

**Student representational competence and self-assessment when solving physics problems**

Patrick B. Kohl and Noah D. Finkelstein

*Department of Physics, University of Colorado, Boulder, Colorado 80309, USA*

(Received 17 May 2005; published 19 October 2005)

Student success in solving physics problems is related to the representational format of the problem. We study student representational competence in two large-lecture algebra-based introductory university physics courses with approximately 600 participants total. We examined student performance on homework problems given in four different representational formats (mathematical, pictorial, graphical, verbal), with problem statements as close to isomorphic as possible. In addition to the homeworks, we examine students' assessment of representations by providing follow-up quizzes in which they chose between various problem formats. As a control, some parts of the classes were assigned a random-format follow-up quiz. We find that there are statistically significant performance differences between different representations of nearly isomorphic statements of quiz and homework problems. We also find that allowing students to choose which representational format they use improves student performance under some circumstances and degrades it in others. Notably, one of the two courses studied shows much greater performance differences between the groups that received a choice of format and those that did not, and we consider possible causes. Overall, we observe that student representational competence is tied to both micro- and macrolevel features of the task and environment.

DOI: [10.1103/PhysRevSTPER.1.010104](https://doi.org/10.1103/PhysRevSTPER.1.010104)

PACS number(s): 01.40.Fk

**I. INTRODUCTION**

Student competence with different representational formats is a popular topic in modern science and mathematics education. By “representational formats,” we refer to the many ways in which a particular concept or problem can be expressed. One can categorize problem representations according to whether they are formal or informal, abstract or concrete, or text based versus graphics based, to name just a few. Studies involving representational formats have taken many approaches to this division of representations, including comparisons of mathematical problems couched in words to those stated primarily in equations,<sup>1</sup> comparisons of learning environments that are virtual to those that are physical,<sup>2</sup> and comparisons between verbal, mathematical, graphical, and diagrammatic formats.<sup>3</sup> Scientists can interpret all of these formats effectively and are able to integrate them, translate among them, and assess their usefulness in different situations. A possible instructional goal is to develop this cross-representational facility in science students. In physics education research (PER), there have been several studies in which students are explicitly taught to handle multiple representations of the same topic.<sup>4–6</sup>

There have been a number of studies (both in<sup>7–9</sup> and outside of PER) that examined student performances on particular representations of physics problems (see Meltzer for a short overview<sup>3</sup>). There have also been studies that compare student performance on problems that involve multiple representations to performance on problems that involve isolated representations,<sup>10,11</sup> and studies that investigate student skill at translating between representations.<sup>12,13</sup> There has, however, been less research that compares student performance on different single representations of a problem. Koedinger and Nathan<sup>1</sup> studied the performance and strategies of middle-school algebra students solving both word- and formula-based representations of problems. He found

significant performance differences between the formats, and noted that these differences correlated with the different strategy choices that students made when handling different representations. In Meltzer's work,<sup>3</sup> he provides students in an introductory level algebra-based physics class with quizzes that have nearly isomorphic problems represented four different ways (verbally, mathematically, graphically, and diagrammatically). His prior lectures and assignments made heavy use of the different representations seen on the quizzes. After averaging over several years of data for the same class, Meltzer found instances where students performed significantly better on one representation of a problem than on another. He also found that students were not necessarily consistent in their performance on the same representations across topics, leaving open the question of why these performance differences exist.

We should note that the meaning of the term “problem” has not been completely specified and agreed upon in these communities. Many researchers implicitly treat problems as always involving some quantitative analysis (including but not limited to Refs. 5 and 14). Others do not make such a restriction.<sup>3</sup> We do not wish to debate the proper use of the term “problem” in this paper, and we will simply use the term to refer to typical physics tasks given to students, for example those found in homework assignments and those studied here. This will include questions that do not require calculation.

In this study, we directly compare student performance on different representational formats in the style of Meltzer. In addition, we perform the study in two different classroom environments (one mostly traditional, and one reformed). We divide study problems into verbal, mathematical, graphical, and pictorial formats. We do not mean to imply that these representations span the space of all possible representations nor that they are mutually exclusive, but we do consider this set to cover most of the representations seen by students in lecture and in standard texts. We also consider these repre-

sentations to be fundamental in physics generally. We should point out that variations are possible within any particular category of representation; for example, our pictorial format includes both literal pictures of diffraction patterns and schematic pictures of Bohr-model electron orbits.<sup>15</sup> As we will see, the data show that student success with one representation of a problem does not necessarily imply success with another, a result that may have significant implications. Conceptual surveys, for instance, may need to carefully attend to the various possible representations of a topic to avoid false positives where a student is assumed to have mastered a concept based on performance on one or two representational formats. Furthermore, we will see that success with a particular representation of a topic does not necessarily imply success with that same representation of a different topic. Note that this last point is somewhat confounded by the fact that representations vary from topic to topic; for instance, a graph associated with one topic might differ significantly in style from a graph associated with another topic. Nevertheless, we consider this division of representations into categories to be productive and consistent with general use in physics.

We also broaden the examination by investigating whether students can assess their own representational competence, what motives they have for handling a problem in a particular representational format given a choice of formats, and whether providing this choice affects their performance compared to students randomly assigned to particular formats. These questions relate to students' "metarepresentational competence," a notion that diSessa and Sherin have developed in their work.<sup>16</sup> There, diSessa and others investigated students' ability to generate their own representations of a concept or situation, sometimes in cases where they have very little formal training.<sup>17</sup> They have also considered students' ability to assess and critique the representations that they generate.<sup>18</sup> Our study differs from these in that we ask students to assess fairly standard representations that we have provided rather than ones they have generated themselves. We also have them assess their own skills and preferences regarding these standard physics representations, in part by choosing which representational format they would like to work with on a quiz. There are a number of outcomes that we might observe here. It may be that students have well-defined learning styles and are aware of them, enabling these students to increase their performance given a choice of representation. It may be that students perform in a relatively consistent way across representational formats but are unaware of their strengths, leading to unchanged or even reduced performance when given a choice. Or it may be that students' performance when given a choice of representations varies and is difficult to predict, with some topics and representations resulting in improved performance and other topics and representations resulting in lower performance. This would suggest a more complicated explanation of how their performance varies as it does, one that must attend carefully to both micro- and macrolevel features of the context, and this is in fact what we find.

In short, we have four primary goals.

(1) To further demonstrate that student performance varies, often strongly, across different representations of physics problems with similar content.

(2) To investigate why students perform differently on these different representations.

(3) To show that giving students a choice of representational format will change their performance either for better or for worse, depending on the circumstances.

(4) To begin to explain how providing a choice of representation results in these performance differences, and to note the possible effects of different instructional techniques.

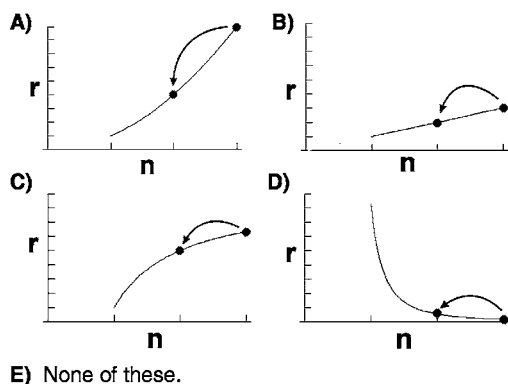
## II. METHODS

We administered our study in recitation to two large (546- and 367-student) algebra-based introductory physics classes at the University of Colorado at Boulder. These courses are composed primarily of students taking the class to satisfy the requirements for life science, social science, and premedical programs. These students are typically in their second or third year of study. College algebra is a prerequisite, though in practice student math skills are quite varied. The first course in our study was an on-sequence second-semester class (Physics 202) held in the spring of 2004. The format was mostly traditional, albeit with some in-lecture qualitative and quantitative concept tests using a personal response system. Students had three one-hour lectures per week, and met for two hours each week in either a recitation and lab. The recitation or laboratory part of the course was directed by a different professor than the lecture portion. The recitations were generally traditional, with students spending most of their time discussing homework and exam questions with a graduate teaching assistant (TA). The labs focused mostly on investigation, testing predictions, and completing open-ended tasks (that is, tasks where the students were given a general goal but no specific directions for how to accomplish that goal). Students' grades were based on exams, labs, homework assignments (both online<sup>19</sup> and long answer), and participation in the concept tests.

The second course was an on-sequence first-semester class (Physics 201) in the fall of 2004. This course precedes 202 in the standard sequence, but this particular 201 section took place the semester following the 202 class mentioned above, and so each group was being exposed to the study for the first time. The 201 course was taught by a different professor, who is familiar with many of the major results of physics education research. The 201 class was largely reformed, with heavy use of interactive concept tests and an emphasis on tightly integrated lecture demonstrations. The students had the same number of lectures, recitations, and labs as the 202 students. The recitation and laboratory section was taught by the lecture professor and another professor working together. The recitations focused on working through problems rich in context in small groups, with some demonstrations and some time reserved for homework and exam questions. The labs were a mixture of directed work, open-ended questions, and testing predictions. Students' grades were determined in much the same way as in 202. For the sake of comparison, we videotaped three lectures from the 201 and 202 courses. 57% of the 201 class time was spent on interactive concept tests versus 23% of the 202 class time, supporting the notion that the 201 course had a

**Question 3 – Graphical**

An electron in a Bohr hydrogen atom jumps from the  $n=3$  orbit to the  $n=2$  orbit. The following graphs show the orbit radius  $r$  as a function of the orbit number  $n$ . Choose the graph that best represents the relative locations of the electron orbits.

**Question 4 – Pictorial**

An electron in a Bohr hydrogen atom jumps from the  $n=3$  orbit to the  $n=1$  orbit. Choose the picture that best represents the relative locations of the electron orbits.

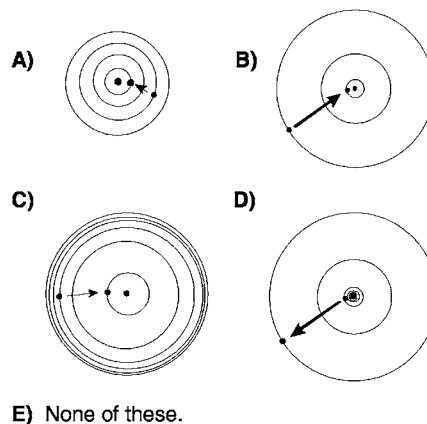


FIG. 1. Isomorphic homework problems (in graphical and pictorial formats) regarding Bohr-model electron orbit radii.

greater commitment to reform-style student engagement.

For the 202 class, we performed the study in two different subject areas: wave optics and atomic physics. The general subject areas were chosen based on which weeks the recitations were available for study; we attempted to avoid weeks with exams or holidays. The students were assigned four multiple-choice homework questions that covered the same concept in four different representational formats, as well as a one-question multiple-choice quiz given in recitation. We selected specific subtopics that were covered in class and were amenable to representation in a number of different formats. The quiz subtopics were also chosen to match material covered in lab in the hopes that the extra time on task from the laboratory would better prepare students to choose between representations. These homeworks were assigned online as prerecitation questions and were turned in at the start of the recitation section. Students were expected to turn in prerecitation homeworks each week and were prepared for the possibility of quizzes, and so these study materials did not represent a significant departure from the norm. The study quizzes were administered by their section TAs. All of the homework and quiz problems are available in the Appendix.

An example of two of four homework problems from one of the two 202 assignments is shown in Fig. 1. After turning in the homeworks, the students were given the one-question quiz in one of four representational formats. These quiz problems were isomorphic from format to format, with the answers and distractors mapping from one format to the next. It is worth noting that we use the word “isomorphic” to mean isomorphic from the point of view of a physicist. A student may have a different view of the similarity (or lack thereof) between these problems.<sup>20</sup> We also refer to isomorphism between problem statements and answer choices. We consider it likely that student solution strategies will be considerably less constant across representation.

Nine of the thirteen 202 recitation sections were allowed to choose from the four representational formats on the quiz

without getting to see the problems before they selected. Our intent was for the students to make their choice based on their previous experience with representations in classes and on the homework assignment. The other four sections had quiz formats randomly distributed to the students; these students served as a control group. We provided more of the recitation sections with a choice of format to ensure that a reasonable number of students chose each format. The choice and control sections did not change from one subject area to the next, and the students in the two groups performed similarly on the study homeworks, the course exams, and in the course overall. Both the quizzes and homeworks included a Likert scale survey on which the students could rate the perceived difficulty of the question, and the quizzes included a section where the students were asked to write about why they chose the format they did (if they had a choice) or which format they think they would have performed best at given the choice (if they had a random assignment). Both the quizzes and the homeworks counted toward the students’ recitation scores for participation but were not otherwise graded.

The study was conducted in much the same way in the 201 class. We covered two subject areas: energy (in particular, kinetic and potential energies and their connection to motion) and pendulums. For the energy and motion topic, the students received a four-question pre-recitation homework and an in-recitation quiz. The 201 class was larger (attrition shrinks the 202 class in relation), and so we were able to designate nine of the 18 recitation sections as control sections, with the remaining nine receiving a choice of quiz format. For the pendulum topic, we gave the students a recitation quiz only (no homework) in order to satisfy schedule constraints. Again, the choice and control groups were the same from one topic to the next, and the two groups performed similarly on homeworks, exams, and the class overall.

In this paper, we restrict our attention to students who completed a homework (when there was a homework) and

TABLE I. Fraction of students answering a homework problem correctly, broken down by representational format and topic. Standard errors vary but are on the order of 0.02.

	Verbal format	Mathematical	Graphical	Pictorial
202 Diffraction and Interference HW ( $N=241$ )	0.52	0.61	0.46	0.54
202 Bohr model HW ( $N=218$ )	0.84	0.83	0.76	0.62
201 Mechanics and energy HW ( $N=333$ )	0.54	0.70	0.50	0.49

the corresponding quiz for a topic, which amounts to roughly 240 and 220 students in the first and second 202 studies, and 330 students in each of the two 201 studies.

### III. DATA AND RESULTS

In this section, we focus on comparisons of student performances on similar problems in different formats and comparisons of student performance in choice and random-assignment (control) recitation sections. We also examine why students made use of the representations they did and how they used multiple representations when they did.

Table I shows the fraction of students (in both choice and control sections) that answered each of the twelve homework (HW) problems (four formats in three different topics) correctly. Table II shows the performances of the students on each format of each in-recitation quiz, grouped by whether they were in a choice or control section. The number of students in each subgroup appears in parentheses.

#### A. Performance across representational format

All statistical significance tests involving student success rates are two-tailed binomial proportion tests. We shall use the following terminology: A difference with  $p > 0.10$  is referred to as not significant,  $p$  between 0.10 and 0.05 is marginally significant,  $p$  between 0.05 and 0.01 is significant, and  $p < 0.01$  is highly significant.

##### 1. Homework problems

In examining the homework data, we note that in several cases there were differences in performance from format to format on a particular assignment. When there was a difference in performance between two formats, the mathematical format was often one of the formats involved. This was the only format to require an explicit calculation. The other for-

mats involved conceptual reasoning supported by descriptive language, graphs, or pictures. We see that on average students were most successful with the mathematical homework format, which is consistent with the notion that first-year university physics students are more comfortable with “plug ‘n chug” types of problems than with conceptual problems.<sup>21,22</sup> We should point out that a mathematical format need not always involve numerical calculation; indeed, one of the math-format quiz questions (to be described later) was best solved through conceptual reasoning supported by the qualitative use of equations. Nevertheless, in this study the mathematical format usually involved direct calculation.

We also see that there are some noticeable performance differences among the more conceptual formats. For instance, consider the graphical and pictorial problems on the Bohr model assignment, shown in Fig. 1. Both require knowledge of how the electron orbit radius varies with the principal quantum number in the Bohr model. The questions differ only in which specific transition is being presented and in whether the problem and solutions are expressed in graphs or pictures. Of the 218 students who answered both problems, 76% answered the graphical problem correctly and 62% answered the pictorial problem correctly. This difference is highly significant statistically ( $p=0.006$ ) and is particularly interesting in that the graphical representation is a rather nonstandard one. Students had not seen any graphs of orbital radius versus quantum number, but the pictorial representation of electron orbits should have been somewhat familiar since it is featured in both the textbook and the lectures that preceded this quiz. Further examination of the individual student answers on these two questions indicates that this performance difference can be attributed almost entirely to the 36 students who answered the graphical problem correctly and missed the pictorial problem by choosing the distractor C. This distractor bears a strong resemblance to the energy-level diagrams seen in the Bohr model section of the

TABLE II. Quiz performance of students from the random-format recitation sections (left) and from the recitations sections that had a choice of formats (right). The number of students taking a quiz is in parentheses. The quiz topics are diffraction, spectroscopy, springs, and pendulums. Standard errors vary and are not shown.

	Control (random-format) group				Choice group			
	Verbal	Math	Graphical	Pictorial	Verbal	Math	Graphical	Pictorial
202 Diff.	0.24 (17)	0.56 (18)	0.25 (16)	0.58 (19)	0.35 (17)	0.37 (57)	0.04 (26)	0.82 (59)
202 Spec.	0.32 (19)	0.13 (15)	0.53 (17)	0.83 (18)	0.81 (21)	0.90 (42)	0.96 (27)	0.39 (58)
201 Springs	0.56 (43)	0.41 (39)	0.69 (42)	0.58 (40)	0.55 (11)	0.57 (102)	0.88 (17)	0.77 (39)
201 Pend.	0.55 (42)	0.30 (40)	0.64 (39)	0.67 (43)	0.62 (21)	0.39 (28)	0.65 (40)	0.78 (80)

**Spectroscopy Problem – Pictorial Format**

The Balmer series of spectral lines is shown below, as seen through a spectrometer:



Now suppose we are in a world where electric charges are weaker, so the electron is not held as tightly by the nucleus and the ionization energy is 13 eV instead of 13.6 eV. Choose the picture that best represents what the new spectrum would look like.

**Spectroscopy Problem -- Verbal Format**

Consider the Balmer series of spectral lines from hydrogen gas. Now suppose we are in a world where electric charges are weaker, so the electron is not held as tightly by the nucleus. This means that the ionization energy for the electron will be smaller. What will happen to the Balmer lines that we see?

B) The spectral lines will all shift to shorter wavelengths (toward the bluer colors).

FIG. 2. (Color) Setup and second answer choice for the verbal- and pictorial-format quizzes given in the second trial. The other distractors match up between the formats as well.

text and lectures. Since the problems are so similar and the same distractors are present in each problem, it appears that in this case representational variations may be traceable to a very topic-dependent cueing on visual features of one of the problems.

## 2. Quiz problems

We can find another example of performance variation across isomorphic problem presentations in the second 202 quiz, which deals with the emission spectrum of a Bohr-model hydrogen atom. The students were prompted to recall the spectrum of hydrogen, and were asked how that spectrum would change if the binding of the electron to the nucleus were weaker. The questions, answers, and distractors were the same on each quiz except for their representation. Figure 2 shows the problem setups and one distractor for the verbal and pictorial formats (performance data are in Table II). Note that one week previous to the quiz, students completed a lab covering emission spectroscopy, and the quiz images match what students saw through simple spectrometers. Nineteen students in the control group were randomly assigned a verbal-format quiz, and 18 were assigned a pictorial-format quiz. 32% of the verbal group answered the question correctly, while 83% of the pictorial group answered correctly. This difference is highly significant ( $p=0.0014$ ). Answer breakdowns indicate that eight of the ten students in the verbal group that missed the question chose the distractor corresponding to the spectral lines moving in the wrong direction (pictured in Fig. 2). Only one student from the pictorial group made this error. It is not clear why there would be such a split, especially since the pictorial format shows numerically larger wavelengths as being on the left, opposite the standard number line convention. A possible hypothesis is that students connect the pictorial format more closely to the lab, giving them additional resources with which to handle the problem. However, as we will see, the students that were

given a choice of format performed significantly worse on this pictorial format despite being more likely to cite the lab in making their choice, and so easy identification with the lab cannot be a complete explanation.

Next, consider the performance of control group students on the mathematical formats of the 201 and 202 quizzes. In three of the four quizzes, the average success rate on the math quiz was significantly lower than the average success rate on the other three formats combined. For the spectroscopy quiz, the average verbal/graph/pictorial score was 0.56 versus 0.13 on the math format, a difference significant at the  $p=0.004$  level. For the 201 spring quiz, the difference was 0.61 vs 0.41 ( $p=0.03$ ), and for the 201 pendulum quiz, the difference was 0.62 vs 0.30 ( $p=0.0004$ ). The difference between the average verbal/graph/pictorial score and the average math score on the 202 diffraction test was marginally significant ( $p=0.09$ ). It is somewhat surprising that students were less successful with the randomly assigned math format given their generally higher performance on the equation-based homework problems, though we should note that the students took the quiz in recitation with a time limit (about fifteen minutes) and without access to a textbook, making the environment much different from that in which they would do a homework problem. We should also note that the math problem on the 201 spring quiz was difficult to solve through explicit calculation, and was more easily handled by using the equations qualitatively. This gives it a different character from the other math-format problems, which is a point we shall return to later.

In closing this subsection, we note that in addition to analyzing homework or quiz problems alone, one can examine whether performance on homeworks is correlated to performance on quizzes in a number of ways. For example, one can ask whether performance on a quiz is correlated to performance on the corresponding homework problem format. Generally, such homework-quiz correlations were very weak, and are not explored further here.

TABLE III. Statistical significance of the quiz performance differences between the format choice and control groups in the 202 and 201 sections. Numbers are  $p$  values using a two-tailed binomial proportion test.  $\times$  denotes a  $p$  value of greater than 0.10. Bold indicates that the choice group had higher performance than the control group.

Quiz subject	Verbal	Math	Graphical	Pictorial
202 Diffraction	$\times$	$\times$	0.04	<b>0.03</b>
202 Spectroscopy	<b>0.002</b>	<b>0.0001</b>	<b>0.0004</b>	0.001
201 Springs	$\times$	<b>0.09</b>	$\times$	<b>0.07</b>
201 Pendulums	$\times$	$\times$	$\times$	$\times$

### B. Effect of student choice of representation

In Table II, we see a format-by-format comparison of the students who received a quiz at random and the students who were allowed to choose a quiz format. There were a total of 16 choice-control comparisons available (four trials with four formats each). Of the eight from the 202 class, six showed a statistically significant performance difference. These data, along with the significances of the choice-control differences (or lack thereof) in the 201 class, are summarized in Table III.

These results are notable in that the effects are in some cases quite strong. For instance, 90% of the 42 students in the choice group answered the math-format question correctly for the spectroscopy topic, while 13% of the 15-student control group answered the same problem correctly. In addition, the direction of the effect can vary. In four of the six cases, giving students a choice of formats significantly increased performance, while in two of the six cases it resulted in a significant decrease. Furthermore, when comparing across content areas we see reversals in the direction of the effect. On the diffraction quiz, students in the choice group do better than the control group on the pictorial representational format and worse on the graphical representational format, while on the spectroscopy quiz the students in the choice group do worse on the pictorial representation and better on the graphical representation. As we can see, giving students a choice of format does not result in consistently increased or consistently decreased performance relative to the control groups. Rather, the direction of the effect appears to vary strongly across both topic and representation, which suggests two things. First, these students do not have the metarepresentational skills necessary to consistently make productive representational choices under these circumstances. Second, a complete explanation of these performance differences will likely be nontrivial and will not be able to rely entirely on broad generalities.

We can further characterize student performance in these cases by considering which distractors they chose. As was mentioned above, the control groups for the pictorial and verbal formats of the 202 spectroscopy quiz (see Fig. 2) showed a significant performance difference, with the errors made by the verbal format control group being concentrated almost entirely on the distractor B in which the spectral lines move in the wrong direction (other distractors include the

lines compressing, the lines staying the same, and none of the above). The corresponding choice groups did the reverse. The verbal-format group had 17 out of 21 people answer correctly, with three choosing the distractor B. The pictorial group had 23 of 58 students answer correctly, with 27 students selecting the distractor B. Thus, we see that the students who chose a verbal-format quiz performed in very nearly the same way as the students who received the pictorial format at random, in terms of both success rate and choice of distractors. Similarly, the students who chose a pictorial quiz performed in the same way as the students who were randomly assigned a verbal quiz.

In general, of the six statistically significant 202 choice-control comparisons, the performance difference in two of them (spectroscopy verbal and pictorial) was mainly attributable to students focusing on a particular distractor. In the other four (spectroscopy mathematical and graphical, diffraction mathematical and verbal) the incorrect answers were split among two or more distractors. Note that the quiz distractors map from one format to the others, so this is not simply a case of some of the problems not having any attractive distractors, though apparently (and perhaps not surprisingly) different representations of a problem can make different distractors more or less attractive.

Next, consider the eight choice-control comparisons from the 201 section (Tables II and III). None of the pairs showed different performance at a  $p=0.05$  significance level. Two were marginally significant (the math-format spring quiz at  $p=0.09$  and the pictorial-format spring quiz at 0.07). There was very little difference in performance between the choice and control groups on the pendulum quiz, which was given four weeks after the spring quiz. The difference between these data and the corresponding 202 data is pronounced. Students in 201 did roughly as well regardless of whether they received their preferred format or a format at random, suggesting that their representational skills are more balanced. That is, they are less likely to have much more trouble with a random representation than with their representation of choice. Since one of the major differences between the 201 and 202 groups was the method of instruction, it may be that the instruction contributed to the effect. We also should note that the 201 and 202 studies involved different topics, which may have contributed to the different performances. We are currently running a comparison of two 202 sections with different professors teaching the same topics. The results will be part of a follow-up paper that allows us to explore the effects of instructor independent of content.

### C. Student self-assessment and assessment of the representations

In this section we consider data intended to address two related questions. First, how do students assess and value different representations? Second, how (and how successfully) do they assess their own representational competence?

The students in the format choice groups were asked to “write a few sentences about why you chose the problem format you did.” We then coded these responses, separating them into categories that developed as the coding proceeded.

TABLE IV. Reasons 202 students gave for choosing a particular representation of a quiz. Only the largest categories are presented here. The numbers in the format column are the numbers of usable responses.

202	Diffraction quiz	202	Spectroscopy quiz
Verbal	3: Preference for qualitative analysis	Verbal	3: Preference for concepts over math/pictures
10	1: Connected it to the prerecitation HW	12	3: Don't like pictures
	1: Preference for concepts over math		2: The format supports the concepts
Math	9: Preference for "plug 'n chug" problems	Math	5: Preference for "plug 'n chug" problems
24	4: Find equations/numbers easy	18	4: Preference for mathematics over concepts
	3: Preference for math over pictures		3: Preference for mathematics over pictures
Graph.	3: Visual learners/people	Graph.	2: Visual learners/people
12	2: Like having a visualization provided	10	2: Preference for visuals over math
	2: Connected it to the prerecitation HW		2: Connected it to the prerecitation HW
Pict.	17: Visual learners/people	Pict.	12: Liked the colors/found it attractive
51	12: Connected it to lab	35	8: Like having a visualization provided
	9: Find other formats difficult		6: Connected it to lab

In Tables IV and V we present the three dominant categories for each quiz. The complete set of data is in the Appendix. Some remarks regarding our categorization methods: Students in the "visual learners" category have explicitly identified themselves as visual learners or visual people. People that expressed a preference for plug and chug problems used language that clearly indicated the insertion of numbers into formulas in a simple fashion, and always used the words plug and/or chug. Students that remarked that they simply found equations or mathematics easier to handle or more straightforward were placed in other categories. In some cases, there is a category for those that chose a format because they were attracted to it and a separate category for those that chose a format because they were avoiding a different one. Many of the responses were too vague to be useful, "pictures are pretty," for example. These were discarded.

There are a few notable trends. First, 72% of all the choice group students (including those who did not make

comments) selected either a math- or pictorial-format quiz. We also see that the vast majority of students who cited their lab in explaining their choice chose the pictorial format. This is despite the fact that the recent lab included representations that corresponded to each quiz format.

There are a fair number of students that chose the mathematical format expecting a plug and chug style of problem, except in the case of the 201 pendulum quiz, which followed the 201 quiz on springs. The 201 quiz on springs was unique in that the mathematical format quiz was difficult to handle through explicit calculation alone, and favored qualitative reasoning supported by equations. Eighteen students taking the 201 pendulum quiz mentioned that they did not like the math format for the earlier spring quiz, with 13 of these choosing the pictorial format the second time. It would appear that in this case there was a mismatch between the students' conception of what constituted a math problem (plugging and chugging) and our conception (either calculation or

TABLE V. Reasons 201 students gave for choosing a particular representation of a quiz. Only the largest categories are presented here. The numbers in the format column are the numbers of usable responses.

201	Spring quiz	201	Pendulum quiz
Verbal	6: Preference for concepts/intuition	Verbal	4: Preference for concepts over math/pictures
9	3: Other formats are difficult	17	4: Clear/ordered presentation
			3: Other formats are difficult
Math	16: Preference for math over concepts	Math	7: Comfortable handling equations
74	12: Like the straightforward/definite nature	22	6: Like the straightforward/definite nature
	11: Comfortable handling equations		2: Preference for math over concepts
	7: Preference for plug 'n chug problems		
Graph.	3: Visual learners/people	Graph.	6: Visual learners/people
11	2: Like having a visualization provided	27	6: Like having a visualization provided
	2: Connected it to the prerecitation HW		3: Didn't like the math format before
Pict.	12: Visual learners/people	Pict.	15: Visual learners/people
26	7: Like having a visualization provided	62	13: Didn't like the math format before
	5: Find other formats difficult (esp. math)		12: Like having a visualization provided

using equations as a conceptual tool), and the students responded accordingly.

In the 202 class, 9% of the people who initially chose a verbal format quiz stayed with that format for the second quiz. Twenty-nine percent of the graphical, 42% of the pictorial, and 46% of the mathematical groups stayed with their format. For the 201 section, 73% of the verbal, 25% of the math, 71% of the graphical, and 79% of the pictorial groups stayed with their choice of format from the first quiz to the second. For all formats but the math (which, for the 201 spring quiz, was different in character from the other math problems in this study), the 201 section was substantially more likely to stay with their choice of format. Of the 76 students in 201 that changed from math on the first quiz to a different format on the second, 11 chose verbal, 22 chose graphical, and 43 chose pictorial. The strong preference for the pictorial format during this switch, the fractions of the class selecting either a math or pictorial quiz, and the student comments overall are all consistent with the notion that students perceive the mathematical and pictorial formats to be dominant and antithetical. That is, when considering the different possible representations of a physics problem, students appear to think primarily of pictorial and mathematical formats (and not so much of others) and to think of these formats as opposites in a sense.

In both the 201 and 202 sections, many of the students who selected a graphical or pictorial format identified themselves as visual learners (15 and 7 of the students on the first and second 202 quizzes, and 15 and 21 students on the first and second 201 quizzes). No students identified themselves as any other type of learner, save one that identified himself as a kinesthetic learner and chose a mathematical format. In the cases of the pictorial formats of the 202 diffraction quiz, the 201 spring quiz, and the 201 pendulum quiz, there were enough self-identified visual learners to compare their performance to the other people choosing the same format. There were 18 self-identified visual learners in the 202 diffraction quiz, who had a success rate of 0.89 as compared to the success rate of 0.78 for the other 41 students. This difference is not statistically significant ( $p=0.33$ ). For the 201 spring quiz, there were 12 self-identified visual learners, who had a success rate of 1.00 as compared to 0.67 for the other 27 students. This difference is significant at the  $p=0.02$  level. For the 201 pendulum quiz, there are 15 self-identified visual learners. These had a success rate of 0.87, as compared to 0.75 for the other 65 students. This difference is not statistically significant ( $p=0.35$ ). Averaging the above, the self-identified visual learners had a success rate of 0.91 versus 0.75 for the other students, which is significant at the  $p=0.02$  level. There are a number of confounding factors that leave us hesitant to draw conclusions based on these data alone. Both the students' abilities to assess their own competencies in this fashion and the overall usefulness of categorizing people as different types of learners are somewhat unclear here, and we also note that the students made these identifications (or did not) *after* having succeeded or failed at visual and/or nonvisual tasks.

#### D. Students' use of multiple representations

The students in this study were provided with single representations of quiz problems, but in many cases the stu-

dents' papers showed that they had made explicit use of supplementary representations in solving the quiz. This was most often a picture that they drew in support of a mathematical- or verbal-format problem. Some students wrote equations in support of nonmathematical formats and a handful wrote out physical principles longhand or drew a graph. It is a common goal in physics education to teach the use of multiple representations,<sup>4,6</sup> so it is interesting here to compare the performance of students that produce supplementary representations to those that do not. We should emphasize that neither of the courses studied here made an explicit attempt to teach the use of multiple representations in the style of the just mentioned references. We should also note that the students who had no explicit supplementary representation may well have used multiple representations in their solution to some extent (it is hard to conceive of a student that can think strictly in terms of one representation and no other), but we shall focus on the students that made these explicit.

On the 202 diffraction quiz, 51% of the 172 choice group students made explicit use of some supplementary representation. These students had an average success rate of 0.47, compared to 0.54 for the students that did not explicitly use a supplementary representation. This difference is not statistically significant ( $p=0.12$ ). Breaking it down by format, 35–40% of each of the verbal, graphical, and pictorial groups used a supplementary representation, and in each case the group using such a representation did not do significantly better or worse than the group that did not explicitly use an additional representation. Of the students that chose a math format, 75% (43 out of 57) used a supplementary representation, which was always a picture. These students had a success rate of 0.28, compared to 0.64 for the 14 that did not. This performance difference is significant at the  $p=0.014$  level.

That the performance difference should favor students that did not draw a picture is surprising, and so we examined the problem in more detail. The math-format quiz question is displayed along with a student drawing in Fig. 3. This is a question regarding the diffraction pattern coming from two finite-width slits illuminated by monochromatic light, a topic that was featured in a lab but was covered minimally in lecture. The pattern will have a narrow peaks separated by a distance  $X$  governed by the slit separation  $D$ , and a longer-period envelope with peaks separated by a distance  $x$  governed by the slit width  $d$ . Given the distance from the slits to the screen  $L$ , the wavelength of the incident light  $\lambda$ , and either  $X$  or  $x$ , one can calculate either  $D$  or  $d$  using  $(d, D)\sin(\theta)=n\lambda$ . Most student errors involved mixing up  $D$  and  $d$ . We examined each student picture (which was often a hybrid of a picture and a graph) and categorized it. Almost no one drew a correct two-slit diffraction pattern, suggesting that this topic was not well understood at this point. Of the 35 students that drew a picture and answered the question wrong, 13 students drew a picture that represented a single-slit diffraction pattern with peaks separated by  $x$ , which led to a mix-up of  $D$  and  $d$  in the equation. There were also eight students that drew such a picture with peaks separated by  $X$  and then used the equation appropriately, answering the question correctly. Fourteen of the students that drew a pic-

**Diffraction Problem -- Mathematical Format**

We have a double-slit experiment with incident light of  $\lambda = 633 \text{ nm}$ . On a screen  $3.0 \text{ m}$  from the slit, we see an intensity pattern with small peaks separated by  $0.5 \text{ cm}$ . The first minimum in the overall intensity envelope is at  $2.0 \text{ cm}$  from the center of the pattern. Calculate the separation of the slits,

d. Circle the appropriate letter.

- A)  $D = 3.8 \times 10^{-5} \text{ m}$
- B)  $D = 3.8 \times 10^{-4} \text{ m}$
- C)  $D = 9.5 \times 10^{-5} \text{ m}$
- D)  $D = 9.5 \times 10^{-4} \text{ m}$
- E) None of the above.

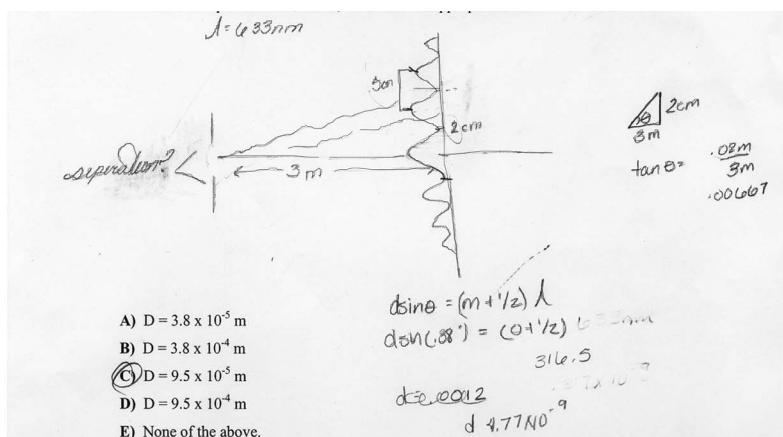


FIG. 3. A student’s use of a supplementary representation (hybrid graph and picture) to solve the math-format 202 diffraction quiz.

ture drew a single-slit diffraction pattern with both  $x$  and  $X$  labeled as follows:  $X$  was marked off between two peaks far from the center, and  $x$  was marked off as the distance from the centerline to the first minimum. These students did not appear to notice that the distance labeled  $0.5 \text{ cm}$  on their paper was roughly twice as wide as the distance labeled  $2 \text{ cm}$ . This drawing was an apparent misinterpretation of the phrase “the first minimum in the overall intensity envelope is at  $2.0 \text{ cm}$  from the center of the pattern” present in the problem statement. This language is similar to that of the text and of the lab that covered this topic, though this is no guarantee that it would be understood. Of these 14 students, two answered correctly and 12 answered incorrectly, calculating  $d$  instead of  $D$ . These 12 students account for much of the performance difference between the picture and no-picture groups.

These data recall Chi *et al.*'s<sup>20</sup> finding that in some cases novice problem solvers draw more pictures than experts while making more errors. This suggests that one motivation for using multiple representations is to work through something perceived to be difficult. However, the students that drew a picture rated the problem to be just as difficult as the students that did not draw a picture. On a Likert scale from 1 (easiest) to 5 (hardest), the students that drew a picture gave this problem an average rating of 3.76 while those that did not draw a picture gave a rating of 3.79. It is thus not clear whether the students that struggled with this problem were more likely to draw a picture. There have been other studies in which including multiple representations of a problem resulted in poorer performance than using single representations. This performance difference was interpreted broadly either as an increase in cognitive load when the representations are separated<sup>11,23</sup> or as an increase in load stemming from an inability to map from one representation to the next.<sup>10</sup> The case here is somewhat different in that it appears to be tied to the specific contextual features of the problem and the problem solvers. It seems that here the higher error rate among students using multiple representations is traceable to a particular misunderstanding of the problem statement that was much more likely to be detrimental if it was expressed in a pictorial manner. If the problem or the general background of the students on this topic had been slightly

different, this likely would not have occurred.

On the 202 spectroscopy quiz, ten out of 148 students in the choice group used a supplementary representation. This is too small a sample for analysis. We do find it notable that there would be such a large difference from topic to topic, with 51% of students using an explicit supplementary representation for the diffraction quiz and only 7% using one for the spectroscopy quiz. The average success rate across all students on this quiz (choice and control) was 0.62, which is significantly greater than the 0.42 for the students on the diffraction quiz ( $p=0.0004$ ). In contrast, the choice and control students gave the spectroscopy quiz a difficulty rating of 3.60 averaged across all formats, compared to the rating of 3.47 for the diffraction quiz. Thus, it appears that the spectroscopy quiz was easier for the students, though they did not rate it as such. It is not clear whether this influenced their decision to use an explicit supplementary representation.

On the 201 spring quiz, 74 of the 169 students in the choice group used a supplementary representation. Sixty-nine of these were students that had chosen a math-format problem. These students had a success rate of 0.55 as compared to 0.61 for the 33 students that chose a math format and did not use a supplementary representation. This difference is statistically insignificant ( $p=0.60$ ). The use of supplementary representations was less common but somewhat more spread out for the control group on the same quiz: 9, 23, 9, and 4 students used a supplementary representation on the verbal, math, graphical, and pictorial formats, respectively. This variation accounts for 45 students out of 164, or 27%. The 23 students that used a supplementary representation on the math format had a success rate of 0.43, which did not differ significantly from the success rate of 0.38 achieved by the 22 students that did not use a supplementary representation ( $p=0.62$ ). These data were very similar for the 201 pendulum quiz.

In summary, students that use explicit supplementary representations on these quizzes are roughly as successful as those that do not, with one case in which they are less successful. This finding is consistent with the cognitive science results mentioned previously in which researchers found that multiple representations do not necessarily increase performance. Rather, multiple representations are tools that students can either use productively or not.

#### IV. DISCUSSION AND CONCLUSION

This study began with a number of goals. First, we wished to know whether student performance on physics problems varies with representational format. We see evidence that it does, often strongly. In the case of the Bohr-model homework problem, the performance difference between the nearly isomorphic graphical and pictorial problems is due to students selecting a particular distractor. This distractor is one that superficially resembles energy-level diagrams that they have seen associated with this material, but only when it is represented pictorially. We also see students in the random-format groups doing much better on a pictorial-format spectroscopy quiz than on a verbal format of that same quiz. It is less clear what might have triggered this. While the choice groups make it clear that students connect the pictorial format more closely to their laboratory experiences, the laboratory did not ask them to consider this particular concept (though it did make use of similar representations). This issue is further confounded by the fact that the students that were allowed to choose a pictorial format, in the process perhaps identifying themselves as students connecting more strongly with the laboratory, did significantly worse than the students that chose a verbal format.

We note that students that were randomly assigned a mathematical quiz did significantly worse in three of the four cases than students assigned any other format. This was true in two cases when the math problem involved simple calculation. This is surprising since the selections and comments of the choice groups, in particular the reasons cited for the move away from the math format in response to the first 201 quiz (a problem that was not plug 'n chug), suggested that many of the students preferred plug 'n chug problems to other sorts. The poorer performance on the mathematical format was also present on the aforementioned 201 quiz where the mathematical format was more easily handled with conceptual reasoning. In that case, the equations appear to have provoked the students to spend time on unhelpful calculations instead of thinking about the problem. Students calculating without thinking has been observed many times before, and has been attributed in part to a lack of metalevel skills.<sup>24</sup>

Given that students do perform differently on different assigned representations of problems, our second goal was to determine why. The data suggest that performance on different representations depends on a number of things, including student expectations, prior knowledge, metacognitive skills, and the specific contextual features of the problems and the representations. This dependence on specific, microlevel features also seems to be responsible for the reduced performance of some 202 students that made use of multiple representations, as compared to students that did not. It may also be that different problem representations are prompting different solution strategies, as Koedinger and Nathan<sup>1</sup> have observed in young algebra students. The strategies of our students cannot be consistently inferred from the data presented here, and so we have interviewed students in depth as they solve these sorts of problems. The results of these interviews will be part of a later paper.

Our third goal was to determine whether allowing students to choose which representation they worked in would

have an effect on their performance. The data show that giving them this choice for a quiz did indeed result in performance differences as compared to the random-format students; however the direction of that effect turned out to be inconsistent. In some cases, students given a choice of representation did much better than the students that were assigned a format at random; in other cases, they did much worse. Furthermore, whether the choice group did better or worse than the control group for a particular format sometimes varied from one quiz topic to the next, as was the case for both the graphical and pictorial groups in the 202 section. This could possibly be explained by the movement of a group of students that is good at choosing from one format to another, but analysis of the students that switch formats shows that this is not the case. Students that stayed with these formats did approximately as well on the second quiz as students that switched to these formats.

Finally, we hoped to explain the effect of student choice. To this end, we examined the students' comments. It appears that, correct or otherwise, students generally view mathematical and pictorial representations as dominant and opposite, at least out of the set of representations presented here. Most students selected one of these two formats. Students that switched from a mathematical format typically switched to a pictorial format. Student comments regarding their choice of quizzes frequently pitted mathematics against pictures, with one being favored versus the other. These same comments suggest that students connect pictorial representations quite strongly with "concepts," which students appear to view as unconnected to the mathematics. By itself, this characterization of student motives does not explain why student choice of representation had the effect that it did.

One could suppose that student choices and performances are guided by intrinsic learning styles. While it is the case that self-identified visual learners performed better on one of the pictorial format quizzes than other students (which is cause and which is effect is not clear here), the arrangement of the performance variations seen here suggests that the bulk of the data cannot be explained by a simple alignment of student choices with some individual learning style. For example, the students that chose pictorial-format quizzes for both the diffraction and spectroscopy topics had success rates of 0.86 and 0.33 on these quizzes. This is a dramatic difference since it appears that the spectroscopy quiz was easier overall, as the choice group had respective success rates of 0.48 and 0.70 averaged across all formats. A learning styles explanation would expect the same students to perform reasonably consistently on the same formats relative to the rest of the class. Considering the complexity of the performance data, it appears likely that an explanation of the effect of giving students a choice of representation will need to carefully attend to the context of the problem, much as we argued above regarding a complete explanation of student performance on different representations.

Another partial explanation of our data is suggested by the fact that the 201 students showed a much smaller performance difference between their choice and control groups than the 202 students. This may be a function of the broader, macrolevel features of the context, including the methods of instruction. As we described before, the 201 class included

more reforms and the features of the course may have provided students with a more varied set of representational skills. This could have leveled out students' performances on their preferred representations as compared to other representational formats. This might also explain the 201 students' much greater tendencies to remain with a format from quiz to quiz, as students with broader representational skills may be less likely to be dissatisfied with a particular representation. We should caution that this explanation is purely speculative at this point, and we note that in this study we cannot separate out the effect of instructional differences from the effect of content differences. The 201 and 202 courses are quite different in subject matter and in representational content. To make such an assertion, we would need to perform a repeat study that uses these quizzes and homeworks in a 202 course taught by the same instructor as this 201 course. Such a study will be the subject of a follow-up paper.

The current study has a number of limitations. First, as was just noted, the two courses studied differed in both subject area and instructional method, making cross-class comparisons difficult. Second, while the courses studied had a few hundred students each, it would still be desirable to replicate the study from year to year. Third, physics, including introductory physics, has a very rich collection of subtopics and associated representations. For the sake of this study, we have defined several representational categories, but such a definition cannot be unique or complete. Further, we have only examined a small selection of possible subtopics in this study, though this is perhaps not a severe limitation. One of our conclusions is that specific problems often will have features that are particular to the representation used that have

significant impacts on student performance. Thus, there are aspects of this study that we would not expect to be invariant across all subtopics. Finally, this study gives us fairly limited insight as to *how* the students solved these problems. Such insight will likely be necessary in order to better understand how and why representation affects performance. We plan to address this with a series of student interviews, which will be the subject of a future paper.

From the above, it appears that a complete understanding of student representational competence will need to attend to the specific and general features of the problems, the courses, and the learners. In this paper, we have taken a detailed look at student performance on specific problems. In the future, we will expand this look at the microlevel with a series of problem-solving interviews. We have also noted the possible effect of instructional method on representational performance. This macrolevel effect will be the subject of our follow-up paper, in which we directly compare the courses studied here with a 202 course taught by the reform-style instructor from the 201 course.

#### ACKNOWLEDGMENTS

This work was supported in part by NSF and by Colorado PhysTEC. Special thanks to the rest of the Physics Education Research group at the University of Colorado at Boulder. Thanks also to Dr. Beale, Dr. Munsat, and Dr. Peterson for their cooperation and aid.

#### APPENDIX: SUPPLEMENTARY AUXILIARY MATERIAL

See separate auxiliary material for the homework, quiz problems, and student reasons for selecting a representational format.

<sup>1</sup>K. R. Koedinger and M. J. Nathan, *J. Learn. Sci.* **13**, 129 (2004).

<sup>2</sup>N. Finkelstein, W. K. Adams, C. J. Keller, P. B. Kohl, K. K. Perkins, N. S. Podolefsky, S. Reid, and R. LeMaster, *Phys. Rev. ST Phys. Educ. Res.* **1**, 010103 (2005).

<sup>3</sup>D. E. Meltzer, *Am. J. Phys.* **73**, 463 (2005).

<sup>4</sup>A. Van Heuvelen and X. Zou, *Am. J. Phys.* **69**, 184 (2001).

<sup>5</sup>A. Van Heuvelen, *Am. J. Phys.* **59**, 891 (1991).

<sup>6</sup>R. J. Dufresne, W. J. Gerace, and W. J. Leonard, *Phys. Teach.* **35**, 270 (1997).

<sup>7</sup>R. J. Beichner, *Am. J. Phys.* **62**, 750 (1994).

<sup>8</sup>F. M. Goldberg and J. H. Anderson, *Phys. Teach.* **27**, 254 (1989).

<sup>9</sup>L. C. McDermott, M. L. Rosenquist, and E. H. Van Zee, *Am. J. Phys.* **55**, 503 (1987).

<sup>10</sup>S. E. Ainsworth, P. A. Bibby, and D. J. Wood, in *Learning with Multiple Representations*, edited by M. W. van Someren, P. Reimann, H. Boshuizen, and Ton de Jong (Pergamon, New York, 1998).

<sup>11</sup>R. E. Mayer, *Learn. Instr.* **13**, 125 (2003).

<sup>12</sup>W. Schnotz and M. Bannert, *Learn. Instr.* **13**, 141 (2003).

<sup>13</sup>R. B. Kozma and J. Russell, *J. Res. Sci. Teach.* **34**, 949 (1997).

<sup>14</sup>P. Heller, R. Keith, and S. Anderson, *Am. J. Phys.* **60**, 627 (1992).

<sup>15</sup>It is also worth noting that the boundaries between representational categories can be fuzzy. For example, pictorial represen-

tations may not have a distinct boundary from graphical representations; however, for this paper, we draw distinct boundaries. We delineate graphical as a representation that includes an independent axis (generally numerical), and a dependent quantity (generally these are Cartesian in nature). We classify pictorial as a literal or iconic depiction of an observable event, or model.

<sup>16</sup>A. A. diSessa and B. L. Sherin, *J. Math. Behav.* **19**, 385 (2000).

<sup>17</sup>A. A. diSessa, D. Hammer, B. L. Sherin, and T. Kolpakowski, *J. Math. Behav.* **10**, 117 (1991).

<sup>18</sup>A. A. diSessa, in *Symbolizing, Modeling and Tool Use in Mathematics Education*, edited by K. Gravemeijer, R. Lehrer, B. van Oers, and L. Verschaffel (Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003).

<sup>19</sup><http://www.lon-capa.org>

<sup>20</sup>M. Chi, T. H. Glaser, and E. Rees, in *Advances in the Psychology of Human Intelligence*, edited by R. J. Sternberg (Lawrence Erlbaum Associates, Hillsdale, NJ, 1982), Vol. 1.

<sup>21</sup>E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Upper Saddle River, NJ, 1996).

<sup>22</sup>E. F. Redish, *Teaching Physics with the Physics Suite* (Wiley, New York, 2003).

<sup>23</sup>M. Ward and J. Sweller, *Cogn. Instruct.* **7**, 1 (1990).

<sup>24</sup>A. Schoenfeld, in *Cognitive Science and Math Education*, edited by A. Schoenfeld (Erlbaum, Hillsdale, NJ, 1987).