

1-2007

Strongly and Weakly Directed Approaches to Teaching Multiple Representation Use in Physics

Patrick B. Kohl
University of Colorado at Boulder

David Rosengrant
Kennesaw State University, drosengr@kennesaw.edu

Noah D. Finkelstein
University of Colorado at Boulder

Follow this and additional works at: <http://digitalcommons.kennesaw.edu/facpubs>

 Part of the [Higher Education and Teaching Commons](#), and the [Science and Mathematics Education Commons](#)

Recommended Citation

Kohl PB, Rosengrant D, Finkelstein ND. 2007. Strongly and weakly directed approaches to teaching multiple representation use in physics. *Phys Rev Spec Top -Phys Educ R* 3(1):010108.

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Faculty Publications by an authorized administrator of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.

Strongly and weakly directed approaches to teaching multiple representation use in physicsPatrick B. Kohl,¹ David Rosengrant,² and Noah D. Finkelstein¹¹*University of Colorado at Boulder, Boulder, Colorado 80309, USA*²*Kennesaw State University, Kennesaw, Georgia 30144-5591, USA*

(Received 11 December 2006; published 12 June 2007; publisher error corrected 18 June 2007; corrected 27 July 2007)

Good use of multiple representations is considered key to learning physics, and so there is considerable motivation both to learn how students use multiple representations when solving problems and to learn how best to teach problem solving using multiple representations. In this study of two large-lecture algebra-based physics courses at the University of Colorado (CU) and Rutgers, the State University of New Jersey, we address both issues. Students in each of the two courses solved five common electrostatics problems of varying difficulty, and we examine their solutions to clarify the relationship between multiple representation use and performance on problems involving free-body diagrams. We also compare our data across the courses, since the two physics-education-research-based courses take substantially different approaches to teaching the use of multiple representations. The course at Rutgers takes a strongly directed approach, emphasizing specific heuristics and problem-solving strategies. The course at CU takes a weakly directed approach, modeling good problem solving without teaching a specific strategy. We find that, in both courses, students make extensive use of multiple representations, and that this use (when both complete and correct) is associated with significantly increased performance. Some minor differences in representation use exist, and are consistent with the types of instruction given. Most significant are the strong and broad similarities in the results, suggesting that either instructional approach or a combination thereof can be useful for helping students learn to use multiple representations for problem solving and concept development.

DOI: [10.1103/PhysRevSTPER.3.010108](https://doi.org/10.1103/PhysRevSTPER.3.010108)

PACS number(s): 01.40.Fk

I. INTRODUCTION

Instructors and researchers in physics education research (PER) have long argued that students can benefit from solving problems that require the use of multiple representations together.¹⁻⁶ The distinction between multiple representation problems and other problems is somewhat artificial, as it is difficult to imagine solving any physics problem or making sense of any physics idea without making use of more than one representation (in thought, if not on paper). Nevertheless, the class of problems referred to as multiple representation problems, and the use of multiple representations in general, is often understood to mean the *explicit* use of more than one representation in solving problems. These kinds of problems are said to require a more complete understanding of the underlying physics than traditional “plug and chug” problems in which only mathematics is explicitly present.^{1,2}

Experts and novices differ significantly in their use of multiple representations. Experts tend to use multiple representations in their problem setups more often than novices, who have a tendency to jump directly to mathematics.^{1,7} Thus, use of multiple representations brings student problem-solving procedures more in line with expert procedures. These differences extend beyond problem solving, as research has shown that novices and professional scientists differ significantly in their ability and willingness to use multiple representations productively in more applied settings such as the laboratory or workplace.^{8,9} It has even been suggested that competency with several representations of a concept is a prerequisite for expertlike understanding,¹⁰ and popular research-based physics assessments have implicitly acknowledged this point by including a spread of representations in their questions and by requiring translations among representations to solve problems.¹¹⁻¹³

Previous work has established that students in traditional physics courses only sometimes use multiple representations,³ and that efforts specifically focused on increasing student use of multiple representations can be successful, even if students are not graded specifically for multiple representation use.^{4,5,14,15} In addition, it has been suggested that this multiple representation use can be associated with increased problem-solving performance, though this correlation is far from perfect.¹⁶ To address this, Rosengrant *et al.* have considered the correlation between the *quality* of multiple representation use and student success.¹⁵ They find that this association is quite strong, a point we return to in the present paper.

The courses in the above studies can be described as taking strongly directed approaches to teaching problem solving with multiple representations. By “strongly directed,” we mean that these approaches teach explicit steps and heuristics for solving multiple representation physics problems and continue to emphasize these steps throughout the course. Another, less studied (but perhaps common) approach is to model good multiple representation problem-solving techniques for students without teaching specific steps. We can refer to this approach as “weakly directed.” Arguments can be made in favor of either the strongly or weakly directed approaches. For example, a strongly directed approach gives students an easy-to-follow checklist, though it might also result in dependence on algorithms executed with little understanding. A weakly directed approach may prevent dependence on checklists, but novice students may be incapable of picking up the appropriate skills in the course of an introductory class without such direction. We are unaware of any studies directly comparing strongly and weakly directed approaches to teaching multiple representation problem solving. In this paper, we perform such a comparison.

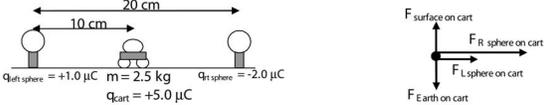
We address these questions regarding the use and learning of representations in two parts. In the first part, we verify that multiple representations aid problem solving, and ask whether we can begin to understand more specifically how multiple representation use is associated with student performance. In the second part, we ask how multiple representation use and success with multiple representation problems varies with instruction, and examine two multiple-representation-rich, PER-based courses that take different approaches to teaching multiple representations: one strongly and one weakly directed. To this end, we study student performance on five multiple representation problems in two introductory large-lecture algebra-based physics courses, one taught at Rutgers, the State University of New Jersey, and one taught at the University of Colorado at Boulder (CU). The problems vary in their difficulty and in their framing. For example, one problem hints that a force diagram might be useful, while another makes no such hint. Four of the problems were given in recitation, and a more difficult “challenge problem” was given as a recitation quiz at CU and as part of an exam at Rutgers.

By examining student solutions and performance in detail, we begin to address our first questions. As we have noted, many studies have established that using multiple representations can improve performance. We find, perhaps not surprisingly, that student use of multiple representations does indeed often correlate with success. However, we find that the correlation is nontrivial. Use of multiple representations alone is insufficient for success and can even be associated with lower-than-average performance. Correct use of multiple representations and close coordination of those representations is much more likely to be associated with high success rates. We also find that problem framing can alter student use of multiple representations; for instance, student solutions to problems might show different uses of free-body diagrams (FBDs) depending on whether the problem used the word “force” or not.¹⁷ Notably, this last result regarding framing is tentative: The data here are insufficient to fully characterize the effect; rather the results serve to suggest that multiple representation use in problem solving may be susceptible to the same kinds of cueing observed in other contexts.^{18–20}

Given these data, the second part of the paper focuses on a cross-course comparison, and on the question of whether one approach or the other is optimal. Most significant was the overall constancy of the results from the first part of our study across both environments. Both courses were successful in promoting multiple representation use, and student performances were very similar. We note some specific differences that emerged, though we emphasize that the major picture was one of strong similarity. The CU students were slightly more likely to use multiple representations on shorter, easier problems, while Rutgers students were more likely to use complete FBDs on the most difficult problem. These differences and others can be plausibly attributed to the differences in instructional environment, and lead us to suppose that elements of each course might be reasonably combined in the future.

To summarize, we ask two primary questions, with the associated findings following:

1. A small (100 g) metal ball with $+2.0 \mu\text{C}$ of charge is sitting on a flat frictionless surface. A second identical ball with $-1.0 \mu\text{C}$ of charge is 3.0 cm to the left of the first ball. What are the magnitudes and directions of the forces that we would have to apply to each ball to keep them 3.0 cm apart?
2. A sphere of 0.3 kg is charged to $+30 \mu\text{C}$. It is tied to a second chargeable sphere by a 20 cm rope, and the spheres sit on a frictionless table. If the rope will break at 4.8 N, what charge needs to be on the second sphere to cause the rope to break?
Hint: It may be useful to draw a force diagram.
3. A frictionless metal cart is being held halfway between two stationary charged spheres. The cart's mass is 2.5 kg and its charge is $+5.0 \mu\text{C}$. The left sphere has a charge of $+1 \mu\text{C}$ and the right sphere has a $-2 \mu\text{C}$ charge. The two spheres are 20 cm apart. At the instant the cart is released, what is the magnitude and direction of the total force on the cart? Refer to the included diagrams for help.
4. A 100 gram ball has a charge of $+40 \mu\text{C}$. The ball is dropped from a height of 2 m into a 7000 N/C electric field pointing up. Draw a diagram showing all the forces involved in the problem, and calculate the magnitude and direction of the net force on the ball.



The diagram for problem 3 consists of two parts. On the left, a top-down view shows a central cart of mass 2.5 kg and charge $+5.0 \mu\text{C}$ positioned halfway between two stationary spheres. The left sphere has a charge of $+1.0 \mu\text{C}$ and the right sphere has a charge of $-2.0 \mu\text{C}$. The distance between the two spheres is 20 cm, and the cart is 10 cm from each. On the right, a free-body diagram for the cart shows four forces: $F_{\text{surface on cart}}$ pointing up, $F_{\text{Earth on cart}}$ pointing down, $F_{\text{L sphere on cart}}$ pointing left, and $F_{\text{R sphere on cart}}$ pointing right.

FIG. 1. Four problems given in recitation. Note the different prompts regarding multiple representation use [(1) no prompt, (2) hint to draw a force diagram, (3) diagrams included, and (4) statement that diagrams are required].

(1) When and how does the use of multiple representations affect student performance on problems involving free-body diagrams? Here, we find a correlation between FBD use and success, but it is not strict: Poor use of multiple representations is no better and possibly worse than no use thereof.

(2) What sorts of instructional methods best foster multiple representation use? We compare two PER-based and representationally rich approaches that differ significantly in their details, with both yielding very high rates of picture and FBD use among introductory students. Neither is clearly superior by our measures.

II. METHODS AND COURSE DESCRIPTIONS

A. Methods: Study problems

In each course, students received a set of four electrostatics problems in recitation that either required calculation of a force or specified forces in the problems. These problems were given after all lecture coverage of electrostatics and students received recitation credit for significant effort. The problems did not otherwise count toward the course grade. All problems are shown in Fig. 1. The problems contained a variety of cues regarding the use of multiple representations. The first problem made no mention of multiple representations. The second problem hinted that it may be useful to draw a force diagram. The third problem included both a picture and a FBD as part of the statement. The fourth problem stated that a FBD was required as part of the solution.

Students were also given a more challenging problem, intended to be very difficult to solve without a FBD. This problem was issued with multiple-choice answers on the first exam in the Rutgers course, and as a free-response quiz in recitation just before the first exam in the CU course. This problem and an example solution are shown in Fig. 2. We shall refer to this problem as the challenge problem.

Student solutions to these problems were coded in several ways in a scheme that extends methods developed

A small metal ball with $Q_1 = +2.0 \mu\text{C}$ of charge hangs at the end of a vertical string. A second, identical ball with $Q_2 = -2 \mu\text{C}$ of charge hangs at the end of a vertical string. The tops of the strings are brought near each other, and the strings reach an equilibrium orientation (no longer vertical) when the balls are a distance $d = 3.0 \text{ cm}$ apart. If the force of the Earth on each ball is $F_1 = F_2 = 30 \text{ N}$, what is the force T of the string on each ball? Write your symbolic answer in terms of $F_1, F_2, Q_1, Q_2,$ and appropriate electrical constants.

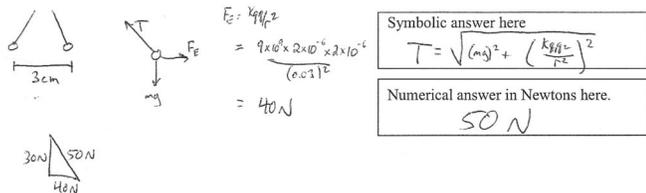


FIG. 2. Challenge problem with example solution. The picture drawn shows the common backwards picture error, as the balls are supposed to attract each other, leading to inward-facing strings.

previously.²¹ The answers were coded as correct or incorrect. Specific answer features were also tracked; for example, if the answer required a number and a direction, each feature was coded separately. Student use of representations was coded using a more complex scheme. Each solution was coded with respect to any picture used and any free-body diagram used. For problems 1, 2, and 4 in the recitation set, the problem was coded as either containing a picture or not containing a picture. A picture was defined as some drawing representing the situation, not to include an isolated free-body diagram (coded separately). The expected elements of each picture were then coded as present or absent. For example, in problem 1 (Fig. 1) the coders looked for the presence of each of the two charges and for a labeling of the distance between them. The pictures for the challenge problem were coded in more detail: The presence and correctness of the picture was evaluated using a 0–3 rubric, where 3 meant a correct depiction of the physical situation, 2 referred to a common error in which the picture was drawn “backward,” 1 referred to an otherwise incorrect picture, and 0 indicated no picture. The expected elements of the picture were then coded as present or absent, as before in the recitation problem. For problem 3 in the recitation section, a picture was provided, so coders noted whether students made their own marks on the given picture, and whether they re-drew a picture of their own.

Free-body diagrams were coded in a similar fashion. For each problem, as many as two to four forces could reasonably be present. For each possible force (a gravitational force, a normal force, etc.), the force was flagged as being present or not, being shown in the correct direction or not (ambiguities were also flagged), and being labeled correctly or not. Coding each element of the FBDs and pictures separately facilitated analysis, as most combinations of features of interest were available in the codings.

An author from Rutgers University coded all of the data from that institution, and an author from CU coded all the data from CU. These authors then both coded two sections of the data chosen at random and compared codings. Agreement varied from 91% to 100%, depending on the category.

B. Course descriptions

The study involved second-semester large-lecture algebra-based physics courses from CU and Rutgers, taught in the

spring of 2006. The instructors for these courses had also taught the first semester of the sequence, and have been involved in PER for many years. Both courses can be described as reformed in nature, making use of many common tools and practices from PER. The courses each had one recitation or lab meeting per week, with two or three full class meetings. Each course was four credit hours. Lecture sections had approximately 300 students each. Each school is a large state university, with similar standardized test scores for incoming students. As these were life-science track courses, the backgrounds and performances of the students within each of the classes varied considerably.

1. Rutgers University

The Rutgers course uses the ISLE curriculum, which is inquiry-based and spends considerable time on the use of multiple representations.²² The instructors use the *Active Learning Guide* workbook in lecture and in recitation, which includes many tasks designed to teach multiple representation use.²³ The recitations have research-based design elements,²⁴ and also use ACTIVPHYSICS computer simulations, which emphasize conceptual development, problem solving, and multiple representations.²⁵ The lectures also use personal response systems (clickers).

For both mechanics and electrostatics problems in the ISLE curriculum, the instructor teaches students an explicit problem-solving heuristic with five main steps, which emphasizes multiple representations and is described elsewhere.²³ Note that this five-step procedure includes within it a subprocedure for drawing free-body diagrams. These procedures are emphasized whenever multiple representation problems are discussed, though rigid adherence to each step of the procedure is not required, and students were never graded specifically on following the steps.

2. University of Colorado

The CU course features such reforms such as clickers and PER-based labs and recitation activities,²⁶ and includes the PhET computer simulations.²⁷ It also includes substantial multiple representation use in lecture and in homework and exam tasks, but little explicit instruction in multiple representation use is given. The instructor taught no specific problem-solving heuristics. In Fig. 3, we see an example of an exam question from the CU course. Such multiple representation questions were common. With substantial multiple representation use in lecture and on exams, students were held accountable for using multiple representations effectively as well as having such use modeled for them.²⁸

C. Environment evaluation

The multiple-representation based reforms present in the Rutgers lectures are well documented.^{22,23} To establish the representational richness of the CU environment, we analyzed the representational content of their lectures using a procedure developed and validated previously.²⁸ To help compare the two course environments, we performed a similar analysis on the exams from each course. We will only summarize the procedure here. To characterize the lecture

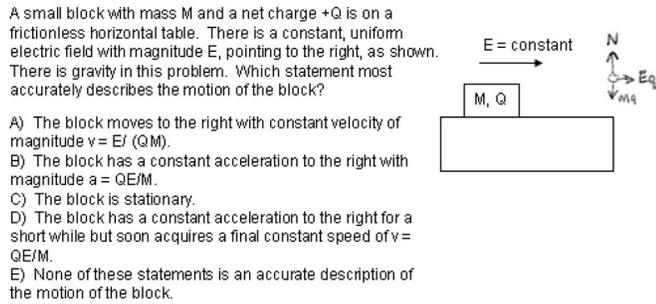


FIG. 3. Multiple representation problem from a University of Colorado exam. The free-body diagram is not part of the problem statement.

content, we take a series of videotaped lectures and break them into one minute intervals. We code each minute according to whether it includes use of verbal, mathematical, graphical, or pictorial representations, with “verbal” including written physics principles, but not spoken language (since spoken language is almost always present). Any interval that has more than one representation is also coded as having “multiple representations.” We then average over all lectures to come up with an average fraction of lecture time spent on each category. We videotaped eight CU lectures in between the first day of class and the first exam (the lectures covering the material used in this study).

For the exam content, we focus on all the exams that lead up to the study material, as these would be the only ones likely to influence student behavior in the study. This means we consider all the exams from the first semester of the course, keeping in mind that all student data presented come from students who took both semesters consecutively. We quantified the fraction of each exam that could be described as verbal, mathematical, graphical, and pictorial in representation on a problem-by-problem basis using a previously developed standard.²⁸ We also quantified the fraction of each exam that explicitly required the use of multiple representations. No effort was made to weight the different representations in a multiple representation problem; for instance, we did not designate a problem as 80% pictorial and 20% mathematical, but rather as 100% of each, and so the sums across all representations included in the data can exceed 100%. Once we characterized each exam in terms of its representational content, we calculated the average representational content of the exams in each of the courses.

III. DATA AND ANALYSIS

We present the data and analysis in three parts. First, we compare the representational content of the courses studied. Second, we examine student performance and representation use on a problem-by-problem basis, comparing across courses when appropriate. Finally, we focus more closely on cross-course analysis.

A. Part I: Environment data

We have claimed that both the CU and Rutgers courses are representation rich, noting the various curriculum re-

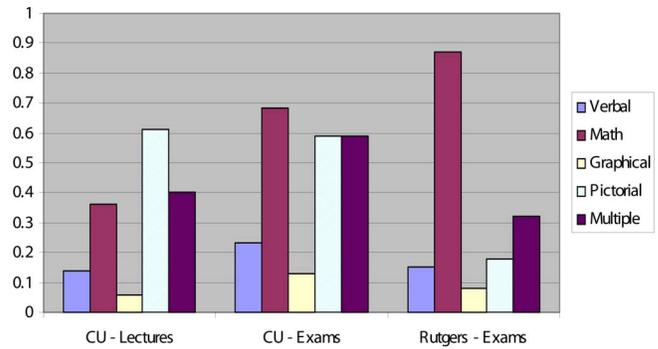


FIG. 4. (Color) Fraction of lectures and exams at CU and exams at Rutgers using verbal, math, graphical, pictorial, and multiple representations.

forms present in each. In particular, we saw that the Rutgers course made use of specific curricula intended to promote the use of multiple representations in lecture. Since CU used no such documented curricula, we present data on the representational richness of CU’s lectures here. For the sake of cross-course comparisons, we also analyze the representational content of the exams in each class.

In Fig. 4, we see the fraction of the sampled lectures that contained verbal, mathematical, graphical, and pictorial representations. The CU data show more representations being used more often than in a similar, traditionally taught class studied previously,²⁸ supporting the claim that this environment is representationally rich, much like the Rutgers lecture environment. Such richness is consistent with the strong, broad similarities observed in representation use among CU and Rutgers students.

In Fig. 4 we also see the fraction of the first-semester exams (those leading up to this study) using each of these four representations and using multiple representations. In these, we see a difference between the courses. The CU exams tended to use more representations more often, and used multiple representations more often, while the Rutgers exams focused more on mathematical representations.³⁰

B. Part II: Performance and representation use

1. Recitation problems

In Table I we see the fraction of the students in each course answering each of the four recitation problems correctly. The numbers in parentheses indicate the number of students sampled for each problem. Problems 1, 2, and 3 had the same sample size.³¹ Problems 1 and 2 are similar in that both require single applications of Coulomb’s law, but with different variables to be solved for (force in problem 1, and charge in problem 2). Students in both courses performed significantly worse on 1 than on 2, with an average fraction correct across courses of 0.37 for problem 1 and 0.53 for problem 2. These differ at a $p < 0.0001$ level using a two-tailed binomial proportion test, but this is likely a result of the extra information requested by problem 1. Problem 1 asks students to note the direction of the force calculated, and examination of student solutions shows that many stu-

TABLE I. Fraction of students answering the four recitation problems correctly at Rutgers and CU. Parentheses indicate sample sizes. Samples for problems 1, 2, and 3 are the same. Standard errors vary but are on the order of 0.03. The 1 (scalar) category refers to the scalar portion of the answer for problem 1.

	Problem 1	1 (scalar)	Problem 2	Problem 3	Problem 4
Rutgers	0.36 (296)	0.51	0.49	0.29	0.38 (155)
CU	0.38 (314)	0.60	0.56	0.43	0.40 (269)

dents simply overlooked or ignored this directive. Thus, we also include in Table I the fraction of students answering the scalar portion of problem 1 correctly, which does not differ significantly from the fraction answering problem 2 correctly. We consider problems 3 and 4 to be less directly comparable to the others since their solutions were substantially different, as were their treatments of multiple representations in the setup.

In Table II we see the fraction of students in each course that drew a picture with their problem solutions. Since problem 3 provided a picture, we instead show the fraction of students in each course that redrew their own picture. Students were quite likely to draw pictures in all cases, with 90% or more of students drawing a picture in four of six cases (not counting problem 3). Students were equally likely to draw pictures for problems 1 and 2. Table II also shows the fraction of students identifying any forces correctly in their solution, using some kind of vector representation. Since problem 3 provided an FBD, we show the fraction of students who redrew some force information on their own. Two data features are notable: First, the vast majority of students drew a complete and correct FBD for problem 4 (almost all who identified at least one force identified both possible forces). Since this problem asked students to draw an FBD as part of their answer and since it was the last problem in the set, we consider this an indication that students were taking the problems seriously throughout the set. Second, students were much more likely to draw some kind of FBD for problem 1 than for problem 2. Forty-five percent of students identified some forces correctly for problem 1, compared to 29% for problem 2 ($p < 0.0001$). Mathematically, these problems were very similar, and it is possible that this difference resulted from some difference in the problem framing, a point we will return to in the discussion.

2. Challenge problem

In the first column of Table III, we see the fraction of students answering the challenge problem correctly in each course. Because the Rutgers problem was given as a five-answer multiple-choice question and the CU problem was given as a free-response question, we do not consider the

difference in performance between Rutgers and CU to be significant or useful for further analysis. In the next three columns, we see the fraction of students identifying exactly one, two, or three forces correctly in their solution. Note again that an FBD was not requested by the problem. More than 98% of students drew a picture. The last two columns show what we refer to as type 2 and type 3 picture use (from the previously described picture rubric). Picture type 3 is complete and correct. Picture type 2 was a common misinterpretation of the problem statement, where students drew the balls as if they were repelling (an example is shown in Fig. 2). Students drawing picture type 1 were otherwise incorrect, are not shown in the table, and will not be considered further.

Once again, we note the very frequent use of multiple representations, with nearly all students in either course drawing a picture, and 83% of students identifying at least one force correctly, despite no request for a picture or FBD in the problem.

3. Relation of performance to representation use

We next consider student performance as a function of multiple representation use. That is, we ask whether students that used pictures and FBDs performed better. We cannot compare problem-by-problem performance between picture-drawing students and non-picture-drawing students since nearly all drew a picture. Instead, we begin by examining student success as a function of correct FBD use. Previous work has shown that students who construct a correct FBD to help them solve problems do significantly better than students who do not construct diagrams or who construct incorrect diagrams.¹⁵ In Fig. 5, we see the success rate for students correctly identifying 0, 1, 2, or 3+ (3 or 4) forces per problem on the challenge problem. Since the CU and Rutgers problems differed in format (CU being free response and Rutgers multiple choice), we have normalized the data to reflect this. Each CU data point has been renormalized by a constant factor so that the CU and Rutgers overall mean scores are identical, allowing for easier trend comparison.³² This scaling does not change the shape of the curve observably. Overlap is very thorough. Student performance drops

TABLE II. Fraction of students drawing a picture for each of the four recitation problems, and fraction of students identifying any forces correctly in their solution. Standard errors vary but are on the order of 0.03.

Pictures	Problem 1	Problem 2	Problem 3	Problem 4	Forces	Problem 1	Problem 2	Problem 3	Problem 4
Rutgers	0.90	0.85	0.00	0.73	Rutgers	0.36	0.26	0.03	0.90
CU	0.92	0.91	0.13	0.90	CU	0.53	0.32	0.11	0.92

TABLE III. Fraction of students answering the challenge problem correctly, fraction correctly identifying one two, or three of the possible forces, and fraction drawing either picture type 2 or type 3. Parentheses indicate sample size.

Challenge	Correct	One force	Two forces	Three forces	Picture 2	Picture 3
Rutgers	0.56 ^a (283)	0.09	0.22	0.51	0.34	0.41
CU	0.29 ^a (280)	0.23	0.31	0.32	0.39	0.45

^aNote that the Rutgers version was multiple choice, while the CU quiz was free response.

from zero to one force identified, and increases to two and finally to three. Uncertainties are relatively large for zero, one, and two forces, but we consider the fact that both schools' curves overlap so closely as to make the observed trend more likely to be real. Averaging the CU and Rutgers data sets results in error bars of approximately half the size (not shown). We can perform a similar analysis for problems 1 and 2 but the trends are much less clear. For those data (not shown), we cannot conclusively claim that more correct use of multiple representations leads to higher performance (and neither can we claim that it does not): the trend is more or less flat. We note here that this analysis would be inappropriate for problems 3 and 4, as problem 3 provides students with a complete FBD already, and problem 4 tells students explicitly to draw a FBD as part of their answer. In the challenge problem and in problems 1 and 2, a free-body diagram is potentially useful but is neither provided nor required.

The above data confirm previous results¹⁵ where FBD use is strongly associated with student success on this problem. This suggests to us a finer-grained analysis here. Is success associated with any more specific pattern of representation use? In Table IV, we show student performance versus the identification of each force, the correct representation of each force, and the correct, labeled representation of each force present in the exam problem. Rutgers and CU data are very similar, so we display only CU data. Student performance is flat along the vertical dimension (which would show a dependence on correctness or labeling), and mostly flat along the horizontal dimension (which would show a dependence on force type). There is a minor excess in the second column, corresponding to the electrostatic force. Notably, this is the force whose correctness can be most easily impacted by drawing a type 2 versus a type 3 picture, so this excess might be more reflective of picture type than anything else. Generally, the weak dependence on any one factor suggests that only correct coordination across all of the forces will be associated with success.

C. Part III: Cross-class comparison

The above data are interesting when viewed from a problem-by-problem perspective. We see that multiple representation use can significantly influence success, especially on more difficult problems, and that complete, correct multiple representation use is associated with high performance. We also note that the data, when viewed from a cross-course perspective, show similarities and differences. Performances overall are quite similar, as is the dependence of performance on representation use. In this section, we investigate those

differences and similarities in more detail, to work toward an understanding of their source.

1. Cross-class performance

First, we compare Rutgers and CU performances on problems 1–4 (Table I). We can compare the performances pairwise, but since this analysis is *post hoc* we must modify the *p* value considered significant (or use an appropriate *post hoc* test). A simple and very conservative approach is to choose the *p* value such that if one were to make *N* *post hoc* comparisons, a difference on any one comparison could be considered significant. We thus choose $p = 1 - (1 - \alpha)^{1/N}$, where α is the desired significance level (0.05) and *N* is the number of *post hoc* comparisons (4).²⁹ This yields $p = 0.013$. The CU and Rutgers performances on problem 3 differ at a $p = 0.0002$ level using a two-tailed binomial proportion test, but no other pair differs significantly. Averaged across all recitation problems, the two courses do not differ significantly in problem performance.

2. Cross-class representation use

Perhaps the most noticeable result is the very large fraction of both course types that made use of pictures and free-body diagrams, despite the significant differences in instruction. Student performance is also very constant across courses, as the performances for problems 1, 2, and 4 are statistically indistinguishable, with the challenge problem performances also similar after accounting for the format differences. Thus, we have a significant performance difference on only one of the five problems studied. However, some differences emerge in representation use. On the recitation problems that neither demand nor provide a free-body diagram (problems 1 and 2), the CU students identify at least one force correctly significantly more often (43% vs 31%, $p = 0.002$, Table II). In contrast, on the exam problem (where the vast majority of students in both courses draw some forces), the Rutgers students are significantly more likely than the CU students to identify all three forces correctly, generating a complete and correct FBD (51% vs 32%, $p < 0.0001$, Table III). Picture use is comparable on problems 1 and 2, but on problem 4 CU students were more likely to draw a picture (90% vs 73%, $p < 0.0001$).

3. Cross-class performance versus representation use

As noted, the dependence of performance on representation use is similar in both classes. The trends of correctness vs FBD use in Fig. 5 are nearly identical, and in neither class

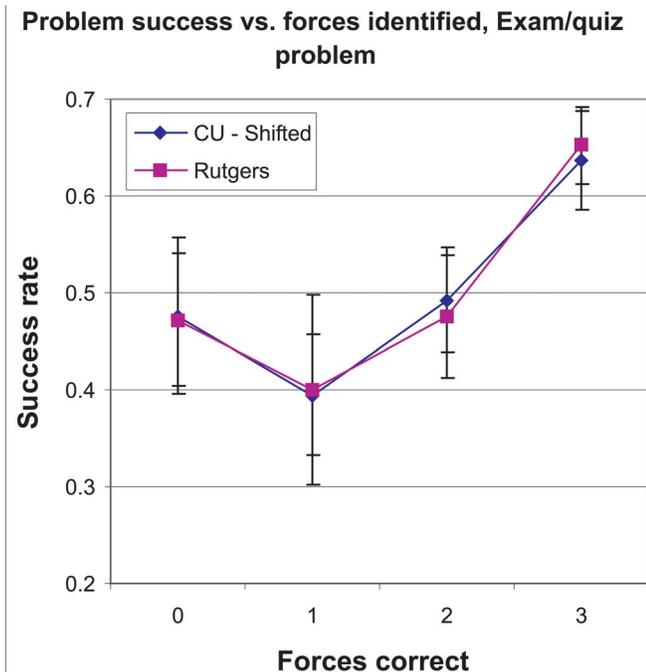


FIG. 5. (Color) Challenge question performance as a function of number of forces identified correctly. Note that CU scores have been shifted to account for free response vs multiple choice difference.

does the performance difference for the challenge problem depend on which specific force was identified (as opposed to how many). Since the data suggest that complete and coordinated use of multiple representations is most relevant, we can continue along these lines by breaking down student challenge problem performance by both FBD use and picture use. While nearly all students drew a picture, not all students drew the same picture. In Table V, we show student performance as a function of picture type drawn (2 or 3) and as a function of the number of forces correctly identified (two or three). We note that, for Rutgers students, the performance difference between using two forces and three forces was minimal, while the difference between using picture 2 and picture 3 was large. Conversely, for CU students the difference between using two forces and three forces was large, while the difference associated with the picture types was small.

TABLE IV. Fraction of students answering the challenge problem correctly, shown for each force available in the problem and broken down according to whether that force was present, drawn correctly, or drawn correctly and labeled. Data for CU and Rutgers are very similar, so we display only those for CU.

	Forces		
	mg	F_e	Tension
Present	0.30 (243)	0.38 (186)	0.32 (222)
Correct	0.30 (243)	0.38 (114)	0.36 (198)
Correct and labeled	0.31 (240)	0.39 (108)	0.34 (191)

TABLE V. Fraction of students answering the challenge problem correctly, broken down by whether they drew picture type 2 or 3 and whether they identified two or three forces correctly.

	Rutgers		CU	
	Two forces	Three forces	Two forces	Three forces
Picture 2	0.44 (39)	0.46 (37)	0.28 (47)	0.38 (21)
Picture 3	0.75 (8)	0.71 (87)	0.22 (27)	0.42 (47)

From the above, we see some specific differences between the two classes: the CU students appear to be more likely to use multiple representations on the simpler problems (specifically, problems 1, 2, and 4), while the Rutgers students are more successful with FBDs on the more difficult challenge problem. Furthermore, the correctness of the picture seems to be a more significant factor for Rutgers students, while correctness of the FBD appears to be the most significant factor for CU students.

IV. DISCUSSION

Our first goal was to ask whether multiple representation use mattered, and if so, how. The challenge problem data confirm what has been observed previously: Students that use free-body diagrams correctly significantly outperform those who do not.^{15,33} However, the trends were less clear for the recitation data, especially for problems 1 and 2. There, the quality of a student’s FBD is not clearly associated with their success, which may be due in part to the relative simplicity of these problems. A student with a good grasp of the material could reasonably solve both of these problems in a “plug ’n chug” fashion, without any additional representations, leading to a less straightforward dependence.

The challenge problem was more difficult, and perhaps benefits more from the use of a picture and free-body diagram. This is consistent with the fact that many more students used both pictures and FBDs for the challenge problem than for the recitation problems, and with the fact that the dependence of performance on representation use was much clearer for the challenge problem. Thus, these data suggest the somewhat intuitive result that for difficult problems, multiple representations can be especially helpful. There are no guarantees, of course: From Fig. 5, we can see that students who drew a FBD that was only partially correct were no more likely to answer the problem correctly than those who drew no FBD at all. Indeed, it is possible (though not conclusive from these data) that the students who drew only one force correctly did worse than those who drew none, which is consistent with earlier work,¹⁵ and is reasonable if we assume that the “no forces” group includes some students who are extremely comfortable with the material and skip diagrams, or keep track of information with mental representations rather than external representations. Along these lines, we note in Table IV that no one type of force in student solutions was a driving factor in student success. Only the successful coordination of all three forces was associated with better-than-average performance.

On a problem-by-problem basis, we note that problems 1 and 2 are very similar in their solution. Each requires a single use of Coulomb's law with one variable missing. From an expert perspective, it is not obvious that one is more difficult than the other, or that one is more likely to benefit from a picture and/or FBD than the other. Students did, in fact, perform very similarly on the scalar-only parts of problems 1 and 2. Yet many more students drew an FBD for problem 1 than for 2 (45% vs 29%). This difference occurs despite the fact that problem 2 hinted directly for students to use a force diagram, and problem 1 did not. We speculate that this variation might have been a result of the problem framing. Problem 1 asked students to calculate a *force* using Coulomb's law, perhaps suggesting a FBD, while problem 2 asked students to calculate a *charge*. This is potentially significant if true: A change in framing (in this case, a language cue) had a significant effect on multiple representation use where an explicit statement did not. In future work, we will vary the problems slightly to look for influences on representation use. For instance, problem 2 could be changed to provide charge and request force magnitude. If the framing is in fact responsible for the difference in representation use between problems 1 and 2, we should expect this change to problem 2 to result in increased use of FBDs.

Our second goal was to compare the two courses directly. Both courses were representationally rich, but with a significant difference. The Rutgers course strongly directed student use of multiple representations, providing specific problem-solving procedures that were emphasized throughout the course. The CU course was representationally rich, presenting a variety of representations in lecture and recitation, and on exams, but did not teach specific procedures. Despite this difference, both courses were very successful in fostering multiple representation use. On all five problems in both recitation and exam environments, students were extremely likely (typically >90%) to use supplementary representations like pictures and FBDs. For comparison, Van Heuvelen observes much less frequent multiple representation use in traditional courses.³ Performance was also quite similar across the courses, and only problem 3 showed a significant difference. The main feature distinguishing problem 3 from the others was the fact that a picture and FBD were included with the problem statement, so students were less dependent on their own supplementary representations.

While we consider the major result to be the strong similarities between the results for both courses, some aspects of the data did differ from course to course. The CU students were more likely to use multiple representations on the shorter, easier recitation problems (particularly problems 1 and 2). The Rutgers students were more likely to use complete and correct FBDs on the challenge problem. This suggests a possible explanation. Since the Rutgers students are being taught (but not graded on) a multistep problem-solving process using multiple representations, they may be less willing to engage in that process in the easier, lower-stakes recitation problems, and more willing to engage in that process for a high-stakes exam problem. In comparison, the CU students have learned to use multiple representations, but without specific procedures or guidelines for their use. This could result in more willingness to use them on lower-stakes prob-

lems, and in relatively less success with them on higher-stakes problems (though those that do succeed in using multiple representations appear to succeed similarly in solving the problem). If this is the case, we might expect there to be less performance dependence on representation use for Rutgers students on high-stakes problems: most students in that case would be using the problem-solving procedure, and the ability and willingness to draw a complete FBD might be less of a discriminator than it would be for CU students. In one sense, we do not observe this. Student performance as a function of FBD correctness is nearly identical in both courses. In another sense, we do see this. Picture correctness is a much more powerful discriminator for Rutgers students than for CU students, which is consistent with the notion that most students, strong or weak, are drawing fairly good FBDs (73% identifying two or more forces correctly), so that some other factor could present itself as a strong discriminator. Either way, these differences in representation use should not detract from the striking broad similarities observed in the data from the two courses.

There is another possible contributor to the surprising observation that CU students solved problems using multiple representations as often as the Rutgers students, whose course appears more likely upon initial inspection to promote multiple representation use. In previous work, we have suggested that facility with multiple representations might best be promoted by infusing all aspects of a course with multiple representation use.²⁸ Here, we saw that the Rutgers and CU environments differed in another way. The CU exams were richer in representations and in multiple representations use, whereas the Rutgers exams were more focused on mathematical representations. The fact that the CU exams were more likely to hold students accountable for being able to interpret a variety of representations might have offset some of the effect of the more detailed Rutgers multiple representations curriculum. Note that the data do not demonstrate this effect clearly; we mention it only as a possible confounding factor. Nor is this a value judgment: The extent to which course exams should focus on nonmathematical representations is dependent on course goals and upon which aspects of the course are meant to promote which goals.

We close our discussion with a summary of the limitations of the study. First, the student populations are different. While both universities were large, fairly selective state universities with comparable SAT and ACT scores, it is quite possible that some systematic differences existed, though none could be clearly identified and correlated with the observed differences in the results. Second, the challenge problem was given in two different formats: multiple choice on an exam for Rutgers, and free response on a recitation quiz for CU. This makes performance comparisons difficult, and introduces an environmental difference between the two groups (quiz vs exam). Third, course language was not necessarily identical. For example, the CU instructor sometimes used the term "force diagram" rather than "free-body diagram." Problem 2 in our study used the phrase "force diagram," though Rutgers students were not exposed to that specific term. Finally, we acknowledge that this study includes only one challenge-style problem, and only one problem in each of the four categories of framing that we define

(hint to use a FBD, FBD required, etc.). Nonetheless, we consider these five problems administered to several hundred students to be adequate to establish the broad similarities between the courses studied, and to establish the existence of the other effects noted.

V. CONCLUSIONS

We can draw two main conclusions from our results, each addressing one of our two primary research questions. First, we confirm that multiple representation use is important in successful physics problem solving as seen in previous work, but find that the dependence is not trivial. Coordinated and correct use of multiple representations on challenging problems can be very helpful, but multiple representation use on simple problems, or poor use of multiple representations, might not have a positive impact on student success. This dependence of performance on representation use was very similar across two different courses. Second, we find that multiple representation use can be taught, and in more than one way. One of the physics courses studied took a strongly directed approach to teaching physics problem solving with multiple representations, while the other took a weakly directed approach. Both courses were very successful in promoting multiple representation use across a variety of problems, and student performances were generally comparable. Notably, both courses were heavily PER influenced. We observed some minor differences between the two courses. The CU students were more likely to use multiple representations

on some of the easier problems, while the Rutgers students were more likely to use multiple representations correctly on the more difficult and higher-stakes challenge problem. This is consistent with the idea that Rutgers students are learning approaches to using multiple representations which, while successful, might not be drawn on in lower-stakes situations.

In addition to the above, we observe that problem framing may have a powerful effect on student use of representations, possibly a more powerful effect than explicit references to multiple representations in the problem. This observation is tentative, but it reinforces previous results of this nature,¹⁷ and we plan to investigate this more thoroughly in future work.

For instruction, we note that multiple representation use can be taught successfully. Furthermore, an instructor can do so in either a strongly or a weakly directed manner. Neither of these approaches was clearly superior for this purpose, so the instructor has some freedom in choosing between them according to other course goals. Alternatively, one might adapt elements of each.

ACKNOWLEDGMENTS

This work was supported in part by a NSF Graduate Award, a NSF CAREER grant, NSF Grants No. DUE 0241078 and No. DUE 0336713, and Colorado PhysTEC. Special thanks to M. Dubson, E. Etkina, A. Van Heuvelen, and the rest of the Physics Education Research groups at the University of Colorado and at Rutgers University.

-
- ¹J. H. Larkin, in *Mental Models*, edited by D. Gentner and A. Stevens (Lawrence Erlbaum, Mahwah, NJ, 1983).
- ²F. Reif, Millikan Lecture 1994: Understanding and teaching important scientific thought processes, *Am. J. Phys.* **63**, 17 (1995).
- ³A. Van Heuvelen, Learning to think like a physicist: A review of research-based instructional strategies, *Am. J. Phys.* **59**, 891 (1991).
- ⁴A. Van Heuvelen and X. Zou, Multiple representations of work-energy processes, *Am. J. Phys.* **69**, 184 (2001).
- ⁵R. J. Dufresne, W. J. Gerace, and W. J. Leonard, Solving physics problems with multiple representations, *Phys. Teach.* **35**, 270 (1997).
- ⁶D. P. Maloney, in *Handbook of Research on Science Teaching and Learning*, edited by D. Gabel (MacMillan, New York 1994).
- ⁷M. Chi, R. Glaser, and E. Rees, in *Advances in the Psychology of Human Intelligence*, edited by R. J. Sternberg (Lawrence Erlbaum, Hillsdale, NJ, 1982), Vol. 1.
- ⁸R. B. Kozma and J. Russell, Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena, *J. Res. Sci. Teach.* **34**, 949 (1997).
- ⁹R. B. Kozma, The material features of multiple representations and their cognitive and social affordances for science understanding, *Learn. Instr.* **13**, 205 (2003).
- ¹⁰L. McDermott, in *Toward a Scientific Practice of Science Education*, edited by M. Gardner, J. Greeno, F. Reif, A. Schoenfeld, A. diSessa, and E. Stage (IOS, Amsterdam 2004).
- ¹¹D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- ¹²R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the evaluation of Active Learning Laboratory and Lecture curricula, *Am. J. Phys.* **66**, 338 (1998).
- ¹³L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010105 (2006).
- ¹⁴P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping, parts i and ii, *Am. J. Phys.* **60**, 627 (1992).
- ¹⁵D. Rosengrant, E. Etkina, and A. Van Heuvelen, Ph.D. thesis, Rutgers, the State University of New Jersey, 2007.
- ¹⁶C. DeLeone and E. Gire, in Proceedings of the 2005 PERC (AIP, Melville, NY, 2006).
- ¹⁷D. Rosengrant, E. Etkina, and A. Van Heuvelen, in Proceedings of the 2006 PERC (AIP, Melville, NY, 2007).
- ¹⁸P. B. Kohl and N. D. Finkelstein, The effects of representation on students solving physics problems: A fine-grained characterization, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010106 (2006).
- ¹⁹J. P. Mestre, T. C. Thaden-Koch, R. J. Dufresne, and W. J. Gerace, in *Research on Physics Education*, Proceedings of the International School of Physics "Enrico Fermi" Course CLVI (IOS, Amsterdam, 2004).
- ²⁰A. Elby, What students' learning of representations tells us about constructivism, *J. Math. Behav.* **19**, 481 (2000).

- ²¹E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 020103 (2006).
- ²²E. Etkina and A. Van Heuvelen, in *Proceedings of the 2001 PERC* (AIP, Melville, NY, 2002), p. 17.
- ²³A. Van Heuvelen and E. Etkina, *Active Learning Guide* (Addison Wesley, San Francisco, 2006).
- ²⁴E. Etkina and S. Murthy, in *Proceedings of the 2005 PERC* (AIP, Melville, NY, 2006).
- ²⁵<http://www.physics.ohio-state.edu/~physedu/tools>
- ²⁶<http://www.colorado.edu/physics/EducationIssues/>
- ²⁷<http://phet.colorado.edu>
- ²⁸P. B. Kohl and N. D. Finkelstein, The effect of instructional environment on physics students' representational skills, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010102 (2006).
- ²⁹D. Sheshkin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 3rd ed. (CRC Press, Boca Raton, 2004).
- ³⁰We emphasize that this is an evaluation of the representations contained within the problem, not necessarily an evaluation of all the representations used by the students. Students were intended to (and often did) use FBDs and pictures in many of their solutions of Rutgers exam problems that were strictly mathematical in presentation.
- ³¹We had initially planned to give problem 4 separately, but changed this partway through the study, resulting in some recitation sections not receiving problem 4.
- ³²We have found that leaving the data unnormalized tends to prompt the conclusion that the Rutgers students substantially outperformed the CU students, neglecting the previously mentioned problem format difference.
- ³³We will not attempt to argue causation here. It could well be that using multiple representations leads to better performance directly. Alternatively, it could be that better students overall are more likely to use multiple representations. Or, most likely (in our opinions), these are both true, and are intertwined.