

## Impact of a short intervention on novices' categorization criteria

Jennifer L. Docktor,<sup>1</sup> José P. Mestre,<sup>1,2,3,\*</sup> and Brian H. Ross<sup>1,4</sup>

<sup>1</sup>*Beckman Institute for Advanced Science and Technology, University of Illinois,  
405 North Mathews Avenue, Urbana, Illinois 61801, USA*

<sup>2</sup>*Department of Physics, University of Illinois, 1110 West Green Street, Urbana, Illinois 61801, USA*

<sup>3</sup>*Department of Educational Psychology, University of Illinois, 1310 South 6th Street, Champaign, Illinois 61820, USA*

<sup>4</sup>*Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, Illinois 61820, USA*

(Received 5 August 2011; published 25 July 2012)

Research on physics problem categorization has established that proficient problem solvers are able to group together physics problems that would be solved by similar principles and use conceptual approaches when solving problems, whereas weak solvers rely more heavily upon surface features (objects, contexts, and quantities provided) to identify specific equations that match to the problem situation. This study explores the degree to which novices are able to shift their categorization strategies toward one that is more principle based as a result of a brief, computer-based intervention designed to highlight the role of principles as categorization criteria. Students finishing an introductory algebra-based mechanics course were presented with a sequence of problem pairs, asked to judge whether each pair would be solved similarly, and provided with feedback. Students in one condition received feedback that was very sparse and only indicated correctness, whereas students in a second condition viewed elaborate feedback that linked problem features to the appropriate concepts and principles to solve each problem. We found an increased use of physics principles in the reasoning provided by students who received the elaborate feedback whereas students who did not view this elaboration primarily cited quantities in the problem statement. Although these results suggest it is possible to increase students' attention to principles when approaching problems, poor performance on the items indicates that considering the appropriateness of principles remains a difficult task for beginning physics students.

DOI: [10.1103/PhysRevSTPER.8.020102](https://doi.org/10.1103/PhysRevSTPER.8.020102)

PACS numbers: 01.40.Fk, 01.40.Ha

### I. INTRODUCTION

A major goal in physics instruction is to help students solve problems using a principled approach—that is, to consider what major principle or concept can be applied to problems [1–3]. However, extracting the major principle or concept from the problem's context, story line, and given quantities can be a very difficult task for beginning physics students. That is likely why novices gravitate toward more formulaic approaches to problem solving, in which they look for equations that match the quantities and unknowns in the problem and manipulate them to get answers [1,4–6]. Although this strategy is somewhat successful at getting answers (and often reasonable grades), it fails to highlight the role that principles and concepts play in solving problems, and does not promote deep conceptual understanding and long-term retention of the important ideas in physics. The study described here explores whether it is possible via a short intervention to shift the tendencies of novices from one that focuses on objects, symbols, and quantities in

problems as input for generating solutions toward one that uses a problem's context to extract the major principle or concept needed for solution; the study also explores whether the accuracy of principle-based categorizations improves as a result of either intervention. That is, at the end of an introductory physics course students have solved a lot of problems and applied many mathematical procedures, but little is known about whether or not it is possible, via a simple intervention, to help students link problem-solving procedures to conceptual knowledge. We begin with an overview of the role of categorization in problem solving and then describe the experiment and findings.

#### A. Overview of research on categorization

In our everyday encounters, we build organizational structures in memory to make sense of experiences and to help us access knowledge efficiently. This process of grouping together related entities is referred to as *categorization*, and perceived connections between concepts are constantly revised and updated in light of new information. Research on concepts and categories has a rich history in the field of psychology (see Ref. [7] or Ref. [8] for a review) including studies with a broad range of topics spanning domain-independent knowledge (e.g., faces, shapes, dot patterns) and domain-specific knowledge (e.g., food, birds, trees). Within the formal domain of

\*mestre@illinois.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

physics learning, the concepts and principles that govern how objects behave in a physical environment are relatively well defined and can be expressed in multiple representational formats, such as verbal descriptions and mathematical symbols (e.g., conservation of linear momentum and  $\vec{p}_{\text{initial}} = \vec{p}_{\text{final}}$ ). Beginning physics students are expected to learn fundamental concepts and to apply those concepts flexibly and appropriately when solving problems [9–12].

Categorization studies within physics seek to understand how populations that vary in their level of physics experience organize their knowledge about concepts applied during problem solving. The most prevalent method for studying how people form categories for physics problems is by using card-sorting tasks [13–21]. In most of these studies, textbook-style physics problems are printed onto index cards or listed on a piece of paper and study participants are asked to place the problems into groups based on solution similarity and to assign a name to each group.

A classic study by Chi, Feltovich, and Glaser [13] found that category labels used by introductory physics student volunteers (novices) were primarily based on the superficial features of the problems, whereas the categories created by advanced physics Ph.D. students (experts) included physics concepts and principles to a greater extent. The literal features cited by novices included a range of problem attributes: objects (inclined planes, springs, and pulleys), quantities (velocity and acceleration), motion (vertical motion, free fall), or other physics terms (friction, kinetic energy). Expert categories included derived, second-order features such as principles (second law, conservation of energy, work-energy theorem, and conservation of angular momentum), and other physics entities (circular motion, statics, linear kinematics, and vectors). The participants overlapped in their use of some category labels, including momentum principles, angular motion, work, and springs. The notion that experts and novices draw upon multiple criteria when grouping problems is consistent with other studies of categorization [17,22–24].

Subsequent studies have extended the findings of Chi *et al.* [13] to determine that while physicists are relatively consistent in their use of principles to group problems, students exhibit a greater range in their use of classification criteria. Some students use surface features and some use principles, suggesting that not all undergraduate students are “novices” [15,16,18,19,25] and not all graduate students can be classified as “experts” [20,21].

Equivalent results have been found in studies that replace card-sorting tasks with similarity judgments [23] or problem similarity ratings [26]. In these studies, participants were asked to judge whether problems would be solved similarly, or were asked to rate the similarity of a pair of problems on a Likert scale (e.g., from 0 to 5). Not surprisingly, introductory students rated problems that

match on surface features but do not share principles as more similar than faculty ratings for those same problems and they rated problems that share only deep structure as lower in similarity. This does not mean that students *cannot* identify principles for problems; rather, it suggests that they do not consider concepts spontaneously and are strongly influenced by surface features [27,28]. Interestingly, student explanations for their ratings suggest that in some cases they recognized that two problems could both be solved using conservation of energy, but the presence of specific kinds of energy (e.g., spring potential energy) that differed across a pair prompted them to rate the problem solutions as dissimilar [26].

Although card-sorting and similarity judgment studies have resulted in consistent findings about expert-novice differences, it is worth mentioning some methodological considerations associated with task design. First, sorting 20–30 problems into groups is a very mentally taxing process. Participants must consider all problems collectively which produces a high cognitive load [23,29]. Second, during sorting the items and item groupings are mutually dependent. Initial groups and category names have a strong influence on the placement of future items; it is difficult and rare for participants to modify their categories once they have formed an initial set and placed some problems into those categories. In contrast, deciding whether or not pairs of problems would be solved similarly gives a high number of independent assessments and allows one to carefully design problems to isolate particular features for investigation. Separate problem comparisons also lend themselves more readily to instructional interventions because the results are straightforward to interpret.

## B. Role of categorization in problem solving

Problem solving is a complex process in which a solver must first read and interpret a problem statement to form an internal *representation* of the problem and then *search* available resources (declarative knowledge, procedural knowledge, strategies, and equations) to decide upon viable solution approaches [4,30,31]. The search process is aided by having knowledge organized around problem “types” based on fundamental physics principles or concepts; once the problem type is identified it is relatively simple, at least for skilled problem solvers, to apply procedures for solving it [1,9,11–13]. The presence of these conceptual *schemas* is considered a fundamental difference between expert and novice problem solvers [29]. Studies indicate that the ability to categorize problems according to the principles or concepts needed for solution and problem-solving performance are positively correlated—better categorizers are better problem solvers and vice versa [15,23]. Clearly, correlation does not imply causation, but problem categorization is assumed to be an important part of problem solving in most extant theories.

Further, it is unreasonable to claim that experts are oblivious to superficial features altogether; it is instead feasible that particular features act as clues that an expert uses to hypothesize and select appropriate principles [24]. For example, an expert might indicate that a change in vertical distance (height) of an object and a compressed spring are clues for using conservation of energy [13]. Novices, who are less skilled in categorization, may require explicit guidance in order to learn relevant clues and how to judge the appropriateness of principles for a particular problem situation [21–24]. There is some evidence suggesting that it is possible to train students to consider principles or concepts when generating problem solutions using techniques such as hierarchical analysis [32,33] and writing qualitative strategies [34]. These studies indicate that after problem-solving training lasting at least several hours during which the role of principles and concepts is emphasized, students become better at selecting the appropriate principle(s) needed to solve problems.

### C. Motivation for this study

Given the important role that ability to categorize problems plays in competent problem solving, we examine what impact a short training session has on principle-based problem categorization. In the study reported here, students who finished an algebra-based mechanics course were given a short training session (less than one hour) consisting of individual trials during which they were asked to determine whether or not pairs of problems would be solved with a similar approach, with feedback provided after each trial. Two types of feedback were explored; one simply told the student after each trial whether their decision was correct. This very *sparse* feedback condition explored whether exposure to pairwise categorization of problems followed by correct or wrong feedback was sufficient to allow novices who had to take the requisite course material to discover or induce criteria for problem solutions (i.e., principles or concepts). The other type of feedback was more elaborate, with the student being told whether or not their decision was correct and, in addition, being provided with an explanation justifying the principle needed to solve the problem based on examination of the problems' context. This *elaborate* feedback more closely modeled for students what experts cue on when deciding on a solution strategy and explored whether explicitly showing students how a problem's context is used to extract the major principle needed for solution is more effective than sparse feedback in promoting principle-based categorization.

At various points during training, students were asked to explain the criteria they were using to make categorization decisions. Thus, we were able to track categorization performance and the criteria that students used in making categorization decisions as a function of type of feedback as well as changes during the training.

Any among the possible outcomes from this experiment would prove informative. One possibility is that students improve their categorization performance as a result of simple right or wrong feedback by reflecting on what features of problems can be used to extract the solution strategy, or perhaps more elaborate feedback cuing students on how surface attributes and story lines of problems would allow them to better determine solution strategies. A finding that one (or both) of the feedback conditions improves categorization performance would have important pedagogical implications. Traditional introductory physics courses devote considerable time toward helping students develop problem-solving skills, but the types of problem-solving skills that students adopt (equation-focused approaches) differ from those that skilled problem solvers use (identifying relevant principles and then applying them via equations). A positive finding would suggest that interventions could be designed to help students develop skilled problem-solving habits without a major restructuring of instructional practices. For example, problem-solving instruction could be supplemented with categorization exercises to alert students to the usefulness of principle-based problem-solving approaches.

On the other hand, given prior findings that categorization according to principles is a difficult, slowly developing index of expertise (even among graduate students), it would be optimistic to expect major changes in categorization *performance* based on less than one hour of training. Perhaps more interesting is the possibility that students might change the *criteria* they use to make categorization decisions as a result of training. If it is possible to get beginning students to think about the underlying structure of problems, and indeed the structure of physics, with a simple categorization exercise, then long-term intervention strategies might be developed around categorization tasks to help students develop conceptual understanding and competent approaches to solving problems by first helping students *change* the approach they use to solve problems (i.e., to attempt a principle-based approach) and then helping them *refine* the approach (e.g., by helping them make accurate categorization decisions and by linking general problem-solving procedures to principles) [35–37].

## II. METHOD

### A. Participants

The study participants were 26 students enrolled in an introductory algebra-based mechanics course at the University of Illinois (10 males and 16 females). Volunteers were solicited by a class-wide Email and paid for their time. Students were scheduled during the last two weeks of the Spring 2010 semester (before their final exam) to ensure that the relevant physics topics had been sufficiently covered in the course.

## B. Materials

Problems were adapted from a variety of sources, including physics problems used in past studies on categorization [23,26] and by modifying typical problems found in introductory physics textbooks not used in the course [38–40]. Problems were paired together to match in one of four ways: on *surface* features (objects) only, on *principles* only, on *both* surface features and principles, or *neither* surface features nor principles [23,26].

There were 32 items common to all participants, equally distributed across the 8 topics of Newton's second law, Newton's second law with centripetal acceleration, work-energy theorem, conservation of energy, impulse-momentum theorem, conservation of momentum, angular impulse angular momentum theorem, and conservation of angular momentum. The 8 problem topics were crossed with the 4 match types (see the table in the supplemental material). An additional bank of 16 problem pairs was constructed for use during initial and final phases of the experiment.

All problems were designed to be solved most efficiently by a single principle. The pairs were structured such that a “theorem” topic was never paired with its corresponding conservation law; for example, an impulse-momentum theorem problem was never paired with a conservation of momentum problem. However, there were items that paired one problem solved using Newton's second law with linear acceleration with a problem solved using Newton's second law with centripetal acceleration (with the expectation that an expert would consider these similar). The problems were written to constrain the applicable principle as much as possible, meaning the problems could only be solved by the application of a single major principle. Since impulse-momentum theorem and work-energy theorem problems often include quantities such as force, time, velocity, and distance, there is some overlap with Newton's second law and kinematics. Constraining these problems to a single theorem often meant explicitly stating the word “impulse” or “work” in the problem, which could be interpreted as a cue for the appropriate principle. This problem feature was taken into consideration when classifying students' responses, as will be explained in a later section.

Nearly all of the problems contained a diagram, and when one problem of a pair included a diagram the second problem did also. Only one problem in each group of eight problems did not include diagrams. Sometimes values for quantities were included on the diagram, but symbols were never provided because this could be interpreted as a cue for equations or principles. The text length of each problem was also approximately matched.

## C. Procedure

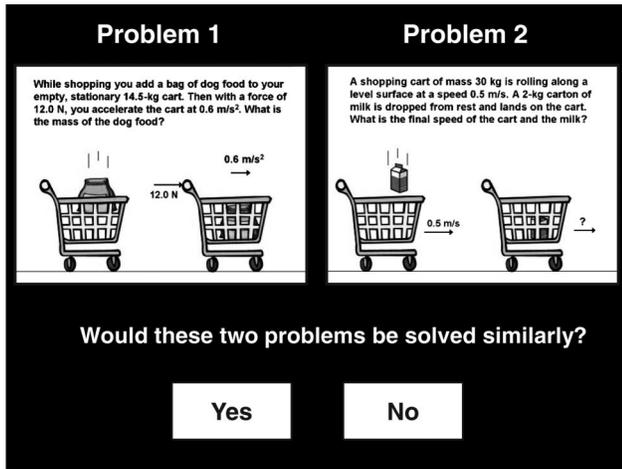
Participating students completed all portions of the experiment on a computer without intervention from the proctor. Instructions to participants are included in

the supplemental material. During the study, each participant was presented with a pair of physics problems side by side on a computer screen along with the text “Would these two problems be solved similarly?” and regions on the screen to click with a mouse for “yes” and “no” (see Fig. 1). Items were presented in a random order, and the two problems in a pair were always kept together (problem statements did not repeat). Immediately after making a selection, the student was presented with either a feedback display screen or a screen that prompted them to explain the reasoning for their choice by typing in a text box [Fig. 1(b)]. If prompted for reasoning, the feedback display screen was shown after completing a typed explanation. The problem statements remained on the screen for each of these portions—the initial decision, reasoning prompt, and feedback display. The reasoning prompt appeared 8 times during the session: after the first and second pair (drawn from the bank of extra problems to include one “yes” and one “no” question), 4 times during the middle 32 problem pairs (after problems 7, 15, 23, and 31), and after each of the final two items (also drawn from the bank of extra problems). In summary, each subject viewed 36 problem pairs and provided typed reasoning criteria for 8 of those items, which varied across participants. Each student had control over the pace of the session by pressing a key to advance to the next display screen, and the time spent viewing each display was recorded.

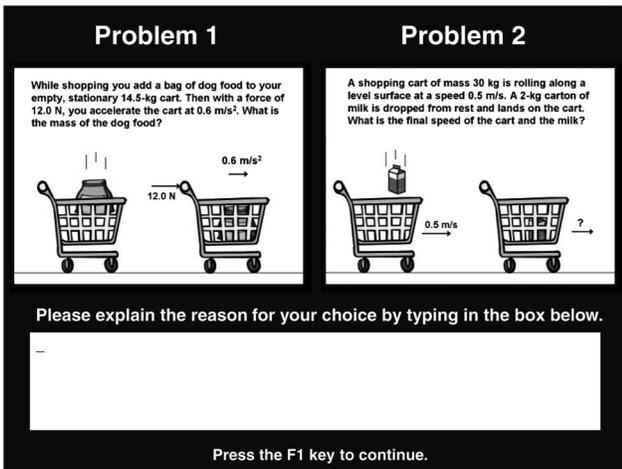
The participants were randomly assigned to one of two conditions corresponding to the type of feedback provided for responses. The conditions will be referred to as *sparse* feedback or *elaborate* feedback. In the sparse condition, the feedback was brief and always formatted in the following way, “(In)Correct. These problems would (NOT) be solved similarly.” In the elaborate condition, the feedback statement included detailed information about the physics principle(s) used to solve each problem [see Fig. 1(c) and the items in the supplemental material]. The elaborate feedback was constructed to explicitly link the quantities and conditions present in the problem to the appropriate concepts and principles, because this qualitative analysis of a problem statement is important but seldom made an explicit part of instruction [18].

In accordance with Ref. [23], a response was considered correct if it was consistent with appropriate principle-based processing. For problems that matched on *principles* or *both* surface features and principles, the appropriate response was “yes”; if they matched on *surface* only or *neither*, the appropriate response was “no.” When students chose correctly, the feedback screen displayed the word “Correct” in bright green, whereas an incorrect response prompted the word “Incorrect” in red.

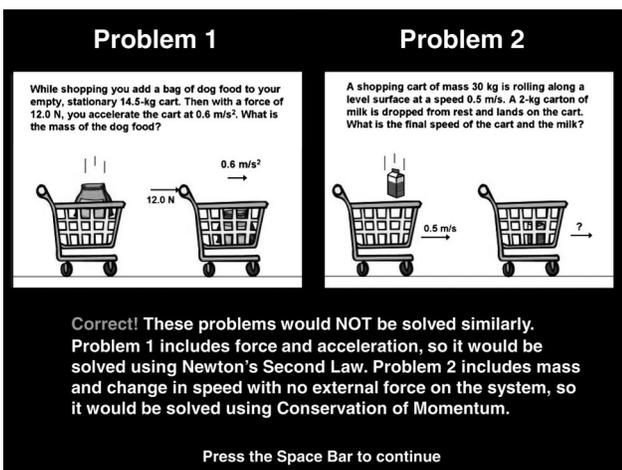
After viewing all 36 items, study participants were presented with a general reflective question, “Did your way of deciding change during the experiment? If so, how did it change?” with a text field in which to type their



(a)



(b)



(c)

FIG. 1. Sample (a) item screen, (b) reasoning prompt screen, and (c) feedback display viewed by participants in the elaborate condition. Participants in the sparse condition viewed only the first line of the feedback. These problems match on surface features (shopping cart) but would be solved using different physics principles.

response. Lastly, they were asked to rate their knowledge of physics on each of the 8 topics on a Likert scale from 1 to 5, “How much do you know about this physics concept:” where 1 was *nothing*, 3–4 was a *fair amount*, and 5 designated *a lot*.

### III. RESULTS AND DISCUSSION

#### A. Reasoning statements

##### 1. Statement classification scheme

The main result of interest is whether the feedback manipulations affected participants’ categorization performance and categorization criteria as measured by use of principles in reasoning about their responses. The reasoning statements that participants typed during the computer session were classified according to what entities were mentioned for the specific items. Previous research [13,23] indicated that participants were likely to cite principles or concepts, quantities, motion descriptors, or objects. When reasoning statements did not fit into one of these classifications a new label was created (e.g., vectors). The eight criteria are stated in Table I with representative examples.

The code for *quantity* primarily designated numerical values or the unknown (target) quantity cited in the problem, but also included terms such as force and impulse if they appeared in the problem statement and implicit quantities such as acceleration of gravity. The classification of *kinematics* or *dynamics* underwent careful consideration, because these terms are sometimes used by experts to describe a set of physics principles [13]. However, none of the items in this study was written to be solved exclusively by motion with constant acceleration (“kinematics”), and oftentimes it was unclear what students meant when they used these terms. As a result, these statements were assigned an independent code. The code for *equation* was kept separate from reasoning based on *principles* because the nature of these reasoning styles was considered qualitatively different. Collapsing them into a single category does not significantly change the results. The *other* code designated statements that were too vague to be classified or when students said they did not know.

Other studies of categorization typically make broad generalizations about participants’ criteria, such as comparing use of surface features to principles [13], model-based compared to theory-based features [17], or classifying them as “good” [21]. In this study an attempt was made to further specify the type of categorization decision made using these eight codes.

The statements were coded for each problem in a pair separately, and most participants addressed each problem independently in their explanations. For example, “*one deals with impulse, one deals with springs*” (subject 11, item no. 48). This was coded as a given *quantity* for the first problem, and as *object* reasoning for the second problem.

TABLE I. Classification scheme for participants' typed reasoning criteria.

Reasoning code	Description and examples
Principle	Principles: Newton's second law, conservation of energy, work-energy theorem, momentum principles, angular momentum principles; Concepts: force, kinetic energy, work
Quantity	Speed, angular speed, mass, height, friction (includes work, impulse, or force if this term appears in the problem); torque, centripetal acceleration, gravity, both problems have the same variables
Motion	(Motion type or direction): Circular motion, angular motion, rotation, centripetal, projectile motion, free fall, collision, simple harmonic motion, pendulum motion, linear, vertical, horizontal, parabolic
Equation	In both of these problems you can use the equation $PE_i + KE_i = PE_f + KE_f$ The first problem uses $F = ma$
Kinematics or dynamics	Kinematics problem, dynamics problem
Vectors	Need to use vector components, both have $x$ and $y$ parts, in two dimensions
Objects	Ramp, spring, string, pulley, roller coaster, car
Other	The problems deal with different things I don't really know

When students referred to both problems in a single statement, the entities were coded twice, once for problem 1 and once for problem 2, "*In both cases there is an  $x$  and  $y$  component to the velocity as the bird is flying, therefore the problems must be solved using vectors*" (subject 4, item no. 3). When the statement only referred to one problem, the second problem was coded as *other*, "*first one momentum and second one something else*" (subject 20, elaborate, item no. 30).

It was also possible for a statement to include more than one type of entity for a single problem, such as including both *motion* and a *quantity*, or *object* and *quantity*; these were designated with multiple codes. When mixed criteria were used, the proportion per problem per subject was split across the categories stated. For example, "*Both problems involve similar set-ups: a string attached at one end. Both ask for the final angular speed.*" (subject 17, item no. 45).

The statements were coded independently by two researchers who were blind to the condition that each statement came from. The coders agreed exactly on 90% of the codings. Each instance of a disagreement was examined for consistency with the rubric, and when disagreements still remained (4% of the statements), the code scores were split across the categories cited by each researcher.

Whenever a statement was principle based, the appropriateness of the principle cited was recorded for separate analysis. The appropriateness was rated using a rubric on a 5-point scale from 0 to 1, and was assessed separately for each problem. In general, a score of 0 or 0.25 meant an inappropriate principle was cited, 0.5 indicated a vague concept (e.g., stating "energy" without specifying the work-energy theorem or conservation of energy), and a score of 1.0 designated an appropriate and complete principle-based statement.

## 2. Prevalence of reasoning criteria

Each of the 26 participants provided reasoning for 8 items (16 problems); however, the first two statements

have been excluded because the feedback manipulation had not yet occurred. That gives a total of 312 codes or 156 per group. The proportion of statements that mentioned a particular entity for each group is reported in Table II. The table also indicates the prevalence of categorization criteria from the table of novice categories in the study by Chi *et al.* (see Ref. [13], p. 129). Category labels from that study were classified according to the current coding scheme: the categories springs, inclined planes, and pulleys were coded as *objects*; vertical motion and free fall were *motion*; *quantities* included velocity and acceleration, friction, and center of mass; and *principles* included kinetic energy, momentum principles, and work. The category "angular motion" included several different criteria, so it was split to include *motion* (angular motion), *quantities* (angular velocity, angular quantities, angular speed), and *principles* (angular momentum) in an estimated 1:3:1 ratio based on the number of terms listed in the label. The proportion value reported in the table is calculated from the number of problems accounted for out of a total of 192 problems (8 novices who each sorted 24 problems).

In the current study, the most stark distinction is the elaborate feedback group's higher use of *principle* criteria [ $t(24) = 3.46$ ,  $p < 0.01$ ] and lower use of *quantity* criteria [ $t(24) = 2.58$ ,  $p < 0.05$ ] compared to the sparse group. For students who received elaborate feedback that mentioned principles and the justification leading to those principles, their statements began to reflect a conceptual focus: "*They both deal (at least in part) with a frictionless area so conservation of energy can be used*" (subject 11, elaborate, item no. 19). "*Problem 1 deals with conservation of momentum while problem 2 uses Newton's second law*" (subject 2, elaborate, item no. 30). In some cases, the reasoning was not as descriptive, "*You can use the same principles to solve for the two different elements of the problems*" (subject 10, elaborate, item no. 7). In contrast, quantities and equations were cited more frequently in statements of students receiving sparse feedback,

TABLE II. Proportion of typed reasoning statements assigned each code for each condition, plus or minus the standard error of the proportion. Participants in the elaborate feedback condition cited principles to a greater extent whereas the sparse feedback condition relied on quantities for their reasoning criteria.

Reasoning code	Sparse feedback <sup>a</sup>	Elaborate feedback <sup>a</sup>	Novices in Ref. [5] <sup>b</sup>
Principle	0.16 ± 0.04	0.47 ± 0.08	0.13 ± 0.02
Quantity	0.48 ± 0.06	0.27 ± 0.06	0.27 ± 0.03
Motion	0.14 ± 0.04	0.09 ± 0.02	0.09 ± 0.02
Equation	0.07 ± 0.03	0.00 ± 0.00	...
Kinematics or dynamics	0.07 ± 0.04	0.05 ± 0.02	...
Vectors	0.03 ± 0.01	0.00 ± 0.00	...
Objects	0.03 ± 0.01	0.02 ± 0.01	0.20 ± 0.03
Other	0.03 ± 0.02	0.09 ± 0.04	0.04 ± 0.01

<sup>a</sup>There were 13 participants in each group who provided reasoning on 6 items of 2 problems each, so the proportion is out of 156 statement code assignments (the first two statements have been excluded because they occurred before the feedback manipulation).

<sup>b</sup>In Ref. [5] there were 8 novices who classified 24 problems, so the proportion is out of 192 problem assignments.

“Problem 2 is finding velocity while Problem 1 is finding tension” (subject 16, sparse, item no. 33). “ $F = ma$  and  $m1v1 = m2v2$ ” (subject 21, sparse, item no. 2).

Subjects in both conditions mentioned descriptions of motion, particularly when objects were moving in a circle, “Both angular problems” (subject 25, sparse, item no. 4). “The first problem is a linear problem while the second is an angular rotation problem” (subject 26, elaborate, item no. 35). In addition, at least two students saw item no. 26 (a neither problem) as being similar because of a parabolic shape in the picture.

One interesting observation is that when the problem included different force terms or a different orientation to the forces, some students commented that they were different problems, which is similar to the finding for energy terms in Mateycik *et al.* [26]. For example, “Both problems deal with rotational kinematics but the first problem looks at normal force. The second problem looks at force of friction” (subject 23, elaborate, item no. 40). “Problem 1 has two forces acting on pulling it straight up, problem 2 has only one and is moving horizontally” (subject 13, sparse, item no. 25).

Objects were rarely mentioned (3% of the time in the sparse condition and 2% in the elaborate condition), but when they were, they usually referred to springs (items no. 1, 33, 48) or a ramp (item no. 29), “one is a rotational problem and the other is a ramp problem” (subject 8, sparse, item no. 29). “Problem 1 deals with impulses and momentum. Problem 2 deals with springs” (subject 4, sparse, item no. 48). This differs from the results reported in Ref. [5], where novice participants placed approximately 20% of the problems into object categories.

There is a perception within the physics problem-solving research community that novices focus on the literal *objects* in problems, when in fact the study by Chi *et al.* [13] suggests they use multiple criteria that include

objects, quantities, motion descriptors, and even sometimes principles. Our findings differ from these previous findings in that they indicate that beginning physics students use quantities in the problems much more frequently than objects. There are a few key aspects of the methodology which may have influenced students’ criteria. First, the particular problems differed from previous studies (for example, there were few blocks and inclined planes.) Second, the two-problem categorization task itself was fundamentally different from card sorting. In addition, the coding scheme for students’ decision criteria is more detailed than previous studies on categorization [13,18,21,23] because it divides the criteria of “surface features” into multiple categories. These differences in coding schemes make it difficult to compare across studies, although an attempt was made to do so here (see Table II). Note the sparse condition and the Chi *et al.* results are similar if one collapses the quantity and object categories.

### 3. Appropriateness of principle-based reasoning

The previous section reported the prevalence of each group’s use of principles as reasoning criteria regardless of whether the principles cited were appropriate for the items. It is possible that participants could focus on identifying principles for use in solving the problems but be unable to identify appropriate principles. Although participants receiving elaborate feedback cite principles significantly more, there are no differences in the average appropriateness rating for those principles ( $M = 0.68$ ,  $SD = 0.24$  for 13 students who cited principles) compared to the sparse condition ( $M = 0.67$ ,  $SD = 0.20$  for 8 students who cited principles). In addition, when students used principles as reasoning criteria, both groups cited appropriate principles for both problems in a pair approximately half of the time.

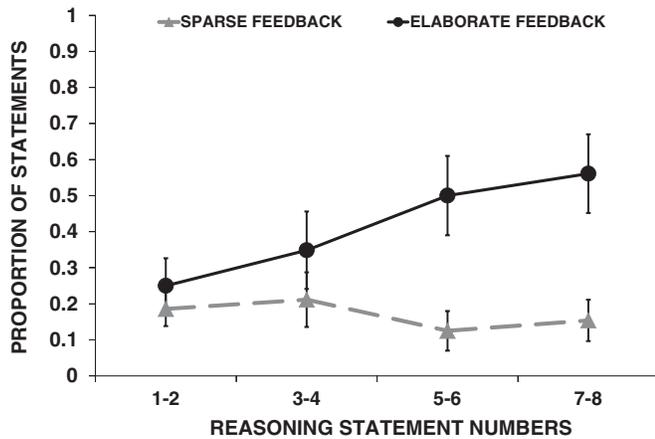


FIG. 2. Proportion of statements coded as principle based over time. Each point represents 13 subjects' statements averaged for two items consisting of two problems (52 codes). Participants in the elaborate group increased their use of concepts and principles significantly as the session progressed.

Our findings are consistent with others who report that ability to categorize physics problems according to principles is difficult [21,23]. Even when participants in this study attempted to use principles or concepts in judging the similarity of problems, they frequently cited principles that were inappropriate for a particular problem.

#### 4. Progression of principle-based reasoning over time

As the session progressed, the proportion of statements that were coded as *principle* for the elaborate condition increased, yet they decreased slightly for the sparse condition. The most significant split happened at the midpoint of the session (see Fig. 2).

#### B. Characterization of overall process

At the conclusion of the experiment, participants were presented with the following question and prompted to type a response, "Did your way of deciding change during the experiment? If so, how did it change?" The purpose of this question was twofold: to elucidate students' self-reported general approach to engaging in the similarity judgment tasks, and as an additional indication of how the feedback influenced this approach. As expected, in the elaborate feedback condition there was a substantially higher proportion of students (9 of 13) who said that their reasoning approach changed. In the sparse feedback condition, only 4 of the 13 participants said their way of deciding changed.<sup>1</sup> There were also qualitative differences in students' descriptions of their decision process. Students in the sparse condition primarily said they looked at quantities and equations:

<sup>1</sup>A one-tailed Fisher exact probability test on these values gives  $p = 0.058$ .

"My way of deciding didn't really change throughout the experiment. I would just look at what was given and what I was looking for" (subject 25, sparse condition).

"Not really. I just tried to think of the equations for both" (subject 21, sparse).

"Yes. Over time I looked first at what the problem was asking for. Then I looked at what was given to me to find that. Then I thought of all the physics set-ups I know in relation to that concept. Although some problems involved similar physics, the problems either asked for different things (thus, need different equation), or the set-up is different between the two" (subject 17, sparse).

In contrast, several participants in the elaborate condition recognized that their reasoning shifted in response to the feedback provided:

"Yes. As I saw the answers and the reasoning behind the answers I was able to answer the later questions using similar terminology. I also learned what to look for in the question to determine what specific theorem was being used" (subject 2, elaborate).

"Yes, it did. At first, I only looked to see if the problems had similar situations. That did not help so I looked to see if the problem were compared based on the type of answer wanted (such as speed, or force or acceleration, etc.). Then I found that the easiest way is to figure out which theorem the problems deal with" (subject 23, elaborate).

This was not true of all elaborate condition participants, however:

"no, my way does not change. I was trying to figure out the 2 problems for all questions" (subject 3, elaborate).

Responses also indicate that some students in the elaborate condition learned to associate particular features or cues from the problem statement with concepts or principles:

"Yes. It changed because I feel like I figured out a pattern...Whenever it said "impulse" the problem was about impulse-momentum, whenever it had speeds and coefficients of friction, it was newtons second law, and so on" (subject 11, elaborate).

"Yes, first I was deciding if they were similar based on what equation to use, then I realized it was asking for you to look at the given variables and think of the concepts behind solving the equations" (subject 20, elaborate).

In both conditions there were participants who mentioned being cautious of surface feature similarities, "I definitely altered my thinking to not judge simply by the presence of the same variables, or very similar pictures" (subject 9, sparse), and "I also became more suspicious of deceptively similar images" (subject 10, elaborate).

At the conclusion of the experiment, participants were also asked to rate how much they know or remember about each mechanics concept on a Likert scale. There were no significant differences between the two groups, although the elaborate subjects rated their knowledge slightly lower (3.7 for elaborate versus 3.9 for sparse out of a possible 5).

Topics that were rated lowest were angular impulse angular momentum theorem and conservation of angular momentum, and topics rated highest were conservation of momentum and conservation of energy.

### C. Accuracy of judgments

The average scores on each problem pair type are reported in the supplemental material. There was not a significant difference in the overall average score for the sparse ( $M = 0.67$ ,  $SD = 0.12$ ) and the elaborate ( $M = 0.63$ ,  $SD = 0.07$ ) conditions;  $t(24) = 1.01$ . Note that performance on items matching on only *surface* features or sharing only *principles* was considerably lower than items matching on *both* or *neither*.

### D. Response time

Although each participant was scheduled for an hour-long session, most participants did not require that much time to finish. The sparse condition participants completed the session in an average of 29 minutes and 38 seconds, whereas the elaborate condition participants took an average of 31 minutes and 51 seconds. It was expected that the elaborate condition would take somewhat longer because of the longer feedback explanations. The average time spent viewing each type of screen (problem pair, reasoning prompt, and feedback display) is reported in Table III.

On average, the participants in the sparse condition spent 30 seconds viewing an item before selecting “yes” or “no,” and the elaborate condition spent a couple seconds longer. This did not vary substantially across the different types of problem pairs (surface, principles, both, and neither). As expected, the feedback in the elaborate condition was longer and took more time to read—approximately two and a half seconds longer than the sparse feedback condition. The short time viewing the feedback screen suggests that participants did not spend substantial (if any) time reflecting on the correctness of their answer before moving on to the next item. However,

TABLE III. Average time spent (in seconds) viewing each screen: reading the problem pair and making a decision (stimulus), typing in an explanation in the reasoning prompt screen, and viewing the feedback display screen. Participants in both groups spent slightly more time viewing the feedback screen when they answered an item incorrectly.

Display screen	Sparse feedback	Elaborate feedback	$p$
	Average time per item (s)	Average time per item (s)	
Stimulus	$30.3 \pm 0.8$	$31.9 \pm 1.0$	0.201
Reasoning	$48.8 \pm 3.8$	$38.0 \pm 3.1$	0.028
Feedback	$1.6 \pm 0.1$	$4.1 \pm 0.2$	<0.001
Correct	$1.4 \pm 0.1$	$3.2 \pm 0.2$	<0.001
Incorrect	$2.0 \pm 0.1$	$5.8 \pm 0.3$	<0.001

participants in both conditions viewed the feedback screen for slightly longer when they answered an item incorrectly.

When typing in the reasoning for their choice, the participants in the sparse feedback condition spent about 10 seconds longer to formulate a response than the elaborate condition. It is possible that the participants in the elaborate condition modeled their responses after the elaborate feedback they viewed, and therefore spent less time deciding how to respond. It is also possible that the groups differed in the length of their typed descriptions. A count of the average number of characters typed by each participant per item indicates that the sparse condition had slightly longer responses ( $M = 101.9$  characters,  $SD = 47.7$ ) than the elaborate condition ( $M = 74.7$  characters,  $SD = 21.0$ ); however, this difference is not significant  $t(24) = 1.88$ ,  $p < 0.10$ .

## IV. GENERAL DISCUSSION

The main goal of this study was to examine whether it is possible to impact categorization criteria, categorization performance, or both, as a result of a brief computer-based intervention. The intervention focused on having students compare pairs of physics problems, which were carefully constructed to match on surface attributes only, on principle or concept only, on both, or on neither, and to state whether the problems would be solved with a similar approach. Students were given feedback on the correctness of their responses, with one group (sparse feedback condition) receiving only right or wrong feedback, and a second group (elaborate feedback condition) in addition receiving feedback on how a problem’s context is used to extract the principle or concept needed for solution.

The most interesting finding of this study, as well as the most relevant in terms of instructional implications, is that a short intervention can help students to increase their tendency to use principle-based categorizations. Even an intervention lasting less than one hour that highlights the role of principles or concepts for deciding on solution similarity (i.e., the elaborate feedback condition) can prove influential in helping students *shift* their focus from quantities to principles or concepts in making categorization decisions. Students in the elaborate feedback condition dramatically shifted the criteria they used in making categorization decisions about halfway through the training (see Fig. 2) toward one that relied more on principles. This is an important result given that novices usually show a strong reliance on thinking in terms of objects, quantities, and equations, not principles. In fact, students receiving sparse feedback in this study generally continued to use surface attributes (quantities and terms) in the problem statements to make categorization decisions, suggesting that simple feedback on correctness is not enough—novices need explicit help in using a problem’s context to extract the principles needed for solution.

Despite the shifts in categorization criteria by the elaborate feedback group, there were no differences in overall categorization performance between the two feedback conditions. Given that previous categorization studies in physics [21,23] consistently reveal that making *accurate* categorization decisions based on principles or concepts is difficult for both novices and graduate students, it would have been very optimistic to expect major shifts in categorization *performance* as a result of a short training session highlighting how principles are used to make categorization decisions. However, before students can begin to improve the accuracy of their principle-based categorization decisions, they first must begin to use principles in making categorization decisions, and our findings suggest that getting students to consider principles in making categorization decisions is not difficult with explicit training.

Finally, it is important to discuss some limitations of this study. Although the brief computer-based intervention resulted in a shift in students' categorization criteria in the elaborate feedback condition, we have no information about the long-term implications of this finding. For example, is improving accuracy in principle-based categorization a long-term process that is difficult to achieve among novices even after they display willingness to shift categorization criteria or is it achievable within an introductory course with the appropriate scaffolding? Although we have shown that a shift in categorization criteria is achievable with a brief intervention, is such a shift ephemeral without continued support and reinforcement? Even with positive answers to these questions, research will need to determine how to structure the appropriate scaffolding and maintain it.

## V. INSTRUCTIONAL IMPLICATIONS

When left to their own devices, students enrolled in introductory physics courses largely prefer equation-focused quantitative approaches to solving problems at the expense of conceptual development. However, it has long been known that experts categorize problems based on the underlying principles and then use these principles

to solve the problems. A central question of physics education research is how to get novices to focus on the principles, preferably without a major change in the established curriculum. Some physics education research studies have also shown that it is possible to improve categorization performance and problem-solving skills among novices with interventions lasting from several hours [32,33] to a whole semester [34] that highlight the role played by principles in solving problems. Although we believe such conceptual-based problem-solving approaches may help novices to overcome their focus on equations, they require considerable investment in time and require students to learn a difficult approach before they understand fully the benefits of the approach. The study described herein takes a simpler quick-feedback approach to try to first get students to understand that problems can be categorized in terms of principles. The results suggest that exercises could be devised and administered via computer (e.g., using today's popular Web-based homework delivery systems) to help novices identify the principle(s) needed to solve problems. One might imagine a set of short computer-based feedback exercises throughout the semester (that could be done outside the classroom) that first focuses learners on thinking about principles, then provides periodic reinforcement of this idea while at the same time giving them practice on conceptual analysis by incorporating newly learned principles into the exercises.

## ACKNOWLEDGMENTS

Work in part supported by the Institute of Education Sciences of the U.S. Department of Education under Grant No. DE R305B070085. We thank Jeanne Brunner and Rachel Aschenbrenner for assisting with coding reliability.

## SUPPLEMENTAL MATERIAL

See separate supplemental material for the study instructions, physics problems, and feedback text.

- 
- [1] J. Larkin and F. Reif, Understanding and teaching problem solving in physics, *Int. J. Sci. Educ.* **1**, 191 (1979).
  - [2] A. Van Heuvelen, Learning to think like a physicist: A review of research-based instructional strategies, *Am. J. Phys.* **59**, 891 (1991).
  - [3] E. Yerushalmi, C. Henderson, K. Heller, P. Heller, and V. Kuo, Physics faculty beliefs and values about the teaching and learning of problem solving. I. Mapping the common core, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020109 (2007).
  - [4] J. H. Larkin, in *Mental Models*, edited by D. Gentner and A. L. Stevens (Lawrence Erlbaum, Hillsdale, NJ, 1983), p. 75.
  - [5] J. Tuminaro and E. F. Redish, Elements of a cognitive model of physics problem solving: Epistemic games, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020101 (2007).
  - [6] L. N. Walsh, R. G. Howard, and B. Bowe, Phenomenographic study of students' problem solving approaches in physics, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020108 (2007).

- [7] D.L. Medin, Concepts and conceptual structure, *Am. Psychol.* **44**, 1469 (1989).
- [8] D.L. Medin and L.J. Rips, in *The Cambridge Handbook of Thinking in Reasoning*, edited by K.J. Holyoak and R.G. Morrison (Cambridge University Press, New York, 2005), p. 37.
- [9] B.-S. Eylon and F. Reif, Effects of knowledge organization on task performance, *Cogn. Instr.* **1**, 5 (1984).
- [10] F. Reif, Teaching problem solving—A scientific approach, *Phys. Teach.* **19**, 310 (1981).
- [11] F. Reif and J. Heller, Knowledge structure and problem solving in physics, *Educ. Psychol.* **17**, 102 (1982).
- [12] R. Zajchowski and J. Martin, Differences in the problem solving of stronger and weaker novices in physics: Knowledge, strategies, or knowledge structure, *J. Res. Sci. Teach.* **30**, 459 (1993).
- [13] M.T.H. Chi, P.J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, *Cogn. Sci.* **5**, 121 (1981).
- [14] A.H. Schoenfeld and D.J. Herrmann, Problem perception and knowledge structure in expert novice mathematical problem solvers, *J. Exp. Psychol. Learn. Mem. Cogn.* **8**, 484 (1982).
- [15] T. de Jong and M.G.M. Ferguson-Hessler, Cognitive structures of good and poor novice problem solvers in physics, *J. Educ. Psychol.* **78**, 279 (1986).
- [16] E. Savelsbergh, T. de Jong, and M.G.M. Ferguson-Hessler, Choosing the right solution approach: The crucial role of situational knowledge in electricity and magnetism, *Phys. Rev. ST Phys. Educ. Res.* **7**, 010103 (2011).
- [17] J.L. Snyder, An investigation of the knowledge structures of experts, intermediates, and novices in physics, *Int. J. Sci. Educ.* **22**, 979 (2000).
- [18] G.H. Veldhuis, Differences in the categorization of physics problems by novices and experts, Ph.D. thesis, Iowa State University, 1986.
- [19] G.H. Veldhuis, The use of cluster analysis in categorization of physics problems, *Sci. Educ.* **74**, 105 (1990).
- [20] S.-Y. Lin and C. Singh, Categorization of quantum mechanics problems by professors and students, *Eur. J. Phys.* **31**, 57 (2010).
- [21] C. Singh, Categorization of problems to assess and improve proficiency as teachers and learners, *Am. J. Phys.* **77**, 73 (2009).
- [22] S.B. Blessing and B.H. Ross, Content effects in problem categorization and problem solving, *J. Exp. Psychol. Learn. Mem. Cogn.* **22**, 792 (1996).
- [23] P.T. Hardiman, R. Dufresne, and J.P. Mestre, The relation between problem categorization and problem solving among experts and novices, *Mem. Cogn.* **17**, 627 (1989).
- [24] T.L. Spalding and B.H. Ross, Concept learning and feature interpretation, *Mem. Cogn.* **28**, 439 (2000).
- [25] T. de Jong and M.G.M. Ferguson-Hessler, Knowledge of problem situations in physics: A comparison of good and poor novice problem solvers, *Learn. Instr.* **1**, 289 (1991).
- [26] F. Mateycik, N.S. Rebello, and D. Jonassen, Students' rating of problem similarity as a measure of problem solving expertise, in *Proceedings of the 2010 Annual International Conference of the National Association for Research in Science Teaching (NARST), Philadelphia, PA*, <http://www.narst.org/annualconference/postconference10/additionalpapers.cfm>.
- [27] L.R. Novick, Analogical transfer, problem similarity, and expertise, *J. Exp. Psychol. Learn. Mem. Cogn.* **14**, 510 (1988).
- [28] D.D. Cummins, Role of analogical reasoning in the induction of problem categories, *J. Exp. Psychol. Learn. Mem. Cogn.* **18**, 1103 (1992).
- [29] J. Sweller, Cognitive load during problem solving: Effects on learning, *Cogn. Sci.* **12**, 257 (1988).
- [30] A. Newell and H.A. Simon, *Human Problem Solving* (Prentice-Hall, Englewood Cliffs, NJ, 1972).
- [31] M.L. Gick, Problem-solving strategies, *Educ. Psychol.* **21**, 99 (1986).
- [32] R.J. Dufresne, W.J. Gerace, P.T. Hardiman, and J.P. Mestre, Constraining novices to perform expertlike problem analyses: Effects on schema acquisition, *J. Learn. Sci.* **2**, 307 (1992).
- [33] J.P. Mestre, R.J. Dufresne, W.J. Gerace, and P.T. Hardiman, Promoting skilled problem-solving behavior among beginning physics students, *J. Res. Sci. Teach.* **30**, 303 (1993).
- [34] W.J. Leonard, R.J. Dufresne, and J.P. Mestre, Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems, *Am. J. Phys.* **64**, 1495 (1996).
- [35] D.M. Bunce, D.L. Gabel, and J.V. Samuel, Enhancing chemistry problem-solving achievement using problem categorization, *J. Res. Sci. Teach.* **28**, 505 (1991).
- [36] F. Reif, Instructional design, cognition, and technology: Applications to the teaching of scientific concepts, *J. Res. Sci. Teach.* **24**, 309 (1987).
- [37] V.L. Willson, Methodological limitations for the use of expert systems techniques in science education research, *J. Res. Sci. Teach.* **27**, 69 (1990).
- [38] D.C. Giancoli, *Physics: Principles with Applications* (Prentice Hall, Upper Saddle River, NJ, 1998), 5th ed.
- [39] K. Cummings, P. Laws, E. Redish, and P. Cooney, *Understanding Physics* (John Wiley & Sons, Hoboken, NJ, 2004).
- [40] J.S. Walker, *Physics* (Prentice Hall, Upper Saddle River, NJ, 2002).