refer to bar charts because they often are grouped together with line graphs, but there are very significant differences between the two representations that would have to be considered in future work. We restrict ourselves primarily to these four representations for practical and logistical reasons. There are certainly other pedagogically significant representations, for example, pictorial representations, computer animations and simulations, haptic (sense of touch) and kinesthetic interfaces and representations, video recordings of actual physical processes, and actual physical objects and systems using laboratory equipment. All of these are under investigation by many research groups (including ours), but they lack the relative standardization (due to long-term use in instruction) and ease and flexibility of implementation that characterize V, D, M, and G. Historically, V, D, M, and G have been ubiquitous in scientific work.
[5]M. T. H Chi, P. J. Feltovich, and R. Glaser, ''Categorization and representation of physics problems by experts and novices,'' Cogn. Sci. **5**, 121–152 (1981); Yuichiro Anzai, ''Learning and use of representations for physics expertise,'' in *Toward a General Theory of Expertise*, edited by K. Anders Ericsson and Jacqui Smith (Cambridge U. P., Cambridge, 1991), pp. 64–92.
[6]Jill H. Larkin, ''The role of problem representation in physics,'' in *Mental Models*, edited by Dedre Gentner and Albert L. Stevens (L. Erlbaum, Hillsdale, NJ, 1983), pp. 75–98.
[7]David P. Maloney, ''Research on problem solving: Physics,'' in *Handbook of Research on Science Teaching and Learning*, edited by Dorothy L. Gabel (Macmillan, New York, 1993), pp. 327–354.
[8]R. Kleinman, H. Griffin, and N. K. Kerner, ''Images in chemistry,'' J. Chem. Educ. **64**, 766–770 (1987); Robert Kozma, Elaine Chin, Joel Russell, and Nancy Marx, ''The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning,'' J. Learn. Sci. **9**, 105–143 (2000).
[9]Frederick Reif, ''Millikan Lecture 1994: Understanding and teaching important scientific thought processes,'' Am. J. Phys. **63**, 17–32 (1995).
[10]Alan Van Heuvelen, *ALPS Kit: Active Learning Problem Sheets, Mechanics; Electricity and Magnetism* (Hayden-McNeil, Plymouth, MI, 1990); Ruth W. Chabay and Bruce A. Sherwood, *Matter & Interactions I, II* (J. Wiley, New York, 2002); Frederick Reif, *Understanding Basic Mechanics* (J. Wiley, New York, 1995); Lillian C. McDermott and the Physics Education Group, *Physics by Inquiry* (J. Wiley, New York, 1996); Randall D. Knight, *Physics for Scientists and Engineers: A Strategic Approach* (Pearson Addison-Wesley, San Francisco, 2004); Eric Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997); Gregor M. Novak, Evelyn T. Patterson, Andrew D. Gavrin, and Wolfgang Christian, *Just-In-Time Teaching: Blending Active Learning with Web Technology* (Prentice Hall, Upper Saddle River, NJ, 1999); Thomas L. O'Kuma, David P. Maloney, and Curtis Hieggelke, eds., *Ranking Task Exercises in Physics* (Prentice Hall, Upper Saddle River, NJ, 2000); Wolfgang Christian and Mario Belloni, *Physlets: Teaching Physics with Interactive Curricular Material* (Prentice Hall, Upper Saddle River, NJ, 2001); Lillian C. McDermott, Peter S. Shaffer, and the Physics Education Group, *Tutorials in Introductory Physics* (Prentice Hall, Upper Saddle River, NJ, 2002-2003).
[11]Fred Goldberg, ''Constructing physics understanding in a computer-supported learning environment,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 903–911 (1997)], pt. 2.
[12]R. R. Hake, ''Promoting student crossover to the Newtonian world,'' Am. J. Phys. **55**, 878–884 (1987); Richard R. Hake, ''Socratic pedagogy in the introductory physics laboratory,'' Phys. Teach. **30**, 546–552 (1992); Ronald K. Thornton and David R. Sokoloff, ''Learning motion concepts using real-time microcomputer-based laboratory tools,'' Am. J. Phys. **58**, 858–867 (1990); Priscilla W. Laws, ''Calculus-based physics without lectures,'' Phys. Today **44(12)**, 24–31 (1991); ''Millikan Lecture 1996: Promoting active learning based on physics education research in introductory physics courses,'' Am. J. Phys. **65**, 14–21 (1997).
[13]Patricia Heller and Mark Hollabaugh, ''Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups,'' Am. J. Phys. **60**, 637–644 (1992).
[14]Robert J. Beichner, ''The impact of video motion analysis on kinematics graph interpretation skills,'' Am. J. Phys. **64**, 1272–1277 (1996).
[15]Robert J. Dufresne, William J. Gerace, and William J. Leonard, ''Solving physics problems with multiple representations,'' Phys. Teach. **35**, 270–275 (1997); Lawrence T. Escalada and Dean A. Zollman, ''An investigation on the effects of using interactive digital video in a physics classroom on student learning and attitudes,'' J. Res. Sci. Teach. **34**, 467–489 (1997); Melissa Dancy, Wolfgang Christian, and Mario Belloni, ''Teaching with Physlets: Examples from optics,'' Phys. Teach. **40**, 494–499 (2002); Anne J. Cox, Mario Belloni, Melissa Dancy, and Wolfgang Christian, ''Teaching thermodynamics with Physlets in introductory physics,'' Phys. Educ. **38**, 433–440 (2003).
[16]David E. Meltzer and Kandiah Manivannan, ''Promoting interactivity in physics lecture classes,'' Phys. Teach. **34**, 72–78 (1996); ''Transforming the lecture-hall environment: The fully interactive physics lecture,'' Am. J. Phys. **70**, 639–654 (2002).
[17]K. A. Burke, Thomas J. Greenbowe, and Mark A. Windschitl, ''Developing and using conceptual computer animations for chemistry instruction,'' J. Chem. Educ. **75**, 1658–1661 (1998); Thomas J. Greenbowe, ''An interactive multimedia software program for exploring electrochemical cells,'' J. Chem. Educ. **71**, 555–557 (1994); Michael J. Sanger and Thomas J. Greenbowe, ''Addressing student misconceptions concerning electron flow in aqueous solutions with instruction including computer animations and conceptual change strategies,'' Int. J. Sci. Educ. **22**, 521–537 (2000); Hsin-kai Wu, Joseph S. Krajcik, and Elliot Soloway, ''Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom,'' J. Res. Sci. Teach. **38**, 821–842 (2001).
[18]Robert J. Beichner, ''Testing student interpretation of kinematics graphs,'' Am. J. Phys. **62**, 750–762 (1994).
[19]John Clement, ''Observed methods for generating analogies in scientific problem solving,'' Cogn. Sci. **12**, 563–586 (1988).
[20]Rolf Plötzner, *The Integrative Use of Qualitative and Quantitative Knowledge in Physics Problem Solving* (Peter Lang, Frankfurt am Main, 1994), pp. 33–46.
[21]Ronald K. Thornton and David 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**, 338–352 (1998).
[22]Alan Van Heuvelen and Xueli Zou, ''Multiple representations of work-energy processes,'' Am. J. Phys. **69**, 184–194 (2001); Xueli Zou, ''The role of work-energy bar charts as a physical representation in problem solving,'' in *Proceedings of the 2001 Physics Education Research Conference*, edited by Scott Franklin, Jeffrey Marx, and Karen Cummings (PERC, Rochester, NY, 2001), pp. 135–138.
[23]Allan Paivio, *Imagery and Verbal Processes* (Holt, Rinehart and Winston, New York, 1971).
[24]Claude Janvier, ed., *Problems of Representation in the Teaching and Learning of Mathematics* (L. Erlbaum, Hillsdale, NJ, 1987).
[25]Richard Lesh, Tom Post, and Merlyn Behr, ''Representations and translations among representations in mathematics learning and problem solving,'' in *Problems of Representation in the Teaching and Learning of Mathematics*, edited by Claude Janvier (L. Erlbaum, Hillsdale, NJ, 1987), pp. 33–40; Paul White and Michael Mitchelmore, ''Conceptual knowledge in introductory calculus,'' J. Res. Math. Educ. **27**, 79–95 (1996); Peter C.-H. Cheng, ''Unlocking conceptual learning in mathematics and science with effective representational systems,'' Comput. Educ. **33**, 109–130 (1999); Shaaron Ainsworth, ''The functions of multiple representations,'' *ibid.* **33**, 131–152 (1999).
[26]Robert B. Kozma and Joel Russell, ''Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena,'' J. Res. Sci. Teach. **34**, 949–968 (1997).
[27]Donald R. Jones and David A. Schkade, ''Choosing and translating be-

tween problem representations,'' Organ. Behav. Human Decision Process. **61**, 214–223 (1995).

[28]Janice R. Mokros and Robert F. Tinker, ''The impact of microcomputer-based labs on children's ability to interpret graphs,'' J. Res. Sci. Teach. **24**, 369–383 (1987); Lillian C. McDermott, Mark L. Rosenquist, and Emily H. Van Zee, ''Student difficulties in connecting graphs and physics: Examples from kinematics,'' Am. J. Phys. **55**, 503–513 (1987); Fred M. Goldberg and John H. Anderson, ''Student difficulties with graphical representation of negative values of velocity,'' Phys. Teach. **27**, 254–260 (1989).

[29]Craig A. Berg and Philip Smith, ''Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments,'' Sci. Educ. **78**, 527–554 (1994).

[30]Italo Testa, Gabriella Mouroy, and Elena Sassi, ''Students' reading images in kinematics: The case of real-time graphs,'' Int. J. Sci. Educ. **24**, 235–256 (2002).

[31]For example, see: G. J. Hitch, M. C. Beveridge, S. E. Avons, and A. T. Hickman, ''Effects of reference domain in children's comprehension of coordinate graphs,'' in *The Acquisition of Symbolic Skills*, edited by Don Rogers and John A. Sloboda (Plenum, New York, 1982), pp. 551–560.

[32]Norman H. Fredette and John J. Clement, ''Student misconceptions of an electric circuit: What do they mean?,'' J. Coll. Sci. Teach. **10**, 280–285 (1981); Samuel Johsua, ''Students' interpretation of simple electrical diagrams,'' Eur. J. Sci. Educ. **6**, 271–275 (1984); Lillian C. McDermott and Peter S. Shaffer, ''Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding,'' Am. J. Phys. **60**, 994–1003 (1992); erratum, **61**, 81 (1993); Peter S. Shaffer and Lillian C. McDermott, ''Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of instructional strategies,'' *ibid.* **60**, 1003–1013 (1992); S. Törnkvist, K.-A. Pettersson, and G. Tranströmer, ''Confusion by representation: On student's comprehension of the electric field concept,'' *ibid.* **61**, 335–338 (1993); Randal Robert Harrington, ''An investigation of student understanding of electric concepts in the introductory university physics course,'' Ph.D. dissertation, University of Washington (UMI, Ann Arbor, MI, 1995), UMI #9537324, Chap. 5; Stephen Emile Kanim, ''An investigation of student difficulties in qualitative and quantitative problem solving: Examples from electric circuits and electrostatics,'' Ph.D. dissertation, University of Washington (UMI, Ann Arbor, MI, 1999), UMI #9936436, Chaps. 4–7; Leith Dwyer Allen, ''An investigation into student understanding of magnetic induction,'' Ph.D. dissertation, The Ohio State University (UMI, Ann Arbor, MI, 2001), UMI #3011018, Chap. 6; Rasil Warnakulasooriya and Lei Bao, ''Towards a model-based diagnostic instrument in electricity and magnetism–an example,'' *Proceedings of the 2002 Physics Education Research Conference [Boise, Idaho, August 7-8, 2002]*, edited by Scott Franklin, Karen Cummings, and Jeffrey Marx (PERC, New York, 2002), pp. 83–86.

[33]J. Ramadas, ''Use of ray diagrams in optics,'' School Sci. **10**, 1–8 (1982); Fred M. Goldberg and Lillian C. McDermott, ''Student difficulties in understanding image formation by a plane mirror,'' Phys. Teach. **24**, 472–481 (1986); ''An investigation of student understanding of the real image formed by a converging lens or concave mirror,'' Am. J. Phys. **55**, 108–119 (1987); P. Colin and L. Viennot, ''Using two models in optics: Students' difficulties and suggestions for teaching,'' *ibid.* **69**, S36–S44 (2001); Philippe Colin, Françoise Chauvet, and Laurence Viennot, ''Reading images in optics: students' difficulties and teachers' views,'' Int. J. Sci. Educ. **24**, 313–332 (2002).

[34]Glenda Jacobs, ''Word usage misconceptions among first-year university physics students,'' Int. J. Sci. Educ. **11**, 395–399 (1989); P. Kenealy, ''A syntactic source of a common 'misconception' about acceleration,'' in *Proceedings of the Second International Seminar: Misconceptions and Educational Strategies in Science and Mathematics III* (Cornell Univ., Ithaca, NY, 1987), pp. 278–292; Jerold S. Touger, ''When words fail us,'' Phys. Teach. **29**, 90–95 (1991); H. Thomas Williams, ''Semantics in teaching introductory physics,'' Am. J. Phys. **67**, 670–680 (1999).

[35]Patricia F. Keig and Peter A. Rubba, ''Translations of representations of the structure of matter and its relationship to reasoning, gender, spatial reasoning, and specific prior knowledge,'' J. Res. Sci. Teach. **30**, 883–903 (1993).

[36]W. L. Yarroch, ''Student understanding of chemical equation balancing,'' J. Res. Sci. Teach. **22**, 449–459 (1985).

[37]Andrew Elby, ''What students' learning of representations tells us about constructivism,'' J. Math. Behav. **19**, 481–502 (2000).

[38]Jiajie Zhang, ''The nature of external representations in problem solving,'' Cogn. Sci. **21**, 179–217 (1997); Maarten W. van Someren, Peter Reimann, Henry P. A. Boshuizen, and Ton de Jong, editors, *Learning with Multiple Representations* (Pergamon, Amsterdam, 1998); Jeff Zacks and Barbara Tversky, ''Bars and lines: A study of graphic communication,'' Mem. Cognit. **27**, 1073–1079 (1999); Bruce L. Sherin, ''How students invent representations of motion: A genetic account,'' J. Math. Behav. **19**, 399–441 (2000); Andrea A. diSessa and Bruce L. Sherin, ''Meta-representation: An introduction,'' *ibid.* **19**, 385–398 (2000); Roser Pintó and Jaume Ametller, ''Students' difficulties in reading images. Comparing results from four national research groups,'' Int. J. Sci. Educ. **24**, 333–341 (2002); Tae-Sun Kim and Beom-Ki Kim, ''Secondary students' cognitive processes for line graphs and their components,'' in *Proceedings of the 2002 Physics Education Research Conference [Boise, Idaho, August 7–8, 2002]*, edited by Scott Franklin, Karen Cummings, and Jeffrey Marx (PERC, New York, 2002), pp. 91–94.

[39]A comparison of this type was made by Fernando Hitt, ''Difficulties in the articulation of different representations linked to the concept of function,'' J. Math. Behav. **17**, 123–134 (1998).

[40]Melissa Hayes Dancy, ''Investigating animations for assessment with an animated version of the Force Concept Inventory,'' Ph.D. dissertation, North Carolina State University (UMI, Ann Arbor, MI, 2000), UMI #9982749.

[41]David E. Meltzer, ''Comparative effectiveness of conceptual learning with various representational modes,'' AAPT Announcer **26(4)**, 46 (1996); ''Effectiveness of instruction on force and motion in an elementary physics course based on guided inquiry,'' *ibid.* **28(2)**, 125 (1998); Antti Savinainen and Jouni Viiri, ''A case study evaluating students' representational coherence of Newton's first and second laws,'' in *2003 Physics Education Research Conference [Madison, Wisconsin, August 6–7, 2003]*, edited by Jeffrey Marx, Scott Franklin, and Karen Cummings [AIP Conf. Proc. **720**, 77–80 (2004)].

[42]Robert B. Kozma, ''The use of multiple representations and the social construction of understanding in chemistry,'' in *Innovations in Science and Mathematics Education*, edited by Michael J. Jacobson and Robert B. Kozma (L. Erlbaum, Mahwah, NJ, 2000), pp. 11–46.

[43]Teresa Larkin-Hein, ''Learning styles in introductory physics: Enhancing student motivation, interest, & learning,'' in Proceedings of the International Conference on Engineering and Computer Education, São Paulo, Brazil (2000), ⟨http://nw08.american.edu/~tlarkin/larkin.htm⟩.

[44]Maria Kozhevnikov, Mary Hegarty, and Richard Mayer, ''Spatial abilities in problem solving in kinematics,'' in *Diagrammatic Representation and Reasoning*, edited by Michael Anderson, Bernd Meyer, and Patrick Olivier (Springer, London, 2002), pp. 155–171; Eun-Mi Yang, Thomas Andre, and Thomas J. Greenbowe, ''Spatial ability and the impact of visualization/animation on learning electrochemistry,'' Int. J. Sci. Educ. **25**, 329–349 (2003).

[45]A preliminary analysis of some of these data has been published previously. David E. Meltzer, ''Issues related to data analysis and quantitative methods in PER,'' in *Proceedings of the 2002 Physics Education Research Conference [Boise, Idaho, August 7-8, 2002]*, edited by Scott Franklin, Karen Cummings, and Jeffrey Marx (PERC, New York, 2002), pp. 21–24.

[46]The ''dominance principle'' (a term used by Halloun and Hestenes) refers to students' tendency to attribute larger-magnitude forces to one or the other object in an interacting pair, based on an ostensibly ''dominant'' property such as greater mass, velocity, or charge. See David P. Maloney, ''Rule-governed approaches to physics—Newton's third law,'' Phys. Educ. **19**, 37–42 (1984); Ibrahim Abou Halloun and David Hestenes, ''Common-sense concepts about motion,'' Am. J. Phys. **53**, 1056–1065 (1985); Lei Bao, Kirsten Hogg, and Dean Zollman, ''Model analysis of fine structures of student models: An example with Newton's third law,'' *ibid.* **70**, 766–778 (2002).

[47]This result suggests that some students' expertise in using vector representations may have increased faster than did their understanding of Newton's third law, because response B is an accurate representation of an answer based on the dominance principle.

[48]Question #2 in this set was designed by Leith Allen, private communication (2002).

[49]J. P. Guilford, *Fundamental Statistics in Psychology and Education*, 4th ed. (McGraw-Hill, New York, 1965), p. 184. This test considers each pair of values to be an independent measurement of the difference between the paired quantities. It is the appropriate test here because there are many

year-to-year variations (in student demographics, course logistics, etc.) but in each individual year, there is no *a priori* reason to expect differences between the paired quantities.

[50]Reference 49, p. 255.

[51]Reference 49, pp. 188–189.

[52]David J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures* 2nd ed. (Chapman & Hall/CRC, Boca Raton, 2000), p. 498.

[53]We have tried to further test this interpretation with interview data [Leith Allen and Larry Engelhardt, private communication (2003)]. Approximately 15 students were interviewed in all; they had volunteered in response to a general solicitation. None of the students interviewed showed any clear evidence of the representation-related difficulties identified in this paper. Our experience (and that of others) has been that most students who volunteer for interviews are well above the average in terms of course performance. It seems that the relatively simple nature of the questions used in this investigation (indicated by the low error rates) was an inadequate challenge for the interview volunteers. It will probably be necessary to target potential interviewees in the future, soliciting students who have already shown (on quizzes or exams) evidence of the learning difficulties being investigated.

[54]Lei Bao and Edward F. Redish, ''Concentration analysis: A quantitative assessment of student states,'' Am. J. Phys. **69**, S49–S53 (2001). Also see Ref. 45.

[55]Although the *V* and *D* versions of the gravitation question (and related Coulomb's law question) include similar options regarding force magnitudes, the *D* version obviously portrays directional information as well. This directional information is an additional bit of complexity which probably contributes to overall confusion, although it is not clear how (or whether) it might make it more difficult for a student to pick out an ''equal magnitudes'' option.

[56]This convention—that the tail of the arrow representing a force exerted on an object is attached to the object—is certainly not universal. However, in the context of question 8, the attractive nature of the gravitational force guarantees that the force exerted on an object must point toward the other object in the interacting pair. This fact makes the assignment of force vector arrows in question 8 unambiguous; regardless of the convention for locating the tails of the arrows, the arrow corresponding to the force exerted on the moon must point toward the earth. Therefore, it is not merely a confusion about notation or vector conventions that leads to the error identified here. [It is notable that not a single student chose either response *G* or *H* on the electrostatic final-exam question (Fig. 2); these responses would be acceptable representations of a dominance-principle answer, or the correct answer, respectively, if one ignored tail location.] This observation leaves open the question of whether the students' confusion was primarily with the tail location, the meaning of the arrow direction itself, the meaning of ''attractive force,'' or some amalgam of these (and possibly other) issues.

[57]Most gender-related differences in this study seem to be smaller than the differences documented to exist between traditional instruction and interactive-engagement instruction; see, for instance, Richard R. Hake, ''Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,'' Am. J. Phys. **66**, 64–74 (1998), ⟨http://www.physics.indiana.edu/~sdi/⟩. Marshall has recently reported on a study that suggests the existence of gender differences in interpretation of electric circuit diagrams: Jill Marshall, ''Gender differences in representations of electric circuits,'' AAPT Announcer **34(4)**, 96 (2004).

[58]However, one must also consider the possibility that specific differences in the way the questions were worded also may have contributed significantly to the discrepancies in responses that were observed.