Problem-solving rubrics revisited: Attending to the blending of informal conceptual and formal mathematical reasoning

Michael M. Hull,[1] Eric Kuo,[1] Ayush Gupta,[1] and Andrew Elby[2]

[1]*Department of Physics, University of Maryland, College Park, Maryland 20742, USA*
[2]*Department of Teaching, Learning, Policy, and Leadership, University of Maryland, College Park, Maryland 20737, USA*

Much research in engineering and physics education has focused on improving students' problem-solving skills. This research has led to the development of step-by-step problem-solving strategies and grading rubrics to assess a student's expertise in solving problems using these strategies. These rubrics value ''communication'' between the student's qualitative description of the physical situation and the student's formal mathematical descriptions (usually equations) at two points: when initially setting up the equations, and when evaluating the final mathematical answer for meaning and plausibility. We argue that (i) neither the rubrics nor the associated problem-solving strategies explicitly value this kind of communication *during* mathematical manipulations of the chosen equations, and (ii) such communication is an aspect of problem-solving expertise. To make this argument, we present a case study of two students, Alex and Pat, solving the same kinematics problem in clinical interviews. We argue that Pat's solution, which connects manipulation of equations to their physical interpretation, is more expertlike than Alex's solution, which uses equations more algorithmically. We then show that the types of problem-solving rubrics currently available do not discriminate between these two types of solutions. We conclude that problem-solving rubrics should be revised or repurposed to more accurately assess problem-solving expertise.

PACS numbers: 01.40.Fk, 01.40.Di, 01.40.G−, 01.40.gb

I. INTRODUCTION

''Alex'' and ''Pat,'' two students in an introductory calculus-based physics course, each thought aloud while solving the following problem (paraphrased for brevity): *Two balls are released from the same height at the same time, but one is released from rest while the other is thrown down at 2 meters per second. Five seconds later, what is the difference in speeds of the two balls?*

Alex started by drawing a diagram and listing the known values. She selected the equation $v = v_0 + at$ to calculate the speed of each ball after 5 s. She found those speeds to be 50 and 52 m/s and subtracted to find the difference in speeds, 2 m/s. Alex's methodical solution reflects a standard approach to problem solving taught in physics courses (see, e.g., Refs. [1,2]).

By contrast, in his answer, Pat invoked $v = v_0 + at$ but did not perform algebraic manipulations or plug in numbers. Instead, he found a shortcut, abridged here but presented more completely later:

So the acceleration is a constant and that means that velocity is linearly related to time and they're both at the same... so the [slope] is the same... so the initial

conditions are off by 2 and then the velocities are changing at the same rate so that should mean they [the difference in speeds] stay at 2.

Notice how Pat's reasoning blends an intuitive conceptual idea—if two things change by the same amount, the difference between them stays the same—with mathematical formalism (a linear relation extracted from an equation, and slope) in a single, integrated line of reasoning.

Is Pat's solution more expert than Alex's? We argue that it is. What is more, we argue that existing problem-solving rubrics in the literature overlook this difference in expertise. To be clear, the rubrics (and associated problem-solving strategies) do emphasize connecting the physical scenario to a mathematical model when setting up the equations and when evaluating the results of the mathematical manipulations. And the rubrics do not *downgrade* Pat-style reasoning; a coder could give Pat credit for both appropriate conceptual reasoning and appropriate mathematical manipulations. Furthermore, we do not claim that the authors of such rubrics would fail to appreciate the quality of Pat's reasoning. Our claim is that the rubrics themselves do not recognize or reward blending of conceptual and formal mathematical reasoning *during* mathematical manipulations, even though such blending reflects expertise.

This is important because a problem-solving rubric helps direct the attention of instructors, researchers, and students. A rubric helps instructors decide how to teach

problem-solving strategies, provides researchers with analytical categories to parse students' work, and informs students how their work will be assessed. Therefore, the lack of explicit emphasis on blending conceptual and formal mathematical reasoning during mathematical processing could lead all three parties to emphasize stepwise problem-solving approaches at the expense of—rather than in addition to—the kind of reasoning Pat displays.

To make our case, we first review previous work on problem-solving expertise and rubrics, to establish that

(1) published rubrics in physics problem-solving encode a vision of expert problem solving in which mathematical manipulations is viewed as a separate step from those involving communication between conceptual reasoning and mathematics (e.g., "translating" a qualitative description of the physical situation into a mathematical model, and checking the final answer for plausibility), even though

(2) expert scientists or engineers and good students sometimes use conceptual reasoning *during* mathematical manipulations.

Then, using a validated and robust problem-solving rubric in the physics education research (PER) literature [3], we present a comparative analysis of Alex's and Pat's solutions to the two balls problem presented above. We show that although Pat's solution demonstrates greater expertise, problem-solving rubrics award approximately equal scores to Alex's and Pat's solutions. We will close with two alternative sets of implications for researchers and instructors. The first, less radical suggestion, is an amendment to current problem-solving rubrics. The second, more radical option, is a more fundamental rethinking of what constitutes problem-solving expertise and the role that globally applicable rubrics can play in assessing it.

II. LITERATURE REVIEW: INSTRUCTION AND RUBRICS TARGETING PHYSICS PROBLEM-SOLVING EXPERTISE

Quantitative problem solving plays a central role in the everyday work of scientists and engineers. For this and other reasons, improving students' problem-solving skills has been a central concern of physics instructors as well as education researchers [1,3–19]. Researchers often proceed by (1) describing deficits in students' problem-solving behaviors (for example, by comparing novices to experts), (2) generating instructional strategies that target those deficits, and (3) evaluating these strategies by measuring gains in students' performance, often using a rubric that closely aligns with how the students were taught to solve problems. Thus, the development of rubrics for assessing students' progress is intertwined with a vision of expert behavior that researchers and instructors want students to emulate. So, we start our review with a brief discussion of the research on physics students' problem-solving behaviors and on instructional strategies

to improve those behaviors. Then we discuss currently available rubrics used to evaluate physics problem solving. We close this section with our claim about what is missing in current rubrics, a claim supported in subsequent sections.

A. Expert-novice differences in problem solving: Diagnoses and interventions

Solving qualitative and quantitative problems on a variety of introductory and advanced physics topics has been a topic of much research, with researchers documenting systematic expert-novice differences in problem-solving behaviors. Some of this research has explicitly investigated expert reasoning [12,20]. In other cases, we can infer what counts as expert reasoning for the researchers based on what they value in students' problem-solving behaviors or choose as instructional targets. In what follows, "novice" refers to learners whose problem-solving behaviors differ from desired problem-solving behaviors. Some of the main conclusions of this body of work are as follows:

- Novices tend to start problems by launching directly into manipulating equations [21], whereas experts are more likely to begin with a qualitative description of the situation [12,13].
- Novices tend to categorize problems according to surface features, such as the presence of inclined planes, pulleys, or rotating objects,
- whereas experts classify problems according to "deep structure" (the relevant underlying physics principles), such as conservation of energy [12,22–29].
- Experts' knowledge is linked in a coherent fashion, whereas novices' knowledge is more fragmented or disconnected [30,31].
- Experts make facile use of sense-making tools such as diagrams [32,33].
- Novices are less likely to evaluate their solutions than experts [34–36].
- Novices are slower to abandon an unproductive line of reasoning and search for an alternative [33,37,38].
- Some researchers argue that novices lack conceptual understanding [39,40], mathematical skills [41–44], or the ability to transfer productive knowledge [45], and that these become barriers to effective problem solving.

Research on expert-novice differences has helped guide the development of instructional strategies to help students become better problem solvers. Some of these strategies focus on a particular aspect of problem solving [22,34,38,46,47], while others provide students with an entire framework for problem solving [9,15–18,31,46,48–57]. For example, to help students attend to deep structure, some researchers advocate making students state the underlying principles they are relying upon, such as conservation of energy or force balancing [22,38,46]. Other researchers advocate teaching students a hierarchical

approach to problem solving: start by outlining the "big picture" of how to solve the problem, then fill in the details, with frequent references back to that overall solution structure [15,16,31,46,48,49].

Several researchers spell out expert problem-solving procedures [9,15,16,18,50–57] and many researchers recommend explicitly teaching such strategies and enforcing their use. As an example, Heller, Keith, and Anderson [9] recommend teaching the following five-step strategy (excerpted from Table II, p. 630):

(1) *Visualize the problem.* Translate the words of the problem statement into a visual representation.
(2) *Describe the problem in physics terms (physics description).* Translate the sketch(es) into a physical representation of the problem.
(3) *Plan a solution.* Translate the physics description into a mathematical representation of the problem.
(4) *Execute the plan.* Translate the plan into a series of appropriate mathematical actions.
(5) *Check and evaluate:* Determine if the answer makes sense.

These five steps capture the general characteristics of the physics problem-solving strategies advocated by most physics education researchers (also see Woods [50] for a review of problem-solving strategies in physics and engineering). These strategies embody the idea that expert problem solving includes the following steps: qualitatively convert the real-world description of the problem into something that can be solved mathematically, then carry out the mathematical manipulations necessary to reach a solution, and finally evaluate the solution for meaning and plausibility.

B. Rubrics for evaluating problem solving

These conceptions of expert problem solving and how to teach it have influenced assessment tools for evaluating students' problem-solving skills. Some assessment schemes code for one particular feature that can indicate a component of problem-solving expertise, such as correctly identifying the underlying principle [46], writing a strategy for how to solve the problem [22], analyzing mistakes and modifying the problem-solving approach accordingly [48], and drawing diagrams [32]. Instructors, of course, constantly develop "whole-solution" rubrics for assessing homework and exam solutions from start to finish, but these rubrics are mostly inaccessible to researchers. Searching the articles cited in this paper and other literature, we could find no citations to whole-solution rubrics for physics problems other those developed by the University of Minnesota [3,9–11,14].

One of the most rigorous scoring schemes for evaluating students' problem solving is presented by Heller, Keith, and Anderson [9] (p. 631), paraphrased here and referred to later as the 1992 rubric:

(1) *Evidence of conceptual understanding.* How well the student's representation of the problem (words, diagrams, etc.) indicates an understanding of the relevant physics concepts and relations.
(2) *Usefulness of description.* Whether the representation of the problem, called the "physics description," includes *all* the information needed for a solution.
(3) *Match of equations with description.* Whether the student's equations are consistent with the physical description.
(4) *Reasonable plan.* Whether the student had an appropriate plan for executing the mathematical manipulations needed to solve the problem.
(5) *Logical progression.* How well the overall solution hangs together, from the selection of laws and principles to their appropriate application in the problem situation. This category also rewards students for performing algebraic simplifications before substituting in numerical values.
(6) *Appropriate mathematics.* Whether the mathematical manipulations and assumptions in the students' solutions are appropriate.

Detailed rubrics such as this provide a reliable means of evaluating students' performance on quantitative problems. As compared to simply coding for correct answers, rigorous rubrics help researchers and instructors attend to details of how students are solving the problem.[1] Over the years, the rubric of Heller *et al.* has been adopted, adapted, and built upon [3,10,11,18]. Most recently, Docktor and Heller [3] designed a rubric to be more easily deployed across a variety of instructional contexts by other researchers and instructors. Their rubric includes five scoring categories: "organizing problem information into a useful description, selecting appropriate physics principles, applying physics to the specific conditions in the problem, using mathematical procedures appropriately, and the overall communication of an organized reasoning pattern" (p. 4). This new rubric preserves the basic ideas from the 1992 rubric, though with some reorganization and streamlining (e.g., no "planning the solution" as a separate step, and no extra points for performing algebraic simplifications prior to numerical calculations). As emphasized above, these rubrics build on (i) perspectives of problem solving drawn not only from physics but also from mathematics and cognitive psychology [16,17,59,60], and (ii) research studies on differences between expert and novice problem solvers [12,13,61].

---

[1]In this paper, we do not distinguish between authentic "problems" and standard or routine "exercises" as some researchers do (e.g., Schoenfeld [58]). We use "problem" to refer to the kinds of homework and exam questions that students regularly encounter in a standard physics class and that rubrics are often applied to.

C. Missing from the rubrics: "Credit" for integrating conceptual and formal mathematical reasoning during mathematical processing

These rubrics were not designed to assess everything. For example, Docktor and Heller, referring to their rubric as "minimal measures," intentionally ignore underlying epistemological, metacognitive, and affective aspects of problem solving to focus on conceptual problem-solving elements. They did so to make the rubric more generally adoptable and independent of specific pedagogical approaches [62]. In this paper, however, we want to draw attention to a *conceptual* aspect of expert problem solving that current rubrics do not evaluate or credit: Pat-style integration of conceptual and formal mathematical reasoning *during* mathematical processing, what we call *blended processing*. We later argue that blended processing warrants attention in instruction as well as assessment.

The blended processing we illustrate in this paper builds on research that identifies students as blending intuitive knowledge and formal mathematical knowledge in certain situations. Sherin [63] observed undergraduate students creating equations about a physical scenario. For example, in one interview, a pair of students was considering the amount of rain that would fall on someone outside in a rainstorm, and they came up with the equation [total rain] = [no. raindrops/s] + C. They explained that the total rain hitting the person comes from two sources: the rain falling on the person's head, indicated by [no. raindrops/s], and the rain hitting the front of the person as they walk forward, indicated by "C"). In Sherin's analysis, these students generated their equation by drawing, in part, on the intuitive conceptual idea that the total is composed of the combination of two parts. Similar examples of students drawing simultaneously on mathematical and intuitive conceptual (nonmathematical) knowledge to create equations have been documented for undergraduate physics [64] and 8th grade algebra students [65].

Pat's solution to the two rocks problem, presented in the Introduction, is another example of blending intuitive conceptual and formal mathematical reasoning. He used a conceptual idea, that two things changing by the same amount will maintain their difference, in conjunction with formal mathematical ideas about linear relationships and slope. Importantly, this blending of conceptual reasoning and formal mathematical ideas did not occur while Pat was selecting the relevant equation. Rather, Pat used such reasoning during the "execute the plan" step (in the scheme of Heller, Keith, and Anderson), when the mathematical equations are typically manipulated to find a solution. We refer to this blending of conceptual and formal mathematical ideas during the mathematical processing stage of problem solving as blended processing.

To be clear, in our definition of blended processing, the conceptual knowledge or reasoning involved in blending is intuitive, everyday ideas about the world, not formal physics concepts. For example, suppose a student is solving for the forces acting on a sofa sliding with constant velocity. If the student generates a force equation by starting with Newton's second law and the definition of acceleration (to conclude that $a = 0$), that's good reasoning, but it is not blended processing, because the student relies on formal concepts. By contrast, if the student generates and processes an equation drawing on the intuitive notions that the sofa *keeps doing what it is doing* and hence competing influences on it must *balance out*, the seamless integration of intuitive conceptual ideas (like *balancing*) and a mathematical objects (in this case, equations) constitutes blended processing.

A common feature of the problem-solving rubrics we have presented is the credit given to formal conceptual reasoning used to *generate* equations but the absence of additional credit for intuitive or formal conceptual reasoning used during the mathematical manipulation *of* those equations. So, blended processing earns no additional credit. In Docktor and Heller's scheme, the category *mathematical procedures* does not explicitly look for conceptual reasoning during the mathematical processing. Similarly, in Heller, Keith, and Anderson's 1992 rubric, there is no mention in *reasonable plan* that students should think about the conceptual meaning of their mathematical manipulations. Of course, these problem-solving rubrics do not *penalize* students for opportunistically blending conceptual reasoning with formal mathematical reasoning during mathematical manipulations. However, they do not *reward* such blending either. This lack of reward sends the implicit message to students and instructors that the opportunistic use of blended processing is not a crucial component of problem-solving expertise.

However, the opportunistic blending of conceptual and formal mathematical reasoning *throughout* the problem solution, including during mathematical processing, is something done by professional engineers [66,67] and it can exemplify "adaptive expertise" [68] because it involves flexibly using available intellectual resources rather than following a fixed procedure. Furthermore, the blending of conceptual and formal mathematical reasoning throughout a solution is an example of connecting meaning to mathematical formalism, which indicates a deeper understanding than simply using the formalism, according to Wertheimer [69], Schoenfeld [58], Redish and Smith [70], Gainsburg [66], and others. And deeper, connected understanding is a component of expertise [31]. For all these reasons, we view the opportunistic blending of conceptual and formal mathematical reasoning throughout the solution, not just at the beginning and end, as an important aspect of expertise in physics problem solving. Our detailed explication below of Alex's and Pat's

solutions to the two balls problem will further buttress this argument.

In summary, thus far, we have argued that rubrics used to assess students' problem solving do not reward the opportunistic blending, in a single step, of conceptual and formal mathematical reasoning during mathematical manipulations, though such blending is arguably a component of expertise. Below, we support this argument empirically, by (i) summarizing prior analysis of Alex's and Pat's solutions to the two balls problem, to show that Pat's solution exhibits greater expertise, and then (ii) showing that the published problem-solving rubrics do not score Pat's work more highly than Alex's.

III. METHODOLOGY

A. Recruitment, selection, and interview context

From fall 2008 to fall 2011, we interviewed 13 students at the University of Maryland in an introductory, calculus-based physics course for engineering students covering mechanics. Our goal was to probe students' mathematical sense making, specifically when and how they make conceptual sense of equations in physics.

We recruited via the course Email list, asking for a one-hour interview for which participants would receive $10.00. We interviewed all students who expressed interest except for three students whose schedules were incompatible with ours.

The two students whose solutions we analyze here, Alex and Pat, were interviewed one month into the course. They both took the course during the same semester with the same instructor. At the time of the interview, the course had covered kinematics and Newton's laws, including projectile motion. An analysis of the 13 interviews showed that Alex's and Pat's solutions to the two balls problem (see below) are characteristic of two common solution paths [71]. We saw no gender or other obvious demographic differences between the groups following those two paths.

B. Interview protocol

We designed our interview protocol [72] to investigate what conceptual and mathematical knowledge students bring to bear when solving quantitative problems in physics. Students were prompted to think aloud while reasoning through problems. To probe their conceptual understanding and reasoning, students were also asked to explain some equations that they had seen in class, and some they had not. They were also asked to solve problems using these equations. In addition, the interview probed students' stances towards the nature of knowledge and learning and the use of mathematics in physics with prompts such as "How do you know when you really understand an equation?"

For this paper, only the first two prompts are relevant. Both focused on the kinematics equation $v = v_0 + at$. All

subjects said they were familiar with that equation. The first prompt was as follows:

> You have probably seen this equation in class:
> $v = v_0 + at$
> Here, v is velocity, v_0 is initial velocity, a is acceleration, and t is time. How would you explain this equation to a friend from your class?

The prompt was printed on paper, with the rest of the sheet blank. Students had access to a pen. Follow-up prompts, intended to cue more or less formal explanations of the equation, included explaining the equation to a 12-year old or on a physics exam.

To investigate how students would apply the equation to a quantitative problem, we asked the *two balls problem:*

> Suppose you are standing with two tennis balls on the balcony of a [high] apartment. You throw one ball down with an initial speed of 2 meters per second; at the same moment, you just let go of the other ball, i.e., just let it fall. I would like you to think aloud while figuring out what is the difference in the speed of the two balls after 5 seconds—is it less than, more than, or equal to 2 meters per second? (Acceleration due to gravity is 10 m/s^2.)

The students were instructed to speak aloud as they think. The interviewer also told interviewees that he was interested in how they approached the problem, not what answer they got. If a student brought up air resistance, the interviewer told the student to neglect it.

We chose the two balls problem because it can be solved without explicit calculation, as Pat did. We were interested in seeing whether and how students would implement such a solution. We asked students to explain the velocity equation immediately prior to posing the two balls problem in order to probe students' ways of thinking about the equation, and possibly even to prime conceptual reasoning. Although we did not instruct students to use the equation in solving the problem, both Alex and Pat did so.

C. Analysis

Our analysis started with a qualitative description of students' reasoning in interviews. One member of the research team watched and roughly transcribed the interviews. Then we and another team member, Brian Danielak, watched the interviews as a group, forming initial characterizations of how each student solved problems and of differences in solutions between students. Since Alex's and Pat's explanations of the kinematics equation and solutions to the two balls problem were starkly different, we selected those interviews for finer-grained analysis. We rely on this fine-grained analysis, summarized in the following section and presented in detail in another paper [71], to make the case that Pat exhibited a component of expert problem

solving that is missing in Alex's step-by-step numerical solution: blending conceptual and formal mathematical reasoning during mathematical processing.

Once we had completed this qualitative analysis, we realized that it held implications for problem-solving rubrics. We thus began to focus on how existing problem-solving rubrics would evaluate Alex's and Pat's solutions. To perform these analyses, we needed valid and reliable rubrics that attend to the students' entire solutions. We also needed the rubrics to be explained in sufficient detail that we could apply them. We found two such rubrics readily available in published literature [3,9], both from the same research group. In the main body of this paper, we apply Docktor and Heller's [3] rubric because (1) it closely aligns with the earlier, highly influential research on problem-solving strategies and problem-solving rubrics; (2) its reliability, validity, and utility were carefully established [14]; and (3) it was designed for ease of use by other researchers in other instructional settings. In the appendix, however, we also apply the rubric of Heller *et al.* [9] to Alex's and Pat's solutions.

There is one major methodological caveat in our use of these rubrics: they were designed to assess students' written solutions, not their think-aloud solutions. To generate a codable written solution for Alex and Pat, we combined their transcribed verbal utterances with their written work. Of course, they might have said or written things in the interview that they would not have written on a homework assignment or a test. Conversely, on an exam, Alex or Pat might have included reasoning not present in an interview. However, for the argument we are making in this paper, it is not important how precisely Alex's and Pat's written solutions would have matched their interview responses. It is only important that these two contrasting *kinds* of solutions could plausibly appear in students' written work.

Despite being asked to think out loud, Alex and Pat do not spell out their reasoning in complete detail. In grading such work, graders tend to project correct reasoning onto a student solution that is objectively lacking, merely because the final answer is correct [73]. To try to avoid that tendency and to provide accountability for our application of the rubric, we explain every decision we made in scoring both students and discuss what we perceived as relevant for determining their scores. This allows the reader to follow and evaluate the nuts and bolts of our decision-making process.

Initially, two of the authors (Hull and Elby) worked together to code both solutions on the five categories in the Docktor and Heller rubric. Then, the second author (Kuo) independently coded the two solutions. The coders agreed on all but three codes out of ten (two students across five categories).[2] In order to resolve the disagreements,

we all met to discuss the appropriate codes. The resulting agreement between coders relied heavily on revisiting the training examples provided in [14].

IV. PREVIOUS DATA AND ANALYSIS: ALEX'S AND PAT'S SOLUTIONS TO THE TWO-BALLS PROBLEM

In this section we summarize our analyses of Alex's and Pat's solutions to the two balls problem [71] to argue that Pat's solution shows greater expertise. We present this qualitative analysis to set up our new, rubric-based analyses of their solutions reported in Sec. V.

A. Alex's solution follows a productive but set procedure

Alex started by drawing a diagram of the two balls and labeling their speeds. At first, she incorrectly labeled the initial speed of the dropped ball as "9.8 m/s." Alex then decided to use the equation $v = v_0 + at$, paused to think about the value of "a," and then wrote down "9.8" as the value of a. The interviewer suggested that she is welcome to use either "10" or "9.8," and Alex decided that "10" is "probably easier." She then crossed out the "9.8" she wrote for the speed of the dropped ball and for the acceleration of the thrown ball and replaced both with "10." She then explicitly calculated the final velocity for each of the two balls correctly and wrote down the difference. When calculating the final velocity of the dropped ball using the equation, Alex correctly used 0 m/s as the initial speed. Figure 1 shows all of her written work.

After completing her calculations, Alex explained her reasoning:

> Alex: Ok, so after I plug this into the velocity equation, I use the acceleration and the initial velocity that's given, multiply the acceleration by the time that we're looking at, five seconds, and then once I know the velocities after five seconds of each of them, I subtract one from the other and get two. So the question asks "is it more than, less than, or equal to two," so I would say equal to two.
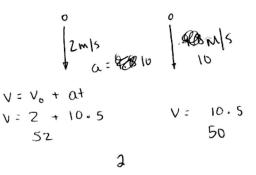

FIG. 1.   The written component of Alex's think-aloud solution to the two balls problem.

---

[2]The specific details of the coders' disagreements are detailed in a footnote after the results of the rubric coding are presented.

When then asked if it would have been possible to answer this question without "working out the numbers" to explicitly calculate the velocities, Alex initially said it should be possible but could then not think of a way to do so.

B. Pat's solution opportunistically blends conceptual with formal mathematical reasoning throughout

Pat started by referring to $v = v_0 + at$, but used the equation differently from the way Alex did.

> Pat: Well, the first thing I would think of is the equations. The velocity, I suppose, is the same equation as that other one [from the previous interview prompt] and I'm trying to think of calculus as well and what the differences do. So the acceleration is a constant and that means that velocity is linearly related to time and they're both at the same... so the first difference is the same. I think it's equal to two meters per second. So the first differences are the same, and if the first differences are the same then the initial difference between the two speeds should not change.

When asked, Pat explained that the term "first differences" came from his high school algebra class, where sets of data points would be analyzed by taking "delta y over delta x," which is called the "first difference." This suggests that "first difference," for Pat, is similar to the idea of a slope.

Pat spontaneously went on to say that "there's a couple of methods of attacking [the problem]," and that he would switch to these other methods if he ran into difficulty in one solution method. Prompted by the interviewer, Pat provided an alternate reasoning:

> Pat: So if I started from thinking about the equations and I'm not quite sure whether the velocities are changing at the same rate, then like sometimes I'll use several and see if they're consistent, then I could switch to thinking about the derivatives of the velocity and I'll think ok, so the initial conditions are off by 2 and then the velocities are changing at the same rate so that should mean they stay at 2.

Pat then started to explain the situation in terms of accelerating reference frames but decided this approach to be inappropriate for the given situation. He then went on to talk about how he could also solve the problem by performing more explicit calculations (see Fig. 2):

> Pat: I would think about one of them is zero and the other one is two and since their accelerations are the same, then I mean if you wanted to attack it on paper clearly you can subtract these and find that everything cancels except this two and that would be another way

FIG. 2. The written component of Pat's think-aloud solution to the two balls problem.

to attack this problem. But thinking about it in my head, I would realize that with you, without actually writing out the subtraction of the equations that the *at* term will be exactly the same for both of them since they say acceleration due to gravity is the same and they're both after five seconds.

Although Pat did a calculation here, he did not calculate the final velocities of the two balls as Alex did. He hinted that the explicit calculation was not necessary to his reasoning when he said, "if you wanted to attack it on paper clearly" and "I would realize that... without actually writing out the subtraction of the equations." Pat used the written work to mathematically communicate his reasoning: the "*at*" terms are identical for both balls and will cancel out when calculating the difference in final speeds.

C. Analysis: Pat's solution displays greater mathematical sense making

Alex's systematic solution follows the problem-solving steps described in common physics textbooks [1,2]. She decided which equation to use, drew a picture to organize the relevant information, substituted the known quantities in the equation, and solved for the unknown to arrive at the answer. She did connect the variables to their physical meanings. For example, she interpreted the symbol v as the physical concept of speed. However, she used the equation as a computational tool, where known values are plugged in to compute unknown values. She showed no evidence of making conceptual sense of the equation as a whole.

By contrast, Pat's solution, though messier and harder to parse, relies on conceptual reasoning that allowed him to minimize calculations. More importantly, his solution connects *mathematical formalism* (the equation $v = v_0 + at$) to that conceptual reasoning—in this case, the idea that if two things (like velocities) undergo the same change, the difference between them stays the same. Whereas Alex connected conceptual interpretations to mathematical formalism at the level of individual variables, Pat connected his conceptual reasoning to the equation as a whole. Pat saw this conceptual idea as applicable partly because of the presence in his reasoning of certain *mathematical* elements such as the linearity of the equation in time and the equality of the slopes.

In addition to differences in which aspects of the equation Alex and Pat physically interpret (the individual variables versus the equation as a whole), the two students also differ in *when* they make their interpretations. Pat's solution, unlike Alex's, involves blended processing, a blending of conceptual meaning with mathematical formalism *during the processing of mathematical equations* to reach a solution. As argued above, this blended processing is a form of mathematical sense making and hence a component of expertise in quantitative problem solving. However, this component is not reflected in problem-solving rubrics, as we now show.

V. RESULTS: CURRENT RUBRICS FAIL TO DISTINGUISH BETWEEN PAT AND ALEX

In this section, we assess Alex's and Pat's solutions to the two balls problem using the rubric developed by Docktor and Heller [3]. In the appendix, we use the rubric developed by Heller, Keith, and Anderson [9]. In both cases, Alex and Pat receive similar scores.

Using the Docktor and Heller rubric, for each of the five coding categories, we will quote the description of the category and then grade Alex's and Pat's solutions on a scale of 0–5 within that category, or NA(S) if that step was not necessary for the solver. Where appropriate, we have assigned a range of scores. The categories are *problem description*, *physics approach, specific application of physics, mathematical procedures*, and *logical progression*. In addition to the instructions contained in [3], we have used examples from Docktor's dissertation [14] (pp. 299–303) and the training materials contained within the dissertation as guides for our rubric scoring.

A. Useful description

Category description: "*Useful Description* assesses a solver's process of organizing information from the problem statement into an appropriate and useful representation that summarizes essential information symbolically, visually, and/or in writing... A problem description could include specifying known and unknown information, assigning appropriate symbols for quantities, stating a goal or target quantity, a sketch or picture of the physical situation, stating qualitative expectations, an abstracted physics diagram, drawing a graph, defining coordinate axes, and/or choosing a system" (Docktor and Heller [3], p. 5).

Alex scores a 4 or NA(S)—her description of the problem is "useful but contains minor omissions or errors" (p. 19 of [3]). She started by organizing all the given information in a sketch of the two balls. Though she did not explicitly put in symbols for the quantities she labeled, her verbal response shows that she understood what variable she was labeling. She did initially make an error in

labeling the initial speed of the dropped ball as "9.8." However, she quickly crossed this out and instead provided the acceleration of the ball ("10"), immediately after doing so for the thrown ball. What does remain in error for her useful description, however, are the units attached to the acceleration of the dropped ball (m/s instead of m/s^2). This error, however, does not propagate to the final solution. For this error, she could receive a 4. Alternatively, the grader could decide that Alex simply did not require a complete technically correct description, since her subsequent solution was correct. In this case, Alex would receive NA(S) for useful description [74].

Pat receives an NA(S) for this category, because he solved the problem correctly without an explicit description of the problem. Namely, when given the problem, he jumped into looking at the equation instead of first writing out the known and unknown values or explicitly describing the physical situation.

B. Physics approach

Category description: "The *Physics Approach* assesses a solver's process of selecting appropriate physics concepts and principles to use in solving the problem. Here the term "concept" is used to mean a general physics idea, such as the general concept of vector or specific concepts such as momentum and velocity. The term 'principle' is used to mean a fundamental physics rule or law used to describe objects and their interactions, such as conservation of energy or Newton's third law" (Docktor and Heller [3], p. 5).

A grader might give both Alex and Pat a 5 in this category. Both of them chose the appropriate equation to determine the final velocity of an object accelerating with a constant acceleration, likely supported by the previous interview question specifically focusing on that equation. Furthermore, all of the additional concepts that Pat identified as relevant (e.g., first differences, rates of change, and accelerating reference frames) can be applied correctly and productively to this problem. Alternatively, one might give both Alex and Pat NA(Problem) for this part. Although we did not instruct students to use the velocity equation in solving the two rocks problem, it was cued up by the previous interview prompt, and hence students did not really need to *decide* which "physics approach" to use.

C. Specific application of physics

Category description: "*Specific Application of Physics* assesses the solver's process of applying physics concepts and principles to the specific conditions in the problem. Specific application often involves connecting the objects and quantities in the problem to the appropriate terms in specific physics relationships... This

category separates the identification of appropriate principles and concepts in the Physics Approach from the actual application of those principles to the specific conditions in the problem'' (Docktor and Heller [3], pp. 5 and 6).

Alex and Pat both score a 5 in this category. Both correctly connected the variables in the velocity equation (v, v_0, a, and t) to the conditions in the two balls problem where needed.

D. Mathematical procedures

Category description: "*Mathematical Procedures* assesses the solver's process of executing the solution with respect to selecting appropriate mathematical procedures and following mathematical rules to obtain target quantities'' (Docktor and Heller [3], p. 6).

Alex scores a 5 here. Alex correctly calculated the final velocity of each ball and then correctly subtracted the two values to obtain the difference.

Pat's scoring here is more complicated. In the first parts of his solution, Pat arrived at an answer in two different ways without explicit mathematical calculations ("... if the first differences are the same then the initial difference between the two speeds should not change''). This could be scored NA(S) or a 5, depending on whether the grader counts the formal mathematical components of Pat's blending of conceptual and formal mathematical reasoning to be "executing ... mathematical procedures.''

However, when he finally wrote out the velocity equation for both balls, plugging in the values of v_0 for both, he accurately found the difference in final speeds of the two balls. For this mathematical calculation, Pat would receive a 5. For comparison purposes, we will score Pat a 5 in this section, because even though explicit mathematical

procedures were not central to his reasoning, they were applied correctly when used.

E. Logical progression

Category description: "[*Logical Progression*] checks whether the overall problem solution is clear, focused, and organized logically. The term "logical" means that the solution is coherent (the solution order and solver's reasoning can be understood from what is written), internally consistent (parts do not contradict), and externally consistent (results agree with qualitative physics expectations)'' (Docktor and Heller [3], p. 6).

Alex would score either a 4 or 5. Overall, her solution is clear and systematic. Her diagram highlights what she thinks is relevant in the problem situation, and her verbal utterances suggest that she had a clear plan of computing the two final velocities and subtracting one from the other to get the difference, which is an appropriate solution. Her solution does not seem qualitatively ridiculous, compared to typical values in the problem. Although she did make an error in writing out the initial velocity of the dropped ball, she remedied this by turning it into an acceleration, and her end product (verbal and written) does not include that error. Her correction, however, came without explanation, and her reasoning about whether "9.8" is the initial velocity or "10" is the acceleration thus seems inconsistent or at least unclear. If a point had already been taken off for this in the *useful description* section, a score of 5 would be given here so as not to double count this error. If NA(S) had been given in the *useful description* section, 4 points would be given here [74].

Pat also scores either a 4 or 5. Although a "cleaned-up'' restatement of his original solution is logical and consistent, his first solution is hard to follow, because of some initially undefined terms (i.e., first differences) and lack of explicit calculations. So, he loses points for clarity, but not

TABLE I. Pat and Alex score similar on the Doctor and Heller [3] rubric. Summary of scores on rubric categories.

Category (from Docktor and Heller)	Alex's score	Pat's score	Justification for deductions
Useful Description	4 or NA(S)[a]	NA(S)	Alex first wrote that the speed of the dropped ball is 10 m/s. Pat did not require a useful description in his solution.
Physics Approach	5 or NA(P)	5 or NA(P)	
Specific Application of Physics	5	5	
Mathematical Procedures	5	5	
Logical Progression	4–5[a]	4–5	Alex's solution was not consistent with her saying that the initial speed is 10 m/s. Pat's first solution is hard to follow, because of some initially undefined terms and lack of explicit calculations.

[a]Alex loses 1 point, either in Useful Description or in Logical Progression.

a significant amount since his eventual explicit calculation is clear.

F. Coding summary

As Table I shows, Docktor and Heller's rubric awards approximately the same scores to Pat's and Alex's solution in each category. The 1992 rubric of Heller *et al.* produced similar results (see Table II in the appendix). We recognize that other graders using this rubric may assign scores differently. Even so, we cannot imagine an application of this rubric that gives Alex and Pat *significantly* different scores.[3] More importantly, we cannot imagine an application of this rubric that rewards Pat for blending conceptual and formal mathematical reasoning *during* the mathematical procedures by giving him a higher score than Alex in that category.

Most of the students interviewed, like Alex and Pat, had little difficulty in solving the two rocks problem correctly. One could argue that this problem was too simple for us to expect to see large differences across students' work. We argue, however, that even on this "simple" problem, there is an important difference in expertise between Alex's and Pat's solutions, and the fact that their rubric scores are similar (on codable categories) is evidence that current rubrics are not sensitive to the opportunistic use of blended processing. Even if the problem were more difficult and current rubrics did yield different scores for the two students, it would not be for the reason we argue for in this paper. Our main point is that, on either "simple" or "difficult" problems, current rubrics are not sensitive to the expertise reflected by blended processing.

VI. DISCUSSION

In this section, we first address some challenges to the validity and relevance of our results. Then we discuss implications for researchers and instructors.

A. Considering possible counterarguments

A skeptic could argue that our result is only marginally relevant for PER because Pat's solution is idiosyncratic; we simply should not expect a problem-solving rubric to "see" his emerging expertise, any more than we expect

---

[3]Even for the three initial disagreements of the independent scores by Hull and Elby and by Kuo, the scores were still close. On two of the disagreements, the differences were within one point: Hull and Elby had given Alex 3 or 4 on Logical Progression, whereas Kuo had given Alex a 4 or 5. Also, Hull and Elby had given Pat a 5 on Logical Progression, whereas Kuo had given 4 or 5. The third disagreement was on whether NA(S) was an appropriate code. Hull and Elby had given Pat a 2 or 3 on Useful Description, whereas Kuo had given NA(S). Most importantly, however, is that the reasons for these differences in scoring were not related to our central issue: the blended processing that we see in Pat's solution.

the math SATs to "see" the difference between the top math student in a high school and a true math prodigy.

We acknowledge that the *details* of Pat's solution (e.g., "first differences") are idiosyncratic; however, blending conceptual and formal mathematical reasoning is common among "good" physics students, although it may not manifest itself in graded class work (like homework and exams). For instance, posing difficult physics problems to pairs of students enrolled in a third-semester physics course for engineers at high-ranked State University, Sherin [63] found that most of them blended intuitive conceptual knowledge and formal mathematical knowledge, drawing on knowledge elements that he called *symbolic forms*.

A symbolic form is a blend of a *conceptual schema* with a *symbol template*. A conceptual schema is a simple and intuitive idea about the physical situation. Two examples of conceptual schemata provided by Sherin are (i) two factors competing with each other, like an airplane engine competing with a strong headwind, and (ii) a whole consisting of multiple parts (e.g., conceptualizing the airplane's total body as composed of the wings, the cockpit, the hull, etc.). A symbol template is the general "form" of an algebraic equation or expression, without attending to the specific variables or terms. For example, for the equation $F_{\text{friction}} = F_{\text{push}}$, which might describe pushing a box at constant velocity, the symbol template is $\square = \square$. Alternatively, the first law of thermodynamics, $\Delta E = Q + W$, has a symbol template of $\square = \square + \square$. It is the blending of these two elements, a conceptual schema and a symbol template, that constitutes a symbolic form. For example, the base + change symbolic form combines the conceptual schema that *the amount you end up with is the amount you start with plus the amount you gain* with the symbol template $\square = \square + \Delta$. Other researchers have also used symbolic forms in modeling how students translate between mathematical solutions and physical understanding [64,65,75,76].

In other work, we argue that Pat used this base + change symbolic form when explaining the equation $v = v_0 + at$; he linked v_0 to the amount (of speed) you start with and at to the amount (of speed) you gain or lose, and hence their sum to the amount (of speed) you end up with. We argue that the use of symbolic forms-based reasoning is part of the blended processing Pat does in solving the two balls problem, as discussed above. In doing so, Pat was *not* idiosyncratic; 4 of the 13 students we interviewed exhibited blended processing when solving the two ball problems and 7 of the 13 interviewed students used symbolic forms-based reasoning when explaining the velocity equation [71].

In summary, the rarity of Pat-like reasoning on students' written work may stem more from how physics courses encourage students to formulate and present their solutions than from students' inability to blend conceptual and

formal mathematical reasoning. Of course, many students need more instructional support than Pat does to excel at such blending. But that is our point: helping students learn to blend conceptual and symbolic reasoning *throughout the solution* should be an important target of problem-solving instruction, because it is a component of quantitative problem-solving expertise that builds on knowledge students already possess (*intuitive* conceptual schemata and basic algebraic symbol templates).

In response, a skeptic could acknowledge that helping students blend conceptual and symbolic reasoning during mathematical processing is a worthy goal but not a feasible one; maybe average physics students cannot be taught this sophisticated skill, given the constraints typical of introductory physics courses. PER has a long record, however, of finding ways to help average students do things that only top students do in traditional courses [77,78]. In the conclusion below, we discuss in more detail the feasibility of instruction that targets blended processing.

In this section, we emphasize the importance of assessment in such instruction. Just as University of Washington Tutorials succeed partly because exams reward the deeper conceptual understanding taught by Tutorials [79], and just as the problem-solving instruction of Heller *et al.* succeeds partly because exams reward students for citing the physical principles they are using to solve a quantitative problem [9], instruction targeting blended processing will work best if students get rewarded for such reasoning within the framework of their physics course. That is why the rubrics used by researchers—and, ultimately, instructors—need to reward blended processing if we hope to help more students reason like Pat.

B. Options for enhancing assessment rubrics

Before discussing how problem-solving rubrics might be changed, we want to emphasize that the Docktor and Heller rubric is very good at what it is designed to do. It emerges from previous research in problem-solving expertise and improves upon other, less nuanced assessment schemes. The rubric looks at more than just final answers, focusing primarily on problem-solving processes (e.g., on whether a student starts with an understanding of the physical situation and translates from that physical description to a mathematical representation).

At the same time, Docktor and Heller recognize a tension and trade-off between a rubric's ability to be applied broadly and its ability to evaluate the full range of problem-solving processes, such as metacognition and epistemology. We echo this tension but also point out, in particular, one specific unexamined dimension of problem-solving expertise: the opportunistic blending of symbolic and conceptual reasoning *during* the mathematical manipulation stage. Given our case for the importance of this aspect of problem-solving expertise, how can this tension be resolved?

In an effort to build on the work of Docktor and Heller and other assessment designers, we will consider two possibilities: (i) continuing to use a "global" rubric (i.e., one meant to apply to all quantitative physics problems) but modifying it, or (ii) adapting a global rubric to each given problem, to create problem-specific rubrics that reward the particular opportunities to demonstrate expertise afforded by each problem.

1. Retain a global rubric

As Docktor and Heller (and others) discuss, retaining a globally applicable rubric has huge advantages: performance on different problems can be compared, researchers and instructors need to master just one scheme, and instruction can consistently be tied to the rubric. To retain these advantages, rubric creators could respond to our argument in this paper by adding an element to the *mathematical procedures* category, something like "when productive, the student blends conceptual and formal mathematical reasoning to minimize calculations."[4]

Although we see this addition as an improvement to current rubrics, rubric developers have at least two good reasons to reject it. First, there is no clean way to specify when it is "productive" to blend conceptual and formal mathematical reasoning. So, our proposed addition to the rubric calls upon coders to use their professional judgment about the productivity of a particular heuristic. This goes against the intent of a rubric, which is to make the coding as objective as possible, helping to ensure that different coders reach the same conclusions. Second, blending conceptual and formal mathematical reasoning is just one of potentially many heuristics that a good problem solver could apply to some problems but not others, such as making appropriate approximations that simplify calculations, or considering limiting cases to decide on relevant variables and evaluate solutions. Cluttering up the rubric with a bunch of heuristics that do not apply to most quantitative physics problems would make it unwieldy.

So, global rubric creators have good reasons not to modify their rubrics in response to this paper. In that case, the main point of our paper is about trade-offs. In exchange for gaining the advantages of a global rubric,

---

[4]Although one may instead consider adding productive blended processing as its own sixth category in the rubric, we feel blended processing fits within mathematical procedures because it is a "process of executing the solution" and "obtaining target quantities" (Docktor and Heller [3], p. 6). Moreover, we feel that part of expertise with mathematical procedures is avoiding unnecessary calculations, as Pat demonstrates with his blended processing. Last, our argument is that the mathematical processing step should not be divorced from conceptual reasoning, as it is typically portrayed in rubrics. Thus, while including productive blended processing as a sixth rubric category may be successful in practice, it could be interpreted as reinforcing the very separation of conceptual reasoning and formal mathematical reasoning that we argue against.

researchers and instructors give up something else: a systematic way to recognize and ''reward'' the heuristics that good problem solvers opportunistically use on some but not all problems, such as blending of conceptual and formal mathematical reasoning during mathematical manipulations. This trade-off is consequential: to the extent that problem-solving rubrics guide instruction and/ or learning, failing to reward blended processing could implicitly devalue the development of such reasoning.

2. Adapt a global rubric to create problem-specific rubrics

We see a preferable way of facing these trade-offs: treat rubrics such as that of Docktor and Heller as adaptable templates rather than as ''finalized'' rubrics. An example will illustrate what we mean. Our research on students' approach to the two balls problem suggests that Pat-style blended processing is part of expert reasoning *on that problem*. So, in creating a rubric for that problem, we would start with Docktor and Heller's rubric but add ''blending of conceptual and formal mathematical reasoning'' to the mathematical procedures category. Other problems might invite other adaptations to the global rubric. On some problems, we might add ''uses approximations to minimize calculations.'' On another, we might add ''considers limiting cases to narrow down which variables are relevant.'' On another, we might add ''generates a non-standard but appropriate representation that allows the student to minimize calculations,'' and so on. This approach has the advantage of retaining all the insight contained in the global rubric while also allowing researchers to recognize and reward the heuristics that experts and good students opportunistically use on some problems but not others.

Our adaptation approach to rubric creation also has many potential disadvantages, which we now address. First, it relies on professional judgment to decide which opportunistic problem-solving heuristics warrant inclusion in the rubric for a given problem. However, this is a problem faced by global rubrics as well. Graders have to make a judgment, based on the guidelines of the rubric, in order to assign a numerical score to student work. In addition to a five-point scale, Docktor and Heller also provide the discretion to attribute NA(Problem) and NA(solver)—basically, the grader decides whether a problem needs to have, for example, a *useful description* and even whether a particular student needs to provide a *useful description*.

One example of how problem-specific rubrics could minimize (though not eliminate) this lack of objectivity is as follows: before deciding to include blending conceptual and formal mathematical reasoning during the mathematical procedures step in the two balls problem rubric, we would show Alex-like and Pat-like solutions to multiple physics professors, probing if they consider Pat's solution to be more expert. To tap into professors' views about disciplinary expertise as opposed to their views about how physics should be taught, we might ask them ''based

on these solutions, which students would you most want for a summer internship in your lab?'' or ''which students would you recruit to grad school?'' rather than ''which solution should be graded higher?''

A second potential disadvantage of creating problem-specific rubrics rather than using a global rubric is the apparent need for laborious, problem-by-problem research, like our exploration of the two balls problem. Such research need not happen, however, if the researcher or instructor delays finalizing the rubric until looking through a representative sample of student responses. For instance, on the two balls problem, if a sophisticated solution shows up in that sample, the researcher or instructor could quickly check with colleagues that the solution does indeed reflect problem-solving expertise, and modify the rubric accordingly.

VII. CONCLUSION

In this paper, we contrasted two students' solutions to a physics problem in order to argue that current problem-solving rubrics do not reward an important component of quantitative problem-solving expertise—the opportunistic blending of conceptual and formal mathematical reasoning during the mathematical processing stage. We argued that researchers and instructors should adapt a global problem-solving rubric to create problem-specific rubrics that capture the heuristics opportunistically used by experts and by good students on particular problems.

Researchers and instructors could question the importance of our results on the following grounds: for most students, the main instructional target needs to be learning to think about physical principles *at all* when solving quantitative problems. By this argument, focusing on sophisticated heuristics such as blended processing is too much, too soon; such a focus becomes a feasible and worthwhile instructional target only after students achieve basic competence in core problem-solving skills such as thinking through the relevant physical concepts, creating productive representations to scaffold sense making, and translating the physical description into equations. Under this view, Docktor and Heller's is a sufficient assessment tool.

We agree that, if researchers and instructors must choose between focusing on these core skills or focusing on blended processing, they should focus on the core skills. We argue, however, that thinking conceptually while manipulating equations is something students can learn to do. As discussed in this paper and in another article [71], we have seen some students who already do this with the velocity equation, in spite of not being explicitly rewarded for such blending and in spite of not being taught this behavior. If students received explicit instruction in combining conceptual reasoning with the formalism of mathematical manipulations, we suspect such reasoning would become more prevalent. Expertise in problem solving in physics and engineering is not limited to mathematical

sense making at the start and end of the problem. If the goal is to develop students' problem-solving expertise, problem-solving strategies and rubrics must similarly not be limited.

Going further, we argue that, in problem-solving instruction centered around both conceptual and mathematical sense- making (cf. Ref. [80]), helping students blend conceptual and formal mathematical reasoning during mathematical processing could support students' development of core problem-solving skills. First, such blending could help a student rethink the relevant concepts involved in a problem even after she has written down equations, thereby increasing her focus on the underlying concepts. Second, blended processing could support students' creation of representations. For example, as part of his blended processing during the two balls problem, Pat talked about how the velocities of both balls linearly increase at the same rate. He did not draw the corresponding velocity versus time graphs, but instruction could prompt students to do so, in which case his blended processing would support the generation of a productive representation for the problem at hand.

Our point is that helping students blend conceptual and formal mathematical reasoning during mathematical processing could support rather than interfere with instructional efforts to develop students' basic problem-solving skills such as using conceptual reasoning at all, producing useful representations, and translating between conceptual and mathematical ideas. At the very least, the utility of instruction that helps students blend conceptual and formal mathematical reasoning during mathematical processing— and hence the utility of problem-solving rubrics that recognize and reward such blended processing, when it is productive—is an empirical issue worth exploring.

We feel future research should be carried out to develop, implement, and evaluate instructional strategies for helping students learn to opportunistically blend conceptual and formal mathematical ideas.

APPENDIX: HELLER, KEITH, AND ANDERSON RUBRIC

Heller, Keith, and Anderson [9] used the following six categories to score students' problem solutions (see p. 631):

(1) Evidence of conceptual understanding: Does the physics description reveal a clear understanding of physics concepts and relations? For example, does the description indicate curvilinear trajectories for projectiles or incorrect straight-line trajectories? Does the solution employ unbalanced forces for an accelerating object, or incorrect balanced forces?

(2) Usefulness of description: Is the essential information needed for a solution present? For example, do the force diagrams include all the relevant forces? For collision problems, are the momentum vectors both before and after an interaction clearly indicated?

(3) Match of equations with description: Are the specific equations used consistent with the physics described? For example, are vector equations used to relate vector quantities? Are the described forces appropriately included in specific force equations?

TABLE II.   Pat and Alex were scored according to two rubrics. Summary of the Heller, Keith, and Anderson rubric assessment [9]. The Docktor and Heller rubric assessment is presented in the body of the paper.

Category (from Heller, Keith, and Anderson, [9])	Alex (%)	Pat (%)	Reason for deductions
Evidence of Conceptual Understanding	100	70–100	Pat did not explicitly articulate information about the direction of the balls' acceleration relative to the thrown ball's initial velocity.
Usefulness of Description	80–100	70–80; 100[a]	No evidence that Alex was thinking of the v_0 for the dropped ball as being zero (as opposed to just forgetting to put it in).
Match of Equations with Description	100	100	
Reasonable Plan	100	100	
Logical Progression	80–100	100	Alex's number substitution occurred *before* computing an algebraic expression for the difference of the two velocities.
Appropriate mathematics	100	100	
Overall Score	93–100	95–97	

[a]Pat must lose points for the reason above. However, if Pat did not receive a 100% in the first category, he receives a 100% in the second category to avoid double counting.

(4) Reasonable plan: Does the solution indicate that sufficient equations were assembled before the algebraic manipulations of equations were undertaken? Does the solution include an indication of how to combine equations to obtain an answer?

(5) Logical progression: Does the mathematical solution progress logically from general expressions of physics principles to a problem-specific formulation using defined variables? Are numbers substituted for variables only after an algebraic solution for the unknown variable was obtained?

(6) Appropriate mathematics: Aside from minor mistakes, is the mathematics used reasonable? Or does the solution employ invalid mathematical claims in order to obtain an answer (e.g., the mass is small, so set $m = 1$)?

The ratings for these six characteristics were equally weighted in determining an overall score for the problem (out of 100%).

1. Scoring Alex's solution based on the Heller, Keith, and Anderson rubric

(1) Evidence of conceptual understanding: Alex's solution reflects a clear description of the problem situation. The picture she drew correctly depicts the trajectories of the two balls and important information about each ball is labeled. However, it might be a bit confusing because she labeled the speed as 2 m/s on the left ball but the acceleration as 10 m/s on the right ball. However, in her numerical solution she used the "10" as an acceleration, not as a speed. She treated the two balls as having the same acceleration due to gravity. Hence, we score her as 100% in this category.

(2) Usefulness of description: Alex identified the accelerations of both balls, but only indicated the velocity of one ball. Even though the other ball has an initial velocity of zero, this knowledge does not appear explicitly in her solution. Although her final answer does not include the initial velocity of the ball dropped from rest, she did not make clear how she set up the problem, and hence we have no evidence that she was thinking of the v_0 for that ball as being zero (as opposed to just forgetting to put it in). On the other hand, her numerical solution has no errors. So a strict coder might give her 80% on this category, while a more lenient coder would give her 100%.

(3) Match of equations with description: Alex scores a perfect 100% in this category.

(4) Reasonable plan: Again, Alex scores a perfect 100%. She used the provided equation twice and then clearly subtracted the two final velocities to arrive at her answer.

(5) Logical progression: This is perfect with the exception that she substituted numbers before computing an algebraic expression for the difference of the two velocities. On the other hand, her solution was relying on computing each velocity separately, and she did write down the equation for that in symbols before putting in numbers (no further algebra was needed for her approach). So a grader who is strict about computing a symbolic expression might award her 80% in this category; a more lenient grader would award her 100%.

(6) Appropriate mathematics: Perfect (100%).

Averaging Alex's scores yields a score of 93%–100%.

2. Scoring Pat's solution based on the Heller, Keith, and Anderson rubric

(1) Evidence of conceptual understanding: Pat actually solved the problem in myriad ways, some of which are more correct than others. However, to give him the benefit of the doubt, we will look only at the parts that would give him the best score. Although Pat made it clear that he is thinking about balls accelerating, and even accelerating at the same rate for the same time, he did not explicitly express an understanding that the acceleration is in the same direction as the starting velocity of the one ball, which is information that a picture would convey. This is a pretty important point, so he could be scored as low as 70%. However, it could be argued that this point was so obvious to him that he did not feel a need to say it explicitly and that his problem solving reflects this understanding, so he could get the maximum credit of 100%.

(2) Usefulness of description: Even giving him the benefit of the doubt, using this rubric, he must lose points for not being explicit about the relative direction of the acceleration. Everything else needed is presented, however, so he would receive 70%–80%. Note that the authors write that their rubric is careful not to take off points twice. Thus, if Pat was penalized in part 1, he would *not* be penalized again here.

(3) Match of equations with description: Perfect (100%).

(4) Reasonable plan: Perfect (100%).

(5) Logical progression: Perfect (100%).

(6) Appropriate mathematics: Perfect (100%).

Because Pat did not explicitly articulate, either in writing or in speaking, the direction of the balls' acceleration relative to the thrown ball's initial velocity, and because this information is critical in correctly solving the problem, Pat would score between 95% and 97%, only marginally higher than the bottom of Alex's score range and below the top of Alex's score range. We argue that Pat clearly performed more expertlike than Alex and that rubrics should identify this.

[1] D. C. Giancoli, *Physics for Scientists and Engineers with Modern Physics* (Pearson Education, Upper Saddle River, NJ, 2008), 4th ed.

[2] H. D. Young and R. A. Freedman, *University Physics with Modern Physics with MasteringPhysics^{TM}* (Addison-Wesley, San Francisco, 2003), 11th ed.

[3] J. Docktor and K. Heller, Robust assessment instrument for student problem solving, http://groups.physics.umn.edu/physed/Talks/Docktor_NARST09_paper.pdf (2009).

[4] L. Hsu, E. Brewe, T. M. Foster, and K. A. Harper, Resource Letter RPS-1: Research in problem solving, Am. J. Phys. **72,** 1147 (2004).

[5] D. P. Maloney, in *Handbook of Research on Science Teaching and Learning*, edited by D. Gabel (MacMillan, New York, 1994).

[6] D. Maloney, *Getting Started in PER* (American Association of Physics Teachers, College Park, MD, 2011), 1st ed.

[7] R. D. Knight, B. Jones, and S. Field, *College Physics A Strategic Approach* (Pearson, San Francisco, 2007), 1st ed.

[8] P. A. Tipler and G. Mosca, *Physics for Scientists and Engineers* (W. H. Freeman and Co., New York, 2008).

[9] P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, Am. J. Phys. **60,** 627 (1992).

[10] J. M. Blue, Ph.D. thesis, University of Minnesota, 1997.

[11] T. Foster, Ph.D. thesis, University of Minnesota, 2000.

[12] 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).

[13] J. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, Expert and novice performance in solving physics problems, Science **208,** 1335 (1980).

[14] J. L. Docktor, Ph.D. thesis, University of Minnesota, 2009.

[15] J. I. Heller and F. Reif, Prescribing effective human problem-solving processes: Problem description in physics, Cogn. Instr. **1,** 177 (1984).

[16] F. Reif and J. Heller, Knowledge structure and problem solving in physics, Educ. Psychol. **17,** 102 (1982).

[17] A. van Heuvelen, Learning to think like a physicist: A review of researchbased instructional strategies, Am. J. Phys. **59,** 891 (1991).

[18] D. Huffman, Effect of explicit problem solving instruction on high school students' problem-solving performance and conceptual understanding of physics, J. Res. Sci. Teach. **34,** 551 (1997).

[19] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, Am. J. Phys. **60,** 637 (1992).

[20] M. T. H. Chi, R. Glaser, and E. Rees, *Advances in the Psychology of Human Intelligence* (Erlbaum, Hillsdale, NJ, 1982).

[21] 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).

[22] W. J. Leonard, R. J. Dufresne, and J. P. Mestre, Using qualitative problemsolving strategies to highlight the role of conceptual knowledge in solving problems, Am. J. Phys. **64,** 1495 (1996).

[23] A. Feil and J. Mestre, in *Proceedings of the 2007 Physics Education Research Conference* (AIP, Greensboro, NC, 2007), pp. 100–103.

[24] A. Feil and J. P. Mestre, Change blindness as a means of studying expertise in physics, J. Learn. Sci. **19,** 480 (2010).

[25] A. Mason and C. Singh, Assessing expertise in introductory physics using categorization task, Phys. Rev. ST Phys. Educ. Res. **7,** 020110 (2011).

[26] 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).

[27] S. Y. Lin and C. Singh, Using isomorphic problems to learn introductory physics, Phys. Rev. ST Phys. Educ. Res. **7,** 020104 (2011).

[28] C. Singh, Assessing student expertise in introductory physics with isomorphic problems. I. Performance on nonintuitive problem pair from introductory physics, Phys. Rev. ST Phys. Educ. Res. **4,** 010104 (2008).

[29] C. Singh, Assessing student expertise in introductory physics with isomorphic problems. II. Effect of some potential factors on problem solving and transfer, Phys. Rev. ST Phys. Educ. Res. **4,** 010105 (2008).

[30] F. Reif, Millikan Lecture 1994: Understanding and teaching important scientific thought processes, Am. J. Phys. **63,** 17 (1995).

[31] F. Reif, *Applying Cognitive Science to Education: Thinking and Learning in Scientific and Other Complex Domains.* (MIT Press, Cambridge, MA, 2008).

[32] A. Mason and C. Singh, Helping students learn effective problem solving strategies by reflecting with peers, Am. J. Phys. **78,** 748 (2010).

[33] J. Tuminaro, Ph.D. thesis, University of Maryland, 2004.

[34] A. R. Warren, Impact of teaching students to use evaluation strategies, Phys. Rev. ST Phys. Educ. Res. **6,** 020103 (2010).

[35] K. L. Malone, Correlations among knowledge structures, force concept inventory, and problem-solving behaviors, Phys. Rev. ST Phys. Educ. Res. **4,** 020107 (2008).

[36] A. S. Dhillon, Individual differences within problem-solving strategies used in physics, Sci. Educ. **82,** 379 (1998).

[37] T. J. Bing, Ph.D. thesis, University of Maryland, 2008.

[38] P. Antonenko, C. Ogilvie, D. Niederhauser, J. Jackman, P. Kumsaikaew, R. Marathe, and S. Ryan, Understanding student pathways in context-rich problems, Educ. Inf. Technol. **16,** 323 (2010).

[39] L. C. Mcdermott, M. L. Rosenquist, and E. H. Vanzee, Student difficulties in connecting graphs and physics: Examples from kinematics, Am. J. Phys. **55,** 503 (1987).

[40] L. C. Mcdermott and P. S. Shaffer, Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding, Am. J. Phys. **60,** 994 (1992).

[41] E. T. Torigoe and G. E. Gladding, Connecting symbolic difficulties with failure in physics, Am. J. Phys. **79,** 133 (2011).

[42] E. Torigoe, Ph.D. thesis, University of Illinois, 2008.

[43] A. Raw, Developing A-level physics students' mathematical skills—A way forward?, Phys. Educ. **34,** 306 (1999).

[44] C. Valiotis, in *Proceedings of the 2008 American Society for Engineering Education Pacific Southwest Annual Conference* (ASEE, Washington, DC, 2008).

[45] J. P. Mestre, Probing adults' conceptual understanding and transfer of learning via problem posing, J. Appl. Dev. Psychol. **23**, 9 (2002).

[46] 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).

[47] R. Mualem and B. S. Eylon, Junior high school physics: Using a qualitative strategy for successful problem solving, J. Res. Sci. Teach. **47**, 1094 (2010).

[48] B.-S. Eylon and F. Reif, Effects of knowledge organization on task performance, Cogn. Instr. **1**, 5 (1984).

[49] J. P. Mestre, R. J. Dufresne, W. J. Gerace, P. T. Hardiman, and J. S. Touger, Promoting skilled problem-solving behavior among beginning physics students, J. Res. Sci. Teach. **30**, 303 (1993).

[50] D. R. Woods, An evidence-based strategy for problem solving, J. Eng. Educ. **89**, 443 (2000).

[51] J. L. Lubkin, *The Teaching of Elementary Problem Solving in Engineering and Related Fields* (ASEE, Washington, DC, 1980).

[52] R. Zajchowski, *How to Study Science and Solve Problems* (John Abbott College, Quebec, 1980).

[53] P. Heller and K. Heller, *Cooperative Group Problem Solving in Physics* (Department of Physics, University of Minnesota, Minneapolis/St. Paul, MN, 1999).

[54] W. Harvey, Ph.D. thesis, University of California, 1981.

[55] J. H. Larkin and F. Reif, Understanding and teaching problem-solving in physics, Eur. J. Sci. Educ. **1**, 191 (1979).

[56] H. C. McAllister, http://www.hawaii.edu/suremath/learn4.html (1997).

[57] K. Heller and P. Heller, *Competent Problem Solver-Calculus Version* (McGraw-Hill, New York, 2000).

[58] A. H. Schoenfeld, *Handbook for Research on Mathematics Teaching and Learning* (MacMillan, New York, 1992), pp. 334–370.

[59] J. R. Hayes, *The Complete Problem Solver* (L. Earlbaum Assoc., Hillsdale, NJ, 1989).

[60] G. Polya, *How to Solve It: A New Aspect of Mathematical Method* (Princeton University Press, Princeton, NJ, 1948).

[61] J. H. Larkin, Processing information for effective problem solving, Eng. Educ. **70**, 285 (1979).

[62] J. Docktor and K. Heller, in *Proceedings of the 2009 Physics Education Research Conference* (AIP, Ann Arbor, MI, 2009).

[63] B. L. Sherin, How students understand physics equations, Cogn. Instr. **19**, 479 (2001).

[64] 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).

[65] A. Izsak, Students' coordination of knowledge when learning to model physical situations, Cogn. Instr. **22**, 81 (2004).

[66] J. Gainsburg, The mathematical modeling of structural engineers, Math. Think. Learn. **8**, 3 (2006).

[67] R. Hall and R. Stevens, *The Cultures of Computing* (Blackwell Publishers, Oxford, U.K., 1995), pp. 118–145.

[68] G. Hatano and K. Inagaki, *Child Development and Education in Japan* (Freeman, New York, 1986), pp. 262–272.

[69] M. Wertheimer, *Productive Thinking* (Harper, New York, 1959).

[70] E. F. Redish and K. A. Smith, Looking beyond content: Skill development for engineers, J. Eng. Educ. **97**, 295 (2008).

[71] E. Kuo, M. M. Hull, A. Gupta, and A. Elby, How students blend conceptual and formal mathematical reasoning in solving physics problems, Sci. Educ. **97**, 32 (2013).

[72] http://hdl.handle.net/1903/12947

[73] C. Henderson, E. Yerushalmi, V. H. Kuo, P. Heller, and K. Heller, Grading student problem solutions: The challenge of sending a consistent message, Am. J. Phys. **72**, 164 (2004).

[74] J. Docktor (private communication).

[75] D. Hestenes, *Modeling Students' Mathematical Modeling Competencies* (Springer, New York, 2010), pp. 13–41.

[76] K. Vanlehn and B. van de Sande, *Development of Professional Expertise* (Cambridge University Press, New York, 2009), p. 356.

[77] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**, 64 (1998).

[78] L. C. McDermott and E. F. Redish, Resource Letter: PER-1: Physics education research, Am. J. Phys. **67**, 755 (1999).

[79] L. C. McDermott and P. S. Shaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 1998).

[80] E. F. Redish and D. Hammer, Reinventing college physics for biologists: Explicating an epistemological curriculum, Am. J. Phys. **77**, 629 (2009).