Editors' Suggestion

Solution evaluation strategies used by first-year physics students

Abolaji R. Akinyemi
Department of Physics and Astronomy, University of Minnesota, Duluth, Minnesota 55812, USA

Michael E. Loverude
Department of Physics, California State University Fullerton, Fullerton, California 92831, USA

John R. Thompson
Department of Physics and Astronomy and Maine Center for Research in STEM Education, University of Maine, Orono, Maine 04469, USA

(Received 28 March 2023; revised 24 February 2025; accepted 18 June 2025; published 11 August 2025)

One expected outcome of physics instruction is that students develop quantitative reasoning skills, including strategies for evaluating solutions to problems. Examples of well-known "canonical" evaluation strategies include special case analysis, unit analysis, and checking for reasonable numbers. We report on responses from three tasks in different physics contexts prompting students in an introductory calculus-based physics sequence to evaluate expressions for various quantities: the velocity of a block at the bottom of an incline with friction, the final velocities of two masses involved in an elastic collision, and the electric field due to three point charges. Responses from written ($N = 580$) and interview ($N = 18$) data were analyzed using modified grounded theory and phenomenology. We also employed the analytical framework of epistemic frames. Students' evaluation strategies were classified into three broad categories: consulting external sources, checking through computation, and comparing to the physical world. Some of the evaluation strategies observed in our data, including canonical as well as noncanonical strategies, have been reported in prior research on evaluation, albeit sometimes with different names and with varying levels of generalizability. We note four major, general observations prompted by our results. First, most students did not evaluate solutions to physics problems using an approach that an expert would consider an evaluation strategy. Second, many students used evaluation strategies that emphasized computation. Third, many students used evaluation strategies that are not canonical but are nonetheless useful. Fourth, the relative prevalence of different strategies was highly dependent on the task context. We conclude with remarks including implications for classroom instruction.

DOI: 10.1103/rsh8-1ftk

I. INTRODUCTION

Solving problems is a core task in both physics instruction and professional practice. One important skill that physics students are expected to develop is the ability to evaluate the solution to a problem. In physics, evaluation can be defined as checking to make sure that the solution of a problem obeys the laws of physics, is reasonable, and satisfies the constraints relevant to the context of the problem [1]. Examples of evaluation strategies include performing unit analysis, considering limiting cases, using approximations, predicting the effects of changes in problems, and identifying errors in solutions [2]. The ability to evaluate a solution is often cited as an example of what it means to "think like a physicist" [3].

To show how evaluation strategies are used and embody mathematical and physical reasoning, consider the masses on a frictionless pulley in an Atwood's machine. While examining the validity of an expression for the acceleration of the masses, first, a student could ensure that the units of the expression reduce to those of the quantity of acceleration (m/s^2). The student could perform a special case analysis. For instance, the student could verify that if the two masses on the pulley are equal, the expression does not diverge or become infinite, and the acceleration is exactly zero. They could also ensure that if one of the masses is removed, the expression reduces to the free-fall acceleration (g).

Despite agreement about their importance, there is relatively little research on how, when, and where students learn and employ evaluation strategies. Prior studies in physics education research (PER) have documented that few students spontaneously evaluate solutions [4], even at the upper

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

level [5–7], but few studies have probed whether and how students evaluate solutions when prompted [2,8–11]. The framing of evaluating a solution as something specific to physics may neglect parallels to adjacent fields. Further, most of the previous studies have not explicitly connected their findings to theoretical work on student reasoning and mathematical modeling in PER or other disciplines.

To add to the body of knowledge on the use of evaluation strategies in physics, we conducted a study in which students were provided with a physical context and a proposed expression for a relevant quantity and asked how they would go about checking to see whether the expression was reasonable. This work explores evaluation strategies as an avenue for students to find connections between mathematical operations, physics concepts, intuition, and lived experience, joining other studies at the confluence of mathematics and physics as it probes how students ground their use of mathematics in physics when using evaluation strategies. We seek to collect and characterize empirical data to document the practices of students at various levels when prompted to evaluate proposed solutions, irrespective of the context in which they are employed. For this reason, our study includes tasks in multiple physics contexts and courses. However, in this report, we describe data collected from first-year students in introductory physics courses, categorizing student responses into different strategies. Some results are consistent with previous reports, but we also document student approaches to evaluation strategies that have not been previously reported.

Section II motivates and explores the significance of evaluation from the perspectives of problem solving in physics, the use of mathematics in physics, and mathematical modeling. We also summarize previous research on evaluation strategies and articulate our research questions. Section III describes the methods of our study, including the research design, methods, and data analysis. Section IV describes the results and categories resulting from our analysis. In Sec. V, we discuss our results and interpret them through the framework of mathematical reasoning in physics. We conclude with some general insights and implications for teaching.

II. BACKGROUND

Evaluation is an important aspect of critical thinking [12,13] and, consequently, problem solving in mathematics and physics [4,14]. Many models of problem solving in physics education explicitly include evaluation [15,16]. For instance, according to Wright and Williams [15], problem solving involves describing what is happening in the problem, isolating the unknown, substituting in the knowns, and evaluating the derived solution. Even when evaluation is not listed explicitly in a rubric, it is acknowledged as an important step in problem solving [4,17,18]. Unsurprisingly, many models of problem solving in mathematics education also explicitly include evaluation [14,19]. According to Polya [19], problem solving involves understanding the problem, devising a plan to solve the problem, carrying out the plan, and looking back to verify that the solution is reasonable for the given problem. The following includes discussions of research on mathematical reasoning in physics and research on student evaluation strategies in physics education research.

A. Mathematical reasoning in physics

Mathematical reasoning in physics has been described explicitly in the literature as, e.g., a cycle in which students model a physical system mathematically, process the mathematics, and then interpret results; the ability to describe the laws of the physical world using mathematical representations (typically referred to as "mathematization") and to determine the physical consequence of mathematical manipulations of these laws [20]; an organizing structure for analysis of student solutions, particularly with respect to the application of mathematical tools [21]. All three of these perspectives include explicit steps corresponding to checking the solution for physical and/or mathematical validity.

Because problem solving using mathematics and mathematical reasoning manifests in multiple approaches, each of which uses different sets of tools and skills, students' perception of the task at hand can determine their approach to a problem. "Framing" describes structures of expectations that dictate how individuals interpret situations or events [22]. Bing and Redish used the lens of epistemological framing to classify students' justifications of their approaches to solving physics problems [23], identifying four epistemological frames that students activated when confronted with a problem: *invoking authority*, *calculation, physical mapping*, and *mathematical consistency*. In *invoking authority*, students cited an authoritative source, e.g., a physics professor or textbook, typically accompanied by recall and the absence of extended chains of reasoning. In *calculation*, students argued that algorithmically following a set of established computational steps leads to a dependable result. This frame is characterized by a focus on technical correctness and mathematical formalism. Students in *physical mapping* claimed that a mathematical representation of a physical system should correctly characterize the physical scenario it is meant to describe. This frame is also characterized by the presence of extended chains of reasoning and attachment of physical information to symbols, signs, and operations. Finally, mathematical concepts were the underlying structure of different physical scenarios in *mathematical consistency*, which is also characterized by analogies with mathematical ideas.

According to Bing and Redish, a critical part of the novice-to-expert transition in physics is learning to integrate different kinds of knowledge into the solution of a problem [23,24]: the more sophisticated and expertlike

students become, the more flexible students are in their framing and the more likely they are to develop hybrid frames.

B. Physics education research on evaluation strategies

A few PER studies have explicitly explored student use of validity checks or evaluation strategies [2,6–11]. In a study using the ACER framework, Wilcox and Pollock reported that very few students spontaneously evaluated their results (reflection is the R in ACER) [5]; those that did seemed to attend to physical connections (e.g., limiting case analysis or unit analysis). Loverude asked physics students in a mathematical methods in physics course to evaluate whether expressions for the acceleration of masses on an Atwood's machine were correct and to explain how they knew [2]. About 10%–15% of the students ($N \sim 50$) attempted to solve the problem, while 20%–50% cited the presence or absence of a variable in the expression. Roughly 15% of the students described the physical mechanism at play in the physical situation and less than 5% employed limiting case analysis.

Other studies have focused on instructional approaches. Warren [8] reported that learning evaluation strategies made students better self-evaluators and problem solvers. Sikorski *et al.* [9] introduced students in a junior-level electricity and magnetism course to three validity checks: unit analysis, limiting cases, and using reasonable numbers. They found that students checked units most frequently, a finding consistent with the prior report by Warren [8] and a later report by Burkholder *et al.* [11]. Sikorski *et al.* reported that while checking limiting cases, most students gave a mathematical form of the limits without explaining why they expected the physical system to act in a particular way. In addition, students had a difficult time with reasonable values because they lacked intuition for what values were reasonable in the context of electricity and magnetism. Lenz *et al.* suggested that students need explicit instruction on how to perform special case analysis if they are expected to use it to evaluate solutions [10].

White *et al.* [25] studied students' use of limiting case analysis (LCA) in the context of an intermediate electricity and magnetism course, using an adaptive expertise framework, and found that students could develop and refine intuition from LCA. The authors also observed significant variation in which variables students used for their LCA. Students seldom chose constants, and familiar quantities were more likely to be chosen, perhaps due to efficiency.

As part of a broader study demonstrating a paradigm for assessment they call "calculation-concept crossover," emphasizing coherence between formal mathematical manipulation rules and conceptual understanding, with "conceptual" referring to either mathematical or physical concepts, Kuo *et al.* administered an evaluation task to students asking whether a provided expression for the angular speed of a ballistic pendulum postcollision was plausible, finding that ∼20% more students used conceptual arguments to check their results in the mathematical sensemaking-focused class than in the control group [26]. They attribute this to students' increased attention to, and expectations for, coherence in the former class.

In a study similar to that of Loverude [2] and to this study, Burkholder *et al.* [11] administered two evaluation tasks to students with questions in two formats. The "verify" format specifically asked students whether the dependence of a particular variable (e.g., the angle of rope holding up a shelf) and the units of the expression make sense, while in the "contrasting cases" format, two different expressions were provided and students were asked which, if either, was correct. The researchers identified strategies including: "identifying components," where students identified force or torque components related to a particular direction, e.g., height; "identifying functional relationship," where students noted the presence or absence of a variable; calculating torque on the system to evaluate solutions; and "do full solution," for which students attempted to rederive the answer. The authors grouped strategies by sophistication, with "evaluating limit" at the high end and "identifying functional relationship" at the low end, arguing that more sophisticated strategies were tried only after less cognitively demanding ones [27].

Evaluation strategies incorporate mathematical reasoning. Loverude asserts that tasks such as unit analysis, testing expressions with limiting cases, using approximations, identifying errors in solutions, and predicting the effects of changes in problems can help students develop mathematical reasoning skills and tie mathematical reasoning to physical contexts [2]. Kuo *et al.* argue that checking solutions is a form of mathematical sensemaking in which students make sure that calculations and conceptual arguments are consistent [26]. Burkholder *et al.* [11] note that some problem-solving strategies involve an "understanding of mathematical relationships between quantities." All three models of mathematical reasoning described in Sec. II A. emphasize and illustrate the grounding of mathematics in a physical system. All three also include at least one step of sensemaking, connecting physical meaning with mathematics.

C. Research on evaluation strategies from other disciplines

Prior PER studies have examined evaluation within PER, but few have examined related scholarship in other fields. Though there are few explicit discussions of evaluation strategies, an examination of the literature suggests several parallel strands of research in related disciplines. To characterize evaluation strategies as an aspect of student behavior during problem solving across different disciplines, we briefly describe a study in the undergraduate mathematics education (RUME) community relevant to this work.

Carlson *et al.* described a general strategy for reasoning with expressions, known as covariational reasoning, as the cognitive activity that entails coordinating two varying quantities while considering the ways in which they change in relation to each other [28]. Carlson's covariation framework includes five mental actions—stages of covariation that are increasingly sophisticated. These mental actions include mental action 1 (MA1), a simple coordination of two variables (e.g., y changes as x changes); (MA2) coordinating the direction of change of one variable with changes in some other variable (e.g., y increases with increasing x); and (MA3) coordinating the amount of change of one variable with changes in some other variable (e.g., plotting points) [29]. Jones recently extended this framework to include reasoning about coordination of quantities in multivariable expressions, which he labeled multivariation [30].

D. Summary of prior work and research questions

In summary, models of problem solving in both physics and mathematics claim that evaluation is important. However, most research in problem solving has focused on deriving the right result or the correct application of physics concepts, how students use mathematics during problem solving, and how students think about different questions that use the same underlying physics concepts [17,21,31–34]. However, research on the use of mathematics in physics has largely focused on helping students attach the correct physical processes to corresponding mathematical tools but not on evaluation. Consequently, there is little research on what students do when explicitly prompted to evaluate solutions.

To focus on students' understanding and use of evaluation strategies, we explore the following research questions:

1. What evaluation strategies do students use when prompted?
2. To what extent are existing frameworks for problem solving and reasoning from PER consistent with students' use of evaluation strategies?

The primary purpose of this work is to explore the use of evaluation strategies as a tool for helping students meld their knowledge of mathematics and physics productively. Our goal is to add to the effort toward understanding how students develop mathematical reasoning by examining how evaluation strategies can help students consolidate their physics and mathematics knowledge, thus making them better problem solvers, self-learners, and physicists.

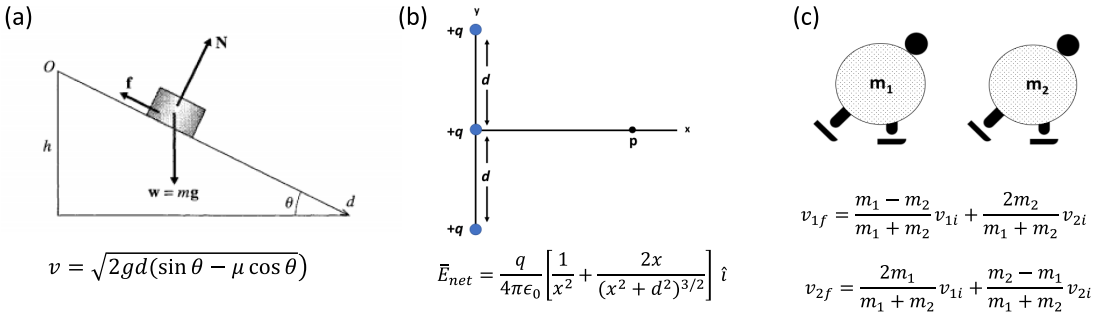
III. METHODS

In this section, we outline our research methodology, including data collection and analysis, and describe the reasoning behind the design of the research tasks.

A. Project design and development

The aim of our project is to explore students' use and understanding of evaluation strategies as sensemaking tools in physics, irrespective of the context where they are employed. To this end, we designed a number of tasks that prompted students to evaluate the solution to a physics problem. We intended that students' responses to the tasks would (a) generate a list of evaluation strategies that students use in different physics contexts and (b) include the use of expert evaluation strategies, including special case analysis, unit analysis, and using reasonable numbers. Tasks were developed for contexts across the physics curriculum; in this report, we focus on three such tasks that were posed in introductory physics courses, which are summarized in Fig. 1. The tasks in the scenarios posed to students were adapted from back-of-the-chapter exercises in the assigned textbooks of the courses associated with each task, e.g., Knight [35] for the tasks in Fig. 1. (We will examine performance across different course levels in a subsequent publication.)

The general layout of each task was as follows: First, a physical scenario and a corresponding problem were described. Next, one or more expressions were provided for relevant quantities that would arise in the solution to the problem posed in the scenario. Finally, students were asked



$$v = \sqrt{2gd(\sin\theta - \mu\cos\theta)}$$

$$\vec{E}_{net} = \frac{q}{4\pi\epsilon_0}\left[\frac{1}{x^2} + \frac{2x}{(x^2+d^2)^{3/2}}\right]\hat{\imath}$$

$$v_{1f} = \frac{m_1 - m_2}{m_1 + m_2}v_{1i} + \frac{2m_2}{m_1 + m_2}v_{2i}$$

$$v_{2f} = \frac{2m_1}{m_1 + m_2}v_{1i} + \frac{m_2 - m_1}{m_1 + m_2}v_{2i}$$

FIG. 1. Figures and given expressions for the assigned tasks: (a) the velocity of a block at the bottom of an incline with friction; (b) the electric field at a point some distance from three point charges of equal magnitude; (c) the final velocities of two masses in an elastic collision. For each task, students were asked, (i) "Without knowing the correct answer, how would you go about checking if your solution is reasonable?" and (ii) "Using the approach(es) you described in part (i), determine whether or not this is likely the correct result."

TABLE I. Summary of previous studies on evaluation strategies. This paper focuses only on introductory students, but a subsequent paper will consider strategies used by more advanced students.

Study	Data type (W = written, I = interview)	Course level; after evaluation instruction?	Prompted	Cues	Tasks
Warren [8]	W	Intro; Y	No	No	Multiple
Loverude [2]	W, I	Upper; N	Yes	No	Single
Sikorski *et al.* [9]	W	Upper; Y	No	No	Multiple
Lenz *et al.* [10]	I	Intro; N	Yes	Not rederive	Single
Burkholder *et al.* [11]	W	Intro; Y	Yes	Units/limiting	Single
White *et al.* [25]	W	Upper; Y	Yes	Yes	Multiple
Kuo *et al.* [26]	W	Intro; N/Y	Yes	Not rederive	Single (part of larger study)
Akinyemi *et al.* (This report)	W, I	Intro; N	Yes	No	Multiple

how they would go about checking whether the expressions were reasonable and to execute the strategies that they described.

The choice to provide students with an expression and prompt them to evaluate it was intentional. First, providing the expression made the time to complete the task shorter and avoided the possibility of conceptual or procedural difficulties in arriving at the solution that were beyond the scope of our study. More importantly, prompting allowed us to focus on how students evaluate expressions, given that students do not generally do so spontaneously [4]. This bypassed the question of whether students who are able to evaluate simply choose not to. We note that other studies in which students were explicitly asked to evaluate the solution of a problem [9–11] have been published since this study was designed.

One limitation of this choice is that the prevalence of evaluation strategies in our study is perhaps not representative of the behavior of students in a natural problem-solving setting. In addition, since students did not go through the steps of generating the solution, their responses may not reflect some strategies that might have been suggested by the problem-solving process itself. In contrast to others, our study did not intend to cue any particular evaluation strategies.

To illustrate how our study design is positioned relative to others in PER, we have summarized the parameters of this and related work in Table I.

We chose physics problems with symbolic expressions that could be easily evaluated using special case analysis and unit analysis. Symbolic solutions are the norm at the upper level of the physics curriculum, allowing for continuity of tasks across levels. However, there are two consequences of the choice to use symbolic expressions. First, most problems given to first-year students have arithmetic solutions (i.e., the solution is a number). As a result, the symbolic expression we provided may have lent itself to symbol-based strategies that first-year students might normally not use. Second, the symbolic nature of the tasks may have made them more adaptable to expertlike strategies like unit analysis and special case analysis.

Questions were chosen with expressions that would require students to enact complex or multistep solutions to reproduce, to dissuade students from solving for the expression again, and to prevent recalling a "correct answer." The choice to ask students to check whether the given expression was "reasonable" was also purposeful. We chose not to ask the students to check whether the given expression was "correct," again to avoid cueing students to rederive the provided expression. We reasoned that the word "correct" might cue students into thinking that the task prompted them to confirm that the provided result was the exact solution to the posed physics problem. We anticipated that the word "reasonable" would elicit students' use of evaluation strategies. Essentially, we sought to investigate whether students would use expert strategies to evaluate the solutions we provided, and so we designed tasks that would be easily evaluated using the expert strategies of unit analysis, special case analysis, and using reasonable numbers. Finally, the expressions provided to students in these tasks were correct, so as to simplify the analysis, giving an upper bound on certain strategies and increasing the available strategies beyond those that would be used for "obvious" flaws (e.g., incorrect units).

The tasks were administered to students in the calculus-based introductory physics sequence at a public research university in New England. By the time the tasks were administered in both interview and written formats, all participants had covered the relevant physics content in class. Instruction consisted of lectures, traditional laboratories, and conceptual tutorials in recitation. However, lectures were taught by different instructors with varying emphasis on quantitative and conceptual explanations. The courses in which data from the inclined plane ($N = 215$) and point charge ($N = 174$) tasks were collected were taught by the same instructor. The weekly homework in these classes was almost completely quantitative. On the other hand, the course in which the conservation of

momentum task data ($N = 191$) were collected had two sections co-taught by different instructors so that students received similar instruction and assessment. In these courses, weekly homework had both quantitative and conceptual components. Written data were collected in a way that the course instructor thought would optimize participation, including short in-class quizzes with or without an offer of extra credit. Interview subjects were volunteers, solicited in the course of interest. Interview data were also collected in different ways to optimize participation, including offers of cash. Interviews were either individual (13 students) or with pairs of students (5 pairs).

All the interviews were conducted in a structured, think-aloud format. The interviews were conducted by the first author and structured like the written tasks. First, students were briefed about the general structure of the interview and then asked to practice thinking out loud by talking through and showing how a stick figure is drawn. The students were then given a copy of the written task—each interview focused on one context—and then asked to respond to the prompts while thinking out loud. Sometimes, the students were asked follow-up questions to clarify their thoughts. Interviews lasted between 16 and 40 min.

B. Data analysis

The analysis of our data was a multistep, iterative process. First, we coded all written responses on the first task administered (inclined plane) using open coding for the general strategy students used, e.g., *solving for the given expression*, *solving for a known result*, or *plugging in numbers*. We also scored the responses from zero to three in order of clarity, zero being not clear at all, and three very clear. Many students suggested evaluation strategies but did not attempt them. It was not clear whether the students did not know how to execute the evaluation strategy or did not have enough time to implement the chosen strategy. Also, some students treated the tasks as one question, i.e., they listed evaluation strategies that they would use, and implemented said strategies in both parts of the task. Consequently, the responses of both questions of each task were coded together.

Second, we refined the code categories through discussions among the coauthors and a constant comparative coding exercise [36]. During this stage, we compared our codes and classifications to Bing and Redish's epistemic frames [23,24]; this comparison helped define and distinguish the highest-level categories of our codes. In order to characterize evaluation strategies as sensemaking tools in physics irrespective of the context where they are employed, as we worked across the multiple contexts, we sought to create categories not specific to a particular problem context but rather generally applicable.

Third, interviews on the inclined plane task helped further flesh out codes. After coding the first set of (13) interviews on the inclined plane task and agreeing on high- and intermediate-level categories, we went back and revisited the written responses. The insight from the interviews helped clarify some students' responses, decreasing the number of responses initially coded as zero and one. Insight from the interviews also helped us reassign a few responses to code categories to which they were better suited. The interviews were designed to be both a data source and a means to clarify written responses.

Next, we compared our results with previous work on evaluation strategies. This comparison helped us collapse some codes into broader categories based on the overall goal of a student's work. For instance, responses that suggested checking whether the solution was reasonable by plugging numbers into the given expression were then coded as *using reasonable numbers* in accordance with previous work on teaching students how to evaluate [8,9].

As our project progressed, student responses from interviews and written responses on subsequent tasks helped distinguish between similar codes and also introduced new ones. Whenever a new code emerged, previously coded data were re-examined for responses consistent with the new code.

IV. RESULTS

In this section, we report results from written and interview responses at the introductory level of the physics curriculum. We also present the relative prevalence of strategies across different tasks.

A. Evaluation strategies observed

We broadly classified the evaluation strategies observed in our data into three categories: *comparing to the physical world, checking through computation*, and *consulting external sources*. One aim of our study is to describe evaluation strategies as sensemaking tools in physics, irrespective of the context in which they are employed. Consequently, these categories cut across the different physics contexts of the tasks we administered. Strategies in the *comparing to the physical world* category involve evaluating the given expression by checking whether it is consistent with physics knowledge, experience, and intuition. Strategies in the *checking through computation* category involve evaluating the given expression using computation without interpreting the physical meaning of the given expression. Finally, strategies in the *consulting external sources* category involve evaluating given expressions by checking with a trusted external source. The first two categories have multiple subcategories. While there were some variations on *consulting external sources*, we did not identify subcategories because this code was not common in our sample. Table II describes the three main categories of strategies.

TABLE II. Descriptions of the three categories of evaluation strategies.

Evaluation category	Description
Comparing to the physical world	Checking for consistency between the expression and one's understanding of the physical world
Checking through computation	Checking using strategies that emphasize computation, e.g., mathematical procedures and algebraic manipulation
Consulting external sources	Consulting an external source (e.g., class notes, textbook, online search) or person (e.g., TA, professor) as an authority on correctness

Each of the top-level categories can be mapped to one of Bing and Redish's epistemological frames. Evaluation strategies in the *comparing to the physical world* category are consistent with Bing and Redish's *physical mapping* frame [23]. When using this group of strategies, students support their arguments by pointing to the quality of fit between the mathematical symbolic representation (the given expression) and the physical situation it is meant to describe. This group of strategies also involves the use of extended chains of reasoning and entails attachment of physical information to symbols, signs, and operations. Similarly, strategies in the *checking with computation category* are generally consistent with the *calculation* frame: students rely on algorithmically following a set of established computational steps to lead to a trustable result. These strategies are also characterized by a focus on technical correctness and include minimal connection between variables and the physical quantities they represent. Finally, the *consulting with external sources* category is consistent with Bing and Redish's *invoking authority* frame, as it treats the information as accurate. A common feature of responses in this category is the absence of extended chains of mathematical reasoning.

Figure 2 shows the different strategies and the classification scheme. For each category and subcategory, we present the strategies roughly in reverse order of sophistication, from most to least sophisticated. For each subcategory observed, we describe defining attributes and key features of corresponding responses.

We refer to the strategies of *special case analysis*, *unit analysis*, and *use of reasonable numbers* as "canonical evaluation strategies" and these three strategies are filed under the *comparing to the physical world* category. *Special case analysis* includes evaluation of limiting cases (labeled as "limiting" by Loverude [2] and as "evaluating limits" by Burkholder *et al.* [11]). We describe the student strategy as unit analysis, as distinct from dimensional analysis [37,38]: the former is a simple check of units (e.g., kg, m, s) while the latter can have more technical meaning. These three canonical strategies have been the focus of prior studies on teaching and learning evaluation strategies. Thus, in describing code categories below, we provide minimal details for these canonical strategies.

Furthermore, in the rest of this paper, we define computation as the act of performing "mathematical moves,"

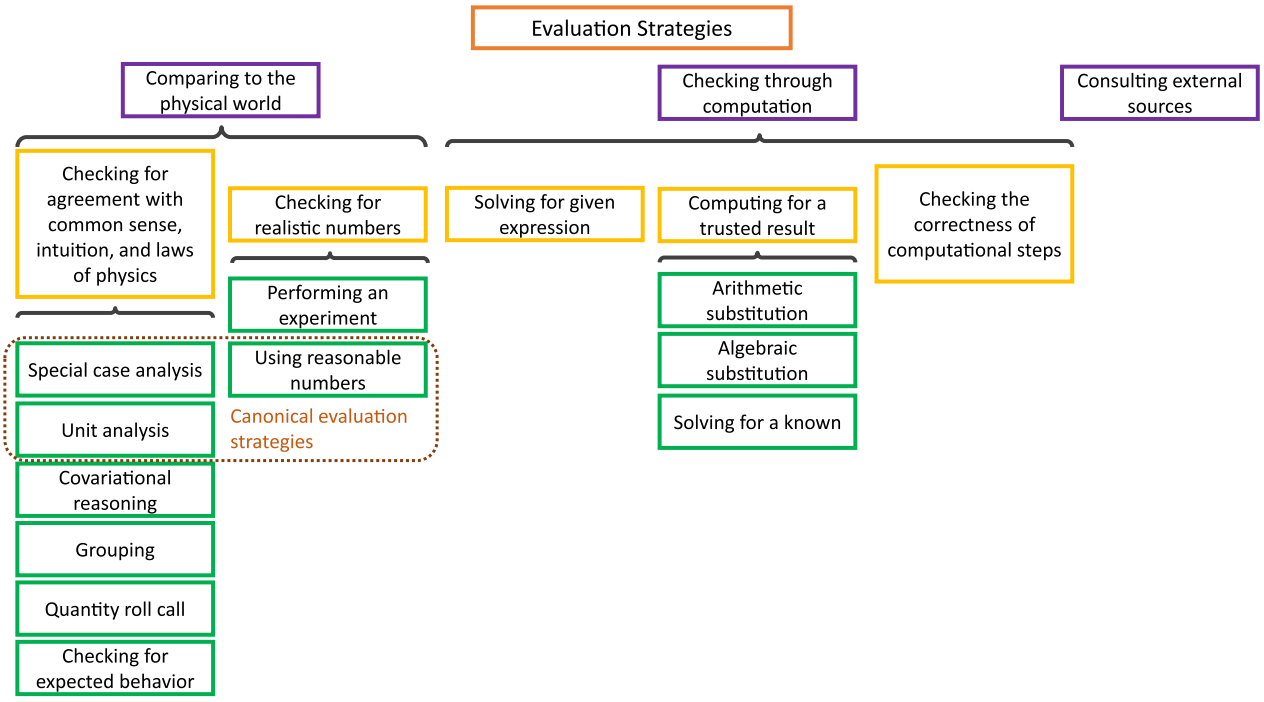


FIG. 2. Category structure of identified evaluation strategies. Categories (purple) can have several subcategories (yellow), which in turn include individual strategies (green).

procedures, and operations such as algebraic manipulation or taking derivatives [34,39,40]. We define algebraic processes as computation that involve manipulation of mathematical symbols, and arithmetic processes as computation that involve manipulation of mathematical expressions with numbers plugged in for variables. We assume competence in the use of mathematics in physics to include knowledge of both computational procedures and mathematical concepts.

While most of the evaluation strategies observed in the data were observed in responses to all of the tasks at the introductory level, a few of them are specific to only one task. However, to explain each code category, we primarily use examples from responses to the inclined plane task, for easier comparisons across categories. A comprehensive list of codes and corresponding sample responses can be found in Supplemental Material [41].

1. Comparing to the physical world

The evaluation strategies in this category involve judging the consistency of the expression with a student's understanding of the physical world. During the implementation of these strategies, students may have used symbolic representations or numerical values of physical quantities. Evaluation strategies using numerical values involve attaching physical meaning to and extracting physical meaning from a numerical answer. This category has two subcategories: *checking for agreement with common sense, intuition, and laws of physics* and *checking for realistic values*; each of these in turn has two or more subcategories. These strategies and their descriptions are summarized in Table III.

a. *Checking for agreement with common sense, intuition, and laws of physics*. This subcategory includes six strategies, two of which are canonical; the focus is on evaluation of the given expression by checking for an agreement between the given expression using lived experience, intuition, and physics concepts. In contrast to the following category, these generally focused on symbolic expressions and quantitative reasoning rather than on numerical values. On the inclined plane task, about 37% of students gave responses that were coded in this subcategory.

(i) *Special case analysis (canonical)*. The *special case analysis* strategy involves checking the behavior of an expression when specific variables are taken to extreme or physically significant values. As described in previous literature [8–11], *special case analysis* involves checking whether the given expression is consistent with real world behavior and the laws of physics under certain specific physical or corresponding mathematical conditions. This group includes students who merely suggested this strategy and those who also predicted what they expected to physically happen at the chosen conditions or limits. We also observed responses that we characterized as arithmetic versions of this strategy. In the arithmetic version, students used numbers in place of the variables in the expression and interpreted the numerical answer in the physical context of the special case being probed. An example is shown in Supplemental Material [41].

(ii) *Unit analysis (canonical)*. A second canonical evaluation strategy, *unit analysis*, is also described in prior work. In *unit analysis*, students evaluated the given expression by checking whether the expression has expected dimensions or units, a canonical strategy identified in previous studies [8–11]. As in *special case analysis*, there were more arithmetic versions of this strategy (see Supplemental Material [41]).

TABLE III. Summary of evaluation subcategories and strategies in the comparing to the physical world category. Strategies marked with an asterisk (*) are labelled as canonical strategies.

Comparing to the physical world: Checking for consistency between the expression and one's understanding of the physical world	
Evaluation subcategory/strategy	Criterion for validity (The expression …)
a. Checking for agreement with common sense, intuition, and laws of physics (Sec. IV A 1 a)	Is consistent with lived experience, intuition, and physics concepts
i. Special case analysis*	Predicts appropriate behavior under special physical conditions
ii. Unit analysis*	Has expected units
iii. Covariation	Shows expected variation between two variables
iv. Grouping	Contains physically significant groups of variables
v. Quantity roll call	Includes expected relevant variables
vi. Checking for expected behavior	Predicts the expected behavior of the system (using a strategy not listed above)
b. Checking for realistic numbers (Sec. IV A 1 b)	Generates physically realistic values for the quantities it contains
i. Performing an experiment	Is consistent with experimental results (imagined or actual)
ii. Using reasonable numbers*	Yields plausible numerical values when computed (with appropriate inputs)

(iii) *Covariation*. The third strategy involves evaluating the given expression by considering an expected variation between the one variable and another in the given expression to check the behavior of the expression. We labeled this strategy *covariation* because it shares the characteristics of covariational reasoning, which is widely used in RUME to describe student reasoning [29]. Unlike the previous two strategies, covariation has received relatively little attention in physics (one example is Zimmerman *et al.* [42]). While covariational reasoning is prevalent in physics reasoning, it has not been documented explicitly as an evaluation strategy. Examples include:

> The velocity increases with the θ increasing which makes sense.

> Solve for velocity again with a value of θ that's close to the original value. If the velocities are similar, then that would prove it's a reasonable answer.

In the first example, the student evaluated the expression by citing that it has the expected covariation between the velocity and the angle of the incline θ, indicative of MA2 [29]. Unlike many mathematical examples, the expression has multiple variables, so the student has chosen θ and v. This example also has characteristics of "dependent multivariation," one of the subcategories of multivariation as identified by Jones [30]. Likewise, in the second example the student stated that the velocities corresponding to similar values of θ should be close. Note that this student did not state the direction (increase or decrease) of expected change in either the velocity or angle of incline but did imply that a change in θ would lead to a coordinated change in the velocity. This type of covariational reasoning is most consistent with MA1 and will be further discussed in Sec. V. As with the previous strategies, students employed an arithmetic version of this strategy as well.

(iv) *Grouping*. Several students made sense of the given expression by identifying a group of symbols within an expression and describing the physical significance of the group. We labeled this *grouping*. Often, groups were separated by mathematical operators or the equal sign like covariation, grouping, has not been documented explicitly as an evaluation strategy. An example of a response coded as *grouping* is:

> $v_f{}^2 = 2gd(\sin\theta - \mu\cos\theta)$. d is the distance of the ramp (Δx). I'm guessing $g \bullet (\sin\theta - \mu\cos\theta)$ is the positive acceleration. The $\sin\theta \bullet g$ would be for a frictionless incline. Because there is friction, the force of friction can be $\mu\cos\theta$ (because $\cos\theta = N$ in this case). So, I guess this makes sense.

This student explicitly identified three groups within the given expression: total acceleration $g \bullet (\sin\theta - \mu\cos\theta)$, acceleration for a frictionless incline $(\sin\theta \bullet g)$, and the "force of friction" $(\mu\cos\theta)$, and checked their physical meanings to validate the equation.

(v) *Quantity roll call*. This strategy involves checking whether expected quantities were present in the given expression as symbols. The quantities the students focused on were those they expected to be relevant in the given task. Examples are:

> Does it have μ, θ, g, and d in it?

> I would […] look to see if it included all the parts I need to be able to calculate velocity and looking at the equation I notice it does not include mass, so I might determine that it might not give me a correct answer.

The first student here listed the quantities they expected the given expression to contain, but not any particular relationship between these variables and/or the velocity. The second student expected mass to be in the equation and cites its absence as a reason why the given expression might not be reasonable. The fact that *quantity roll call* accounts for the mere presence of a quantity, and not for whether or how the quantity affects another quantity in the given expression, distinguishes it from *covariation*.

(vi) *Checking for general expected behavior*. This code was assigned to responses in which students evaluated the given expression by probing aspects of the expected behavior of the physical system without using one of the strategies above. Responses in this category were often incomplete and/or focused on a single aspect of the expression, such as its sign or direction. A sample student response is,

> Make sure the solution is negative since the block is sliding in the $-y$ direction.

The seeming vagueness of the category also results from our desire for categories to be context independent, as many responses were associated with specific tasks. For example, responses citing the direction of the electric field in the point charge task were coded under this category; responses stating that momentum or energy should be conserved fall into this category for the conservation of momentum task. We suggest that context-specific categories identified by other authors, e.g., the *calculating torque* category from Burkholder *et al.* [11], would fit in this category.

b. *Checking for realistic numbers*. The second major subcategory of *comparing to the physical world* included responses focused on checking that specific values of the given physics term or variable were realistic: since the expression was supposed to describe the real world, the result of the expression for appropriate inputs should be "lifelike." As noted below, not all responses provided operational definitions of terms like "feasible" or "reasonable." The responses in this category are generally focused on the values of quantities rather than on their variation, in contrast to arithmetic versions of categories above. This subcategory includes two subcategories: *performing an*

experiment and *using reasonable numbers*. On the inclined plane task, about 25% of students gave responses coded in this subcategory.

The strategies in this subcategory differ from the arithmetic versions of strategies above. Strategies in the *checking for realistic values* use numbers to check whether the given expression is feasible or physically realistic. However, the arithmetic versions of strategies in the *checking for agreement with common sense, intuition, and laws of physics* subcategory are aimed at checking for consistency between the given expression and either lived experience or predictions generated with the laws of physics.

(i) *Performing an experiment*. Responses coded as using this strategy involved suggesting that the given expression could be evaluated by performing an experiment of the situation depicted in the task to see whether the outcome is reasonable for the physical context. An example response is,

> Without knowing the correct answer, just do an actual experiment, measure the velocity, using known values check to see if the results match with what you got.

This response included a suggestion to perform an experiment and to compare the result of the experiment with the velocity calculated using the given expression.

(ii) *Using reasonable numbers (canonical)*. This strategy, which is one of the canonical strategies, involved evaluating an expression by substituting "reasonable" numbers (as determined by the student) into the expression and checking whether the result is also reasonable. This category also includes responses that suggested checking the magnitude of the numerical value of the given quantity to see whether it was feasible given the context of the task.

2. Checking through computation

This category contains strategies that emphasize computation, which we define as the act of performing "mathematical moves," procedures and operations such as algebraic manipulation or taking derivatives. *Checking through computation* is a broad category covering a spectrum of responses. As a result, we divide this category into three subcategories: *solving for the given expression*, *computing for a trusted result*, and *checking the correctness of computational steps*. These strategies and their descriptions are summarized in Table IV.

a. *Solving for the given expression*. This subcategory of response involved solving the original problem from first principles rather than evaluating the given expression. Responses in this subcategory have been present in all tasks; this subcategory was the most frequently coded at the introductory level. In addition to a single rederivation of the expression, this subcategory includes responses in which students suggested (and/or attempted) to obtain the given expression using two different methods. On the written inclined plane task, about 54% of students gave responses coded in this subcategory; additionally, 9 of 13 interviewed students similarly considered starting over on this task. On the other hand, about 7% of the written responses involved students comparing the value of or expression for v obtained by two methods, e.g., comparing the velocity obtained from conservation of energy to velocity derived using kinematics. Examples are:

> Splitting up the forces into x and y axis, then using $\sum F = ma$ and using $a = \frac{v\frac{m}{s}}{ts}$ to find the velocity.
> $\sum F_Y = mg\sin\theta - N\mu\sin\theta$.
> $\sum F_y = \sin\theta(mg - N\mu)$.

> I would check if my solution is reasonable by first determining the velocity of the block using a different verified equation.

The first student described solving for the velocity from first principles: first using Newton's second law to determine the acceleration, and then using kinematics to solve for the velocity using the already calculated acceleration. The second response was coded as using *solving for the given expression* because it included the suggestion to solve for the given expression using another means ("a different verified equation").

Finally, a few responses in this subcategory proposed integrating the given expression to get an "original equation." Responses in this category were difficult to interpret until interview results showed that some students interpreted the word "derived" in the task prompt as "took the derivative of," i.e., differentiated. This clarified responses in which students suggested integrating to get the original (given) expression for velocity:

> You could check your answer the derivative by taking the integral of $v = \sqrt{2gd(\sin\theta - \mu\cos\theta)}$ and seeing if you get the original equation $\int (2gd(\sin\theta - \mu\cos\theta))^{\frac{1}{2}}dv$.

b. *Computing for a trusted result*. Responses in this subcategory involved performing mathematical operations (algebraic or arithmetic) to see whether the mathematical operations produced a known result. All responses in this subcategory entailed a sizeable amount of mathematical computation aimed at confirming that a law of physics (e.g., conservation of energy) was obeyed or that the value of a constant (e.g., g) was equal to its established or accepted value. These strategies are distinct from those above because responses in this category generally resulted in checking whether two sides of an equation were indeed equal, rather than, say, comparing a numerical result to known quantities. On the written inclined plane task, about 8% of students gave responses coded in this subcategory. There are three strategies in this subcategory: *arithmetic substitution*, *algebraic substitution*, and *solving for a known*.

TABLE IV. Summary of the (subcategory) strategies in the "checking through computation" category.

Checking through computation: Checking using strategies that emphasize computation, e.g., mathematical procedures and algebraic manipulation	
Evaluation subcategory/strategy	Criterion for validity (The expression …)
a. Solving for given expression (Sec. IV A 2 a)	Is equal to the result of solving the problem from first principles
b. Computing for a trusted result (Sec. IV A 2 b)	Can be rearranged to solve for another known expression or constant
i. Arithmetic substitution	Obeys a known law of physics when variables are replaced with numbers
ii. Algebraic substitution	Is still consistent with a known law of physics when substituted into another expression
iii. Solving for a known	Yields a known variable or constant when rearranged
c. Checking the correctness of computational steps (Sec. IV A 2 c)	Is the result of properly following procedures in the derivation process

(i) *Arithmetic substitution*. This strategy involved evaluating the given expression by assigning numbers to quantities in the given expression and using them to check that, numerically, a physics rule or concept holds for the given expression. Here, the trusted rule or physics concept being probed depended on the context of the task. In the inclined plane task, students generally checked for conservation of energy, while in the conservation of momentum task students checked for conservation of momentum, conservation of energy, and adherence to the laws of 1-dimensional kinematics. In context of the conservation of momentum task context, an *arithmetic substitution* response invoking conservation of momentum is shown in Fig. 3(a). This student chose values for the masses and initial velocities of the balls (skaters) and plugged the numbers into the given expression to find the final velocities of the skaters. The student then used the masses and initial and final velocities of the skaters to find the initial and final momenta of the skaters. The student concluded that the given expression is correct because the initial and final momenta of the skaters were equal.

(ii) *Algebraic substitution*. This strategy involves evaluating the given expression by substituting the given expression into another equation to confirm that a trusted physics concept or rule was followed. As with the arithmetic version of this strategy, students generally checked against a trusted physics rule or principle relevant to the context, but the comparison was symbolic rather than arithmetic. Usually, the equations representing these physics rules included the quantity or term that was being evaluated. For instance, in the inclined plane, any law of physics into which students substituted the equation contained v. Likewise, in the conservation of momentum context, the expression to be substituted contained v_{1f} and/or v_{2f}. If the given expression was consistent with a known law or equation of physics, then it was judged to be reasonable. An example of an *algebraic substitution* response from the conservation of momentum task is shown in Fig. 3(b); the student substituted the given expression into the equation for conservation of energy to check if the left- and right-hand sides of the equation were equal.

(iii) *Solving for a known*. This strategy involves evaluating the given expression by using a given quantity or its numerical value to solve for a known quantity or one that is assumed to be given. To determine whether the original expression is reasonable, the student compared the value or expression of the calculated quantity to its given or known value. In an interview on the inclined plane task, a student suggested using the calculated value of the velocity to solve for quantities g (acceleration due to gravity) or d (the length of the incline):

> [Rearranges given expression to solve for g] … I was going to solve for a constant that we already know … so if we plug in the numbers that we supposedly got [for v], we should technically get something around that [points at 9.81] answer.

Note that this strategy uses circular reasoning.

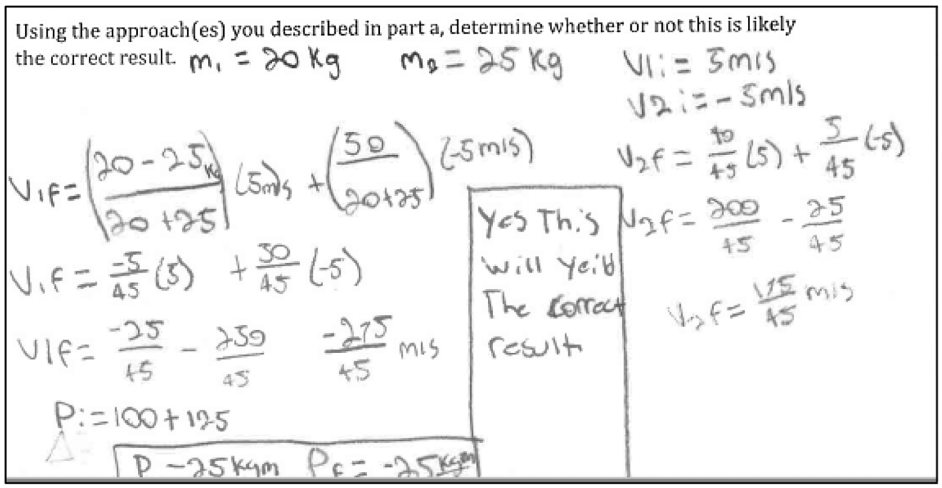
c. *Checking the correctness of computational steps*. In this strategy, students proposed checking the correctness of the steps taken to compute the given mathematical expression. Examples of actions proposed in the responses that used this strategy include checking to see that the correct numbers are plugged into the calculator, ensuring that numbers have the correct sign, and checking that mathematical operations are carried out correctly. Examples include:

> I would check to see if the values of each variable were put into the equation correctly. … ask to see friend's steps throughout his attempt to solve the equation. (in point charge task)

3. Consulting external sources

The third category is characterized by consulting an external source (e.g., class notes, textbook, online search) or person (e.g., TA, professor) as an authority on correctness. Responses in this category were present in written and/or interview data in all three tasks, with a maximum of

(a)



(b)

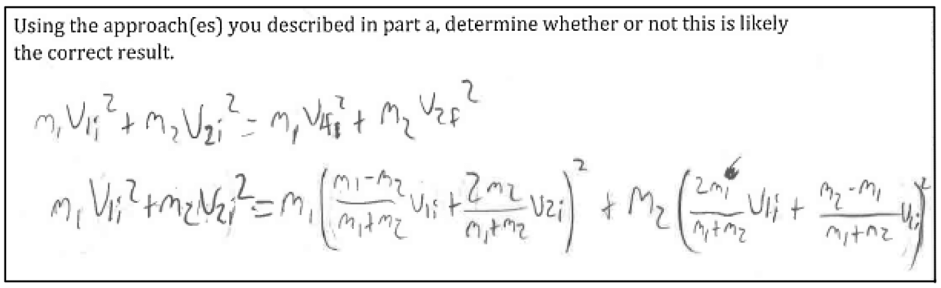


FIG. 3. Examples of student responses from the conservation of momentum task categorized as (a) arithmetic substitution and (b) algebraic substitution.

about 13% of student responses being coded as such on the written point charge task. Examples include:

> I would ask my TA or go to the [Physics Learning Center] for help
>
> … I'd probably check online
>
> I would check my notes to see if the derived expression matched what I learned in class.

B. Prevalence of observed categories

The prevalence of the observed strategies for each task, at the introductory level, is shown in Fig. 4. Note that, as in previously reported studies, multiple codes could be attributed to the same student on one task, so the totals for a given task often add to more than 100%. In general, the majority of strategies used were either computation-based or, we argue, novice versions of more expert strategies that connected the expression to the physical scenario.

The most prevalent subcategories of strategies were *solving for the given expression*, which is in the *checking through computation* category, and *checking for agreement with common sense, intuition, and physical laws*, which is part of the more sophisticated comparing to the physical world category. The major contribution from the latter category was in *checking for expected behavior*, which as mentioned above largely consisted of simple but reasonable checks, e.g., matching signs to directions. In the discussion section, we argue that the latter is a novice version of more expert strategies. The canonical strategies commonly mentioned in prior literature as desirable, namely, *unit analysis, special cases*, and *checking for reasonable numbers*, were not at all common in our dataset, with less than 10% of responses identified with each category for most of the tasks. Overall, only 30%, 7.5%, and 15% of the students used or suggested a canonical strategy on the inclined plane, point charge, and conservation of momentum tasks, respectively. Relatively few students in our sample were categorized as *consulting external sources*; this response appeared primarily in the point charge task.

Of note as well is the wide variation within any single strategy across the three tasks. Some strategies were used only on one task or much more on one task than the others, while other strategies were used on all three tasks but with very different prevalence. As we will report in a subsequent publication, we did not observe any additional strategies in upper-division courses, though prevalence of strategies did vary across levels.



FIG. 4. Prevalence of evaluation strategies used by first-year students in each task (a) across the major categories and subcategories and (b) within the comparing to the physical world category.

V. DISCUSSION

In this section, we first compare the strategies we coded in our data to previous empirical work in PER. Second, we present four observations that summarize our empirical findings and their implications for researchers and instructors. Finally, we compare our empirical results to previously published research on mathematical reasoning in physics.

A. Comparison to previous PER findings

Many of the evaluation strategies observed in our data, including the canonical strategies of *special case analysis*, *unit analysis*, and *using reasonable numbers*, as well as noncanonical strategies of *solving for the given expression*, *covariation*, and *quantity roll call*, have been reported in prior research on evaluation [2,8–11], albeit sometimes with different names.

Loverude coded four main strategies in his "evaluate the expression" tasks administered in a junior math methods course in the context of an Atwood's machine [2]; three of these are identified in the current study as well. Loverude's "solution" code is the same as our *solving for a given expression*; his "variables" code corresponds to *quantity roll call*; and his "limiting" refers to limiting case analysis, part of the *special case analysis* category. His fourth code, "mechanism," overlaps with the other codes, but overlays a description of physical mechanism beyond the mathematical relationship. Loverude also noted the low performance of students in unit identification on "determine the units" tasks for quantities other than distance and time. Burkholder *et al.* [11] found high prevalence of successful use of unit analysis (> 75%) on their verify tasks, in which students were explicitly prompted to consider the units of the expression, and lower prevalence (60%–65%) on their contrasting cases tasks, on which no mention of units was made. They found higher prevalence in checking units of height than of force, consistent with Loverude's earlier findings that unit identification is context dependent. Lenz *et al.* [10] focused on special case analysis, reporting that none of the 11 students in an interview sample spontaneously used this strategy. Instead, students used other strategies, many specific to the inclined-plane context (free-body diagrams, reasoning with friction), which would likely be coded as *checking for expected behavior*. Four of eleven performed a unit check.

Compared to the results of both Loverude's study with upper-division students [2] and the Burkholder *et al.* study [11], a greater fraction of our students attempted *solving for the given expression*. Whereas Warren [8] and Sikorski *et al.* [9] both reported that unit analysis was the most prevalent canonical strategy, our results did not show a similar trend. To our knowledge, the strategies of *arithmetic substitution*, *algebraic substitution*, *solving for a known*, *grouping*, *consulting external sources*, and *performing an experiment* have not been reported in any prior study.

B. General observations

There are four general observations prompted by our results with introductory students. First, most students did not evaluate solutions to physics problems using an approach that an expert would consider an evaluation strategy. Second, many students used evaluation strategies that emphasized computation. As a result, either due to the limitations of their knowledge or in the strategy, many students were unable to complete the evaluation task successfully. Third, many students used evaluation strategies that are not canonical but are nonetheless useful. We also discuss the impact of the task context on the relative prevalence of different strategies.

1. Most introductory students did not use expertlike evaluation strategies

On every task at the introductory level, only a few students used the canonical evaluation strategies, as seen in Fig. 4(b). The most prevalent canonical strategy was *using reasonable numbers*, used mainly on the inclined plane task, by about 20% of students. *Special case analysis* and *unit analysis* were seen at the 0%–10% level, and primarily on the inclined plane and conservation of momentum tasks. This is consistent with the findings of Loverude [2], Lenz *et al.* [10], and Burkholder *et al.* [11], in which only ∼10% of students evaluated limits when not prompted, and at most 27% did so when prompted in a specific format.

Even when students did use canonical evaluation strategies, their implementation was seldom expertlike. For instance, within responses coded as *using reasonable numbers*, there was a range of sophistication in terms of how specific students were about what "reasonable" meant. The response "see if answer feels possible" is not specific about what "possible" means. However, this response is clearer: "you should be able to tell if v is too slow or too fast for example, if $d = 10$ m, a v of $100,000$ m/s wouldn't make sense." While *using reasonable numbers* is a potentially productive direction, its utility is diminished when students do not know what is considered "realistic" or if they lack the physical intuition in some cases. Notably, this strategy was considerably less common in the point charge task, consistent with the findings of Sikorski *et al.* [9] that students have limited intuition about what are reasonable numbers in electricity and magnetism.

2. Many students used evaluation strategies that emphasized computation

Our second observation was that instead of using canonical evaluation strategies, many students attempted to evaluate the given expression using strategies that were based on checking the correctness or accuracy of computational procedures. Even more expert strategies were often implemented arithmetically rather than symbolically. This preference is shown in the frequency of *checking through computation* strategies, e.g., *solving for the given expression, checking the correctness of computational steps*, and *computing for a trusted result*. Specifically, 60%, 67%, and 48% of first-year students used at least one of the evaluation strategies that emphasized computation in the inclined plane task, point charge, and conservation of momentum task, respectively. Essentially, these strategies reflect the notion that if the given expression is correct, it must be reasonable, and if the calculation is reproducible then it is correct. On every task, about half of the students suggested and/or attempted solving the problem from first principles, which was the most popular response to the prompt at this level.

Students' preference for computation is further shown by the existence of arithmetic versions of strategies like *covariation*, *unit analysis*, and *special case analysis*. In

the arithmetic versions of these strategies, students made decisions based on arithmetic computation instead of more symbol-based processes. For instance, in the arithmetic version of *covariation*, a few students plugged in numbers for quantities to check whether the expression produces the expected covariational behavior. The existence of arithmetic versions of these strategies may reflect the lack of qualitative tools available to students and the familiarity with arithmetical substitutions from most course assessments. This is consistent with other literature on novice understanding of problem-solving documenting that novices are less comfortable manipulating terms in multisymbol expressions [43], and thus more likely to plug in numbers sooner in a problem than experts would [4]. Implementing these computation-intensive strategies can be quite cumbersome; with this amount of computation also comes the potential to make arithmetic errors. For instance, on the conservation of momentum task, none of the students who attempted *algebraic substitution* finished the computation. Similarly, this finding extends the results of studies such as that of Torigoe and Gladding [44], in which the performance of introductory students was significantly better on numeric versions of questions than on symbolic versions.

We note that in the Burkholder *et al.* study [11], students still attempted their do full solution strategy, analogous to *solving for the given expression*, even when the prompt included "(without solving)," emphasizing the comfort level for this strategy among students. Similarly, Kuo *et al.* [26] found that only about half of their population used what they refer to as conceptual evaluation strategies, even after instruction focused on coherence between calculations and concepts. Whereas about half of our students tried to rederive the expression as a means of evaluating, 27% of their sample did this in their contrasting cases format, despite the parenthetical instruction not to do so.

Finally, it should be noted that some of these primarily computational strategies nevertheless reflect a degree of sensemaking. While *checking with computation* involves little to no physical interpretation of the given expressions, both *algebraic* and *arithmetic substitution* strategies include identifying laws of physics at play, such as the law of conservation of energy for the inclined plane task and the law of conservation of momentum for the conservation of momentum task.

3. Students used productive evaluation strategies that have not previously been noted as canonical

Our third observation was that some students used strategies that are not among the list of previously described canonical strategies, but that can be productive. We consider some of the observed noncanonical strategies—*covariation* and *grouping*—to be sophisticated and useful and propose that they might be previously unrecognized expert strategies. Specifically, a total of 11%, 13%, and 6% of first-year students used one of these noncanonical but sophisticated strategies (*covariation* and/or *grouping*) in the inclined plane task, point charge, and conservation of momentum task, respectively.

Others, particularly *quantity roll call* and *checking for general expected physics behavior*, were unsophisticated but nonetheless an effort toward expert evaluation. A total of 21%, 57%, and 46% of first-year students used one of these strategies (*quantity roll call* and/or *checking for general expected behavior*) in the inclined plane task, point charge, and conservation of momentum task, respectively.

a. *Noncanonical, sophisticated strategies: Covariation and grouping*. We noted above that no prior studies have connected *covariation* and student evaluation strategies. However, there is reason to believe that the strategy, if not the term, is used by physics experts [42,45]. We believe that this constitutes an expertlike evaluation strategy as well: when students use covariational reasoning, not only do they load meaning onto the symbols and equation, but they also leverage the quantitative relationships between physical quantities [20,46]. We believe that covariation is distinct from special case analysis and would be a productive addition to the set of canonical evaluation strategies. Interpreting an expression as a dynamic mathematical description of what happens physically, consistent with the covariational reasoning literature [28], is a sophisticated, expertlike behavior.

In our study, some students coordinated the change in value of one quantity with changes in the other, which is the type 1 mental action (MA1) of covariation. This use of MA1 is shown in the example "solve for velocity again with a value of θ that's close to the original value. If the velocities are similar, then that would prove it's a reasonable answer." Other students' responses were classified as MA2, as they considered the direction of change of one variable with the changes in the other variable: "the velocity increases with the θ increasing which makes sense."

For the sake of comparison, in a study with high-achieving students in second-semester calculus, Carlson *et al.* reported [28] that the majority of students were able to use categories MA1 through MA3, but the higher levels were used inconsistently. Their sample had difficulty with "level 5" (L5) reasoning, which not only includes reasoning using MA5 but with the other four mental actions as well, prompting them to recommend that curricula "take into consideration the complexity of acquiring L5 (instantaneous rate) reasoning" ([28], p. 374). Given that covariation is not readily taken up by students in mathematics contexts [28,30], it cannot be taken for granted that novices will apply this reasoning in physics tasks.

PER would benefit from more explicit attention to covariational reasoning. As noted in Sec. II, the "shelf-verify" task in the study by Burkholder *et al.* [11] asked specifically about the "angle dependence" in the expression given to students rather than explicitly asking to check

limiting cases. The authors themselves note this and suggest that more explicit guidance needs to be given for students to check limiting cases. We suggest in contrast that their prompt would, in fact, result in the use of *covariation* rather than *special case analysis*. They further argue that evaluating limit is difficult because it requires prediction of "how changing certain variables will change the behavior of the system," which is the definition of covariational reasoning.

Another sophisticated but not identified-as-canonical strategy was *grouping*. Grouping involves (often mentally) reformatting and/or picking apart the given expression to get a form that looks familiar and can be interpreted in the context of the task. By grouping, students can extract the story that the equation tells about the physical scenario it describes, a characteristic of expert behavior. It is worthwhile to distinguish this version of grouping from the similarly named procedural resource defined by Wittmann and Black [47]. In their study, students were grouping terms to then perform primarily algebraic operations in a separation of variables problem. The grouping identified here, while symbolically similar, is motivated by physical significance rather than algebraic convenience. Burkholder *et al.* [11] label an identifying components strategy in their statics-focused tasks; this appears to be a task-specific case of *grouping*, since students associate sine with height or the vertical component of a force and dismiss cosine for that same reason. Because we administered multiple tasks across contexts, we were able to generalize strategies and recognize a broader strategy at play.

We argue that *grouping* is sophisticated because it involves the recognition that the variables in the given expression are not just numbers but represent physical measurements [45]. Our data suggest that its prevalence as a response category changes as students advance through the physics curriculum. This strategy will be further explored in a future publication.

b. *Noncanonical, unsophisticated yet useful strategies: Quantity roll call and checking for expected behavior.* *Quantity roll call* and *checking for expected behavior* are noncanonical, unsophisticated evaluation strategies. These strategies allow students to decide whether the given expression is false, but they do not help a student verify that the expression is reasonable. However, we propose that these may be useful steps toward more expertlike evaluation and thus should be considered approaches that can be cultivated by instructors.

Quantity roll call involves checking to see whether the quantities that are expected to affect the derived physical quantity are present. In this sense, *quantity roll call* might be considered "proto-covariation" or a less sophisticated form of *covariation*, as the given quantity is a function of other physical quantities represented in the expression as variables. However, it does not rise to the level of even the first mental action in *covariation*, i.e., students did not explicitly describe how the given quantity would change with a change in another variable in the equation. Both of these strategies have analogs among Sherin's symbolic forms [48], which describe pairings of a "symbol template," the arrangement of symbols in an expression or equation, with a "conceptual schema," the relationship among the symbols when students are constructing physics equations during problem solving. *Quantity roll call* is consistent with two forms in Sherin's "dependence" cluster, *dependence* and *no dependence*, which concern the presence or absence of a symbol, respectively. Like in *quantity roll call*, the *dependence* form does not include the specific relationship of the present symbol; including that specificity would correlate with one of several other symbolic forms in the dependence cluster, based on the specific dependence, and would be categorized as a *covariation* strategy.

Burkholder *et al.* [11] identified a category named identifying functional relationships. While the title seems nominally analogous to *covariation,* the examples provided contain a simple acknowledgment of the existence of one or multiple variables rather than a description of how the isolated quantity varies as a function of that variable. Thus, we argue that identifying functional relationships is more akin to our *quantity roll call* than to *covariation*.

In this dataset, strategies coded as *checking for expected behavior* entailed broad checks like confirming the direction of vectors or signed quantities, e.g., direction of the net electric field in the point charge task. This strategy requires some knowledge of physics and the physical interpretation of the mathematical expression, e.g., connecting negative signs to a particular direction. However, like *quantity roll call*, *checking for expected behavior* is not necessarily an accurate or complete evaluation strategy: it can be used to rule out a proposed solution but cannot speak fully to whether the solution is reasonable. For instance, an expression for the electric field of a point charge could have the correct vector direction, decrease with r, and include appropriate quantities (r, q, k), but go as $1/r$ or $1/r^3$ instead of $1/r^2$: all relevant quantities are present, with appropriate effects on the outcome, but the specific dependence is incorrect. This would require a unit analysis to discover. Nevertheless, both *quantity roll call* and *checking for expected behavior* involve efforts to make meaning with mathematics and physics.

4. Strategy prevalence was context dependent

Prior discussions of evaluation (and of problem solving in general) appear to treat skills as generically applicable [4,8–11] but this is likely to be an oversimplification. A key observation from our study is that the strategies chosen by students seem to be highly dependent on the task; some strategies appeared at much greater rates in certain contexts than in others. This trend is in keeping with many prior PER results suggesting the importance of context, e.g., [49,50].

As an example, *grouping* was much more prevalent in the inclined plane and point charge tasks than in the conservation of momentum task. We hypothesize that the grouped quantities in the former tasks are more easily recognized (e.g., as acceleration in the inclined plane or electric forces and components in the point charge task), while the potentially groupable quantities in the conservation of momentum task lack obvious physical meaning (other than the total mass).

In contrast, *checking for reasonable numbers* was a somewhat common strategy in the inclined plane and conservation of momentum tasks, but uncommon in the point charge task, consistent with the findings of Sikorski *et al.* [9]. *Checking for reasonable numbers* is likely to be a more accessible strategy for classical mechanics problems than it is for electricity and magnetism, for which students have much less intuition about reasonable values of quantities. *Unit analysis* may similarly be less straightforward for students on tasks that use unfamiliar and esoteric units in later parts of the physics curriculum [2].

Based on our findings, the assumption that evaluation skills transfer readily across the physics curriculum, even within a single semester, may need to be examined. The dependence of strategies on task and context suggests the need for using a variety of physics contexts to explore more fully the breadth and prevalence of problem-solving strategies. A single question asked of a single population should not be expected to span the space of any study of broad skills or knowledge.

C. Connecting results to research on mathematical reasoning in physics

While it is useful to characterize evaluation strategies themselves, it is also important to consider them as behaviors that are just one element of students' attempts to make sense of physical systems with mathematics. In this section, we put our empirical work on evaluation strategies in the broader body of scholarship in PER on modeling and mathematical reasoning in physics.

By incorporating the interpretation of the given expression in the context of the physical system it is supposed to describe, evaluation strategies in the *comparing to the physical world* category implicitly involve mathematical reasoning; they involve interpreting the equation as a mathematical representation of the laws of the physical world and determining the physical consequences of mathematical manipulations of these laws. In this way, these evaluation strategies are consistent with the idealization of the models articulated by Redish and Kuo [46] and Uhden *et al.* [20], connecting the postcomputational physical-mathematical model (the given expression) to the real world (the physical scenario). Again however, we note that many students do not enact these phases and appear to be stuck in mathematical formulations.

While using these strategies, students load physical meaning onto symbols and equations, a characteristic that differentiates physicists from mathematicians [46]. Unlike strategies in *checking with computation*, strategies in the *comparing to the physical world* category always involve attaching or extracting physical meaning from a result. It could be argued that arithmetic versions of *comparing to the physical world* strategies are less sophisticated than their qualitative counterparts because they bypass decisions such as deciding what approximations are valid or appropriate. However, whether in arithmetic or algebraic forms, strategies in the *comparing to the physical world* category are expertlike because they involve recognizing and interpreting the given expression as a mathematical representation of the physical world. Thus, we interpret these students' attempts as applying novice tools to an expertlike strategy. By tying physical meaning to mathematical symbols and operations, this category of strategies also involves grounding mathematics in physics: integrating physics and mathematics in ways that make sense physically [20,21,46].

All the strategies in *comparing to the physical world* explicitly or implicitly involve the use of knowledge of mathematics, mathematical computation, knowledge of physics concepts, and familiarity with and intuition about the physical world. For instance, when students use *special case analysis* in the context of the inclined plane, their associations for inclined plane problems not only include Newton's second law and conservation of energy, but also expands to Newton's first law, kinematics, and motion under gravity. In this way, the students also develop expertlike traits as they see the underlying physics laws at work in a scenario instead of just surface attributes (in this case, an object on an inclined plane) [51]. Special case analysis also requires students to check that the given expression accurately predicts the real world when the incline is upright ($\theta = 90°$) and when it is flat ($\theta = 0°$). This is consistent with findings of White *et al.* that performing limiting case analysis helped students meld their knowledge of mathematics and physics [25]. Performing validity checks requires students to tie their knowledge of mathematics, mathematical computation, and mathematical formalism to the physical context of problems with which they are confronted [2].

VI. CONCLUSIONS AND IMPLICATIONS FOR INSTRUCTION

We have documented the responses of students to tasks in which they were prompted to evaluate an expression that is the result of an imagined derivation. While there have been publications in this area, previous work generally either documented that students did not evaluate without prompting or assessed instructional interventions designed to promote evaluation strategies. Two contemporaneous studies also prompted students but generally focused on a

single task with a single population, whereas we administered multiple tasks in a variety of physics contexts and courses [2,8–11].

Previous studies have shown that students do not check their work spontaneously [4]. Our study instead focused on how, and indeed whether, students evaluated the solution when explicitly prompted to do so. We can corroborate this earlier result with what is perhaps an even stronger finding: even when explicitly asked to evaluate the solution to a physics problem, many introductory students did not do so. Essentially, the evaluation strategies experts recommend are, for the most part, not what first-year students employ to perform the same task.

We documented evaluation strategies previously described, including three that we consider "canonical": *unit analysis, special case analysis*, and *using reasonable numbers.* Our results add to previous findings and show that even when prompted, introductory students generally do not use these canonical strategies. Most students' evaluation strategies were characterized by attempts to solve the problem again from scratch as well as by computational strategies that were largely insufficient for the task. Furthermore, we have documented additional strategies not previously reported, including *grouping* and *covariation,* that are not part of the documented canon but that we characterize as expertlike and suggest should be considered canonical. Strategy choice was context dependent, suggesting the need for broader studies than have previously been reported.

We divided the observed evaluation strategies into three groups: *consulting external sources*, *checking through computation*, and *comparing to the physical world*. The analysis of our results showed our characterizations of these groups to be consistent with prior research in the use of mathematics in physics. From the epistemological framing perspective, a productive frame is one where mathematics and physics were blended in a productive manner, e.g., in *physical mapping*. This suggests that instructional activities that have students learn to use these checks can develop knowledge integration between worlds. All are consistent with evaluation as a crucial skill that should be developed and encouraged in students.

Whereas many prior studies seem to have implicitly adopted a deficit model with respect to use of canonical strategies, we identified additional, noncanonical strategies, including some that are expertlike and others in which novice behavior shows evidence of attempts at sensemaking and may represent developing expertise. In the former category, strategies of *grouping* and *covariation* are sophisticated sensemaking strategies that we believe represent expertlike behavior and indeed might well be included in an expanded list of canonical strategies. In a subsequent publication we will delve further into *grouping* and connect it to studies from psychology.

Other sensemaking lite strategies, like *quantity roll call*, were incomplete; however, we believe these strategies are productive attempts at evaluating as they are in agreement with mathematical reasoning in physics and may serve as building blocks for more sophisticated reasoning. In short, while students often seemed to have an incomplete toolset, they made moves to perform meaningful evaluation upon which instructors might productively build. Consequently, when asking students to evaluate a solution, instructors should look for these unusual evaluation strategies in student work, acknowledge their productivity to the students, and connect them to more sophisticated strategies (or more sophisticated levels inside these strategies, e.g., mental actions in covariational reasoning).

Strategies in the *checking through computation* category and arithmetic versions of strategies in the *comparing to the physical world* category make up a significant portion of students' responses at the introductory level. We claim that the prevalence of *checking through computation* reflects the culture of many traditional physics classrooms. Instruction and assessment often focus on computed answers to routine exercises, leading students to believe that physics means following a set of procedures and algorithms [23,34]. When prompted to evaluate a solution, many students perform mathematical operations and follow procedures of solving physics problems because these practices reflect established norms in their courses. Competence in mathematical computation is important in physics, but a physicist also needs to be able to blend mathematical formalism and computation with physics concepts and physical meaning. As a result, one goal of physics instruction should be to help students ground their knowledge and use of mathematics and mathematical procedures in the physical world. In discussing their findings, Kuo *et al.* [26] argue that the use of "conceptual" strategies in problem solving falls along the innovation axis on an expertise plot [52], which would suggest that strategies such as covariation and special case analysis are indicators of adaptive expertise. Our findings support theirs that novices can develop efficiency alongside innovation given appropriate instruction that focuses on mathematical sensemaking and coherence between calculations and concepts. However, the uptake of any broader strategies may be more difficult due to the mismatch of students working in a more arithmetic regime and instructors being more symbolic in their approaches. Instructional approaches that delay arithmetic substitutions would, in principle, help support this uptake.

Our findings on the prevalence of *consulting with external sources* as a validation strategy suggest that one outcome of an introductory physics course is that a few students think that the one way to know if a solution is reasonable is by asking an external authority. Since

self-evaluation is a necessary aspect of self-regulated learning [53] and a possible catalyst for the development of authentic scientific reasoning [12], a teaching goal should include explicit instruction on how to evaluate one's own work [54]. Instruction should also include dissuasion from relying on the instructor and other "authorities" as the sole source for evaluation and encouragement to consider alternate methods of evaluation.

ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation under Grants No. PHY-1405726, No. PHY-1406035, No. PHY-1912087, and No. PHY-1912660, part of collaborative projects with Warren Christensen (North Dakota State University). The authors are grateful to the members of the University of Maine Physics Education Research Laboratory for their feedback and insights on the project.

[1] L. Greenfield, Engineering student problem solving, in *Cognitive Process Instruction, Research on Teaching Thinking Skills*, edited by J. Lochhead and J. Clement (Franklin Institute Press, Philadelphia 1979), pp. 233–254.

[2] M. E. Loverude, Quantitative reasoning skills in math methods, in *Presented at PER Conference* (2015), pp. 203–206, 10.1119/perc.2015.pr.046.

[3] A. Van Heuvelen, Learning to think like a physicist: A review of research-based instructional strategies, Am. J. Phys. **59**, 891 (1991).

[4] J. L. Docktor, J. Dornfeld, E. Frodermann, K. Heller, L. Hsu, K. A. Jackson, A. Mason, Q. X. Ryan, and J. Yang, Assessing student written problem solutions: A problem-solving rubric with application to introductory physics, Phys. Rev. Phys. Educ. Res. **12**, 010130 (2016).

[5] B. R. Wilcox and S. J. Pollock, Upper-division student difficulties with separation of variables, Phys. Rev. ST Phys. Educ. Res. **11**, 020131 (2015).

[6] T. Tu, C. F. Li, J. S. Xu, and G. C. Guo, Students' difficulties with the Dirac delta function in quantum mechanics, Phys. Rev. Phys. Educ. Res. **19**, 010104 (2023).

[7] T. Tu, C.-F. Li, Z.-Q. Zhou, and G.-C. Guo, Students' difficulties with partial differential equations in quantum mechanics, Phys. Rev. Phys. Educ. Res. **16**, 020163 (2020).

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

[9] T.-R. Sikorski, G. D. White, and J. Landay, Uptake of solution checks by undergraduate physics students, in *Presented at PER Conference* (2017), pp. 368–371, 10.1119/perc.2017.pr.087.

[10] M. Lenz, P. J. Emigh, and E. Gire, Surprise! Students don't do special-case analysis when unaware of it, in *Presented at PER Conference* (2018), 10.1119/perc.2018.pr.Lenz.

[11] E. Burkholder, L. Blackmon, and C. Wieman, Characterizing the mathematical problem-solving strategies of transitioning novice physics students, Phys. Rev. Phys. Educ. Res. **16**, 020134 (2020).

[12] National Research Council, *Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics Education* (National Academies Press, Washington, DC, 2011), 10.17226/13099.

[13] M. Scriven and R. Paul, A statement by Michael Scriven & Richard Paul, in *Proceedings of the 8th Annual International Conference on Critical Thinking and Education Reform* (1987), retrieved from criticalthinking.org.

[14] A. H. Schoenfeld, *Mathematical Problem Solving* (Academic Press, New York, 1985).

[15] D. S. Wright and C. D. Williams, A wise strategy for introductory physics, Phys. Teach. **24**, 211 (1986).

[16] 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).

[17] J. I. Heller and F. Reif, Prescribing effective human problem-solving processes: Problem description in physics, Cognit. Instr. **1**(2), 177 (1984).

[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] G. Polya, *How to Solve It* (Princeton University Press, Princeton, NJ, 1945).

[20] O. Uhden, R. Karam, M. Pietrocola, and G. Pospiech, Modelling mathematical reasoning in physics education, Sci. Educ. **21**, 485 (2012).

[21] B. R. Wilcox, M. D. Caballero, D. A. Rehn, and S. J. Pollock, Analytic framework for students' use of mathematics in upper-division physics, Phys. Rev. ST Phys. Educ. Res. **9**, 020119 (2013).

[22] J. Tuminaro, A cognitive framework for analyzing and describing introductory students' use and understanding of mathematics in physics, dissertation, University of Maryland, 2004.

[23] T. J. Bing and E. F. Redish, Analyzing problem solving using math in physics: Epistemological framing via warrants, Phys. Rev. ST Phys. Educ. Res. **5**, 020108 (2009).

[24] T. J. Bing and E. F. Redish, Epistemic complexity and the journeyman-expert transition, Phys. Rev. ST Phys. Educ. Res. **8**, 010105 (2012).

[25] G. White, T.-R. Sikorski, J. Landay, and M. Ahmed, Limiting case analysis in an electricity and magnetism course, Phys. Rev. Phys. Educ. Res. **19**, 010125 (2023).

[26] E. Kuo, M. M. Hull, A. Elby, and A. Gupta, Assessing mathematical sensemaking in physics through calculation-concept crossover, Phys. Rev. Phys. Educ. Res. **16**, 020109 (2020).
[27] M. Kryjevskaia, P. R. L. Heron, and A. F. Heckler, Intuitive or rational? Students and experts need to be both, Phys. Today **74**, No. 8, 28 (2021).
[28] M. Carlson, S. Jacobs, E. Coe, S. Larsen, and E. Hsu, Applying covariational reasoning while modeling dynamic events: A framework and a study, J. Res. Math. Educ. **33**, 352 (2002).
[29] M. Carlson, *A Cross-Sectional Investigation of the Development of the Function Concept* (American Mathematical Society, Providence, RI, 1998), pp. 114–162, 10.1090/cbmath/007/04.
[30] S. R. Jones, Multivariation and students' multivariational reasoning, J. Math. Behav. **67**, 100991 (2022).
[31] 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).
[32] J. L. Docktor and J. P. Mestre, Synthesis of discipline-based education research in physics, Phys. Rev. ST Phys. Educ. Res. **10**, 020119 (2014).
[33] A. F. Heckler, The role of automatic, bottom-up processes: In the ubiquitous patterns of incorrect answers to science questions, Psychol. Learn. Motiv. **55**, 227 (2011).
[34] 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).
[35] R. Knight, *Physics for Scientists and Engineers: A Strategic Approach with Modern Physics*, 4th ed. (Pearson Education Inc., 2016).
[36] G. Knoblich, S. Ohlsson, H. Haider, and D. Rhenius, Constraint relaxation and chunk decomposition in insight problem solving, J. Exp. Psychol. Learn. Mem. Cogn. **25**, 1534 (1999).
[37] M. Lenz and E. Gire, Faculty views of and expectations for dimensional analysis, in *Presented at PER Conference* (2016), pp. 196–199, 10.1119/perc.2016.pr.044.
[38] E. F. Redish, Using math in physics: 1. Dimensional analysis, Phys. Teach. **59**, 397 (2021).
[39] Y. Chen, P. W. Irving, and E. C. Sayre, Epistemic game for answer making in learning about hydrostatics, Phys. Rev. ST Phys. Educ. Res. **9**, 010108 (2013).
[40] R. R. Bajracharya and J. R. Thompson, Analytical derivation: An epistemic game for solving mathematically based physics problems, Phys. Rev. Phys. Educ. Res. **12**, 010124 (2016).
[41] See Supplemental Material at http://link.aps.org/supplemental/10.1103/rsh8-1ftk for a comprehensive list of codes and corresponding sample responses.
[42] C. Zimmerman, A. Olsho, S. W. Brahmia, A. Boudreaux, T. I. Smith, and P. Eaton, Exploring student facility with "goes like" reasoning in introductory physics, in *Presented at PER Conference* (2020), pp. 605–610, 10.1119/perc.2020.pr.Zimmerman.
[43] P. Her and M. E. Loverude, Examining student understanding of matrix algebra and eigentheory, in *Presented at PER Confernce 2020* (2020), pp. 210–215, 10.1119/perc.2020.pr.Her.
[44] E. Torigoe and G. Gladding, Same to us, different to them: Numeric computation versus symbolic representation, AIP Conf. Proc. **883**, 153 (2007).
[45] E. Redish, Using math in physics: 5. Functional dependence, Phys. Teach. **60**, 18 (2022).
[46] E. F. Redish and E. Kuo, Language of physics, language of math: Disciplinary culture and dynamic epistemology, Sci. Educ. **24**, 561 (2015).
[47] M. C. Wittmann and K. E. Black, Mathematical actions as procedural resources: An example from the separation of variables, Phys. Rev. ST Phys. Educ. Res. **11**, 020114 (2015).
[48] B. L. Sherin, How students understand physics equations, Cognit. Instr. **19**, 479 (2001).
[49] J. Mestre, Is transfer ubiquitous or rare? New paradigms for studying transfer, AIP Conf. Proc. **790**, 3 (2005).
[50] R. Dufresne, J. Mestre, T. Thaden-Koch, W. Gerace, and W. Leonard, Knowledge representation and coordination in the transfer process, in *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. Mestre (Information Age Publishing, Charlotte, NC, 2005), pp. 155–215.
[51] M. T. H. Chi, P. J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices*, Cognit. Sci. **5**, 121 (1981).
[52] D. L. Schwartz, J. D. Bransford, and D. Sears, Efficiency and innovation in transfer, in *Transfer of Learning: Research and Perspectives (Current Perspectives on Cognition, Learning, and Instruction)*, edited by J. Mestre (2005).
[53] B. J. Zimmerman and M. M. Pons, Development of a structured interview for assessing student use of self-regulated learning strategies, Am. Educ. Res. J. **23**, 614 (1986).
[54] M. D. George *et al.*, Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology, J. Coll. Sci. Teach. **26**, 163 (1997).