Summer 8-20-2021

An Investigation of Students' Use and Understanding of Evaluation Strategies

Abolaji R. Akinyemi
University of Maine, abolaji.akinyemi@maine.edu

AN INVESTIGATION OF STUDENTS' USE AND UNDERSTANDING OF EVALUATION STRATEGIES

By

Abolaji R. Akinyemi

B.S. Minnesota State University, 2015

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Physics)

The Graduate School

The University of Maine

August 2021

Advisory Committee:

John R. Thompson, Professor of Physics, Co-Advisor

Michael E. Loverude, Professor of Physics, California State University Fullerton, Co-Advisor

Asli Sezen-Barrie, Assistant Professor of Education

MacKenzie R. Stetzer, Associate Professor of Physics

Michael Wittmann, Professor of Physics

AN INVESTIGATION OF STUDENTS' USE AND UNDERSTANDING OF EVALUATION STRATEGIES

By Abolaji R. Akinyemi

Dissertation Co-Advisors: Dr. John R. Thompson and Dr. Michael E. Loverude

An Abstract of the Thesis/Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Physics)
August 2021

One expected outcome of physics instruction is that students develop quantitative reasoning skills, including evaluation of problem solutions. To investigate students' use of evaluation strategies, we developed and administered tasks prompting students to check the validity of a given expression. We collected written (N>673) and interview (N=31) data at the introductory, sophomore, and junior levels. Tasks were administered in three different physics contexts: the velocity of a block at the bottom of an incline with friction, the electric field due to three point charges of equal magnitude, and the final velocities of two masses in an elastic collision. Responses were analyzed using modified grounded theory and phenomenology.

In these three contexts, we explored different facets of students' use and understanding of evaluation strategies. First, we document and analyze the various evaluation strategies students use when prompted, comparing to canonical strategies. Second, we describe how the identified strategies relate to prior work, with particular emphasis on how a strategy we describe as grouping relates to the phenomenon of chunking as described in cognitive science. Finally, we examine how the prevalence of these strategies varies across different levels of the physics curriculum.

From our quantitative data, we found that while all the surveyed student populations drew from the same set of evaluation strategies, the percentage of students who used sophisticated evaluation strategies was higher in the sophomore and junior/senior student populations than in the first-year

population. From our case studies of two pair interviews (one pair of first years, and one pair of juniors), we found that that while evaluating an expression, both juniors and first-years performed similar actions. However, while the first-year students focused on computation and checked for arithmetic consistency with the laws of physics, juniors checked for computational correctness and probed whether the equation accurately described the physical world and obeyed the laws of physics.

Our case studies suggest that a key difference between expert and novice evaluation is that experts extract physical meaning from their result and make sense of them by comparing them to other representations of laws of physics, and real-life experience.

We conclude with remarks including implications for classroom instruction as well as suggestions for future work.

DEDICATION

To The Fam

ACKNOWLEDGEMENTS

I'd like to thank my family both related and chosen. Thank you for all your love, prayers, and encouragement. Thank you for your investment in my graduate school career and for cheering me on, even when I wanted to quit. Thank you for all the Saturday zoom meetings, and thank you for being home to me. I'd like to thank my dissertation committee for their guidance, support, encouragement, and feedback. I especially want to thank Dr. Michael Loverude for his outstanding leadership as a co-advisor from far away, even before zooming was in vogue. I'd also like to thank the past and present members of the UMaine PERL group: Dr. John Thompson, Dr. MacKenzie Stetzer, Dr. Michael Wittmann, Dr. Saima Farooq, Dr. Thanh Le, Dr. Carolina Alvarado, Dr Kevin Van Der Bogart, Dr. Benjamin Schermerhorn, Dr. Caleb Spiers, Mary Jane Brundage, William Riihiiluoma, Anthony Pina, Thomas Fittswoods, Mikayla Mays, Ryan Moyer, Meghan Brooks, Trevor Robinson, and Em Sowles. Thank you for being the best academic community anyone could ever ask for. You are great physics education researchers and more importantly, great people. I'd like to thank Charlotte Zimmerman for insightful conversations about chunking. Our hours of readings yielded fruit! I also want to say a huge thank you to my other friends within and outside the physics department; Pat Byard, Dr. Prakash Rout, Zach Smith, Komala Shivanna, Andrea Mercado, Sarah Joughin, and The Unclaimed Treasures.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1. INTRODUCTION

1.1 Origin Story

As a means of motivating this project, the author offers a brief anecdote in the first-person singular ("I") to illustrate the development of her interest in the phenomena that are studied. For most of the dissertation, the voice will shift to first-person plural ("we") to reflect the inherently collaborative nature of the work.

During my first year as a teaching assistant at UMaine, I spent a few hours a week tutoring students in the Physics Learning Center (PLC), a drop-in tutoring center operated by the physics department. Students who visited the PLC were taking introductory physics classes (both algebra- and calculus-based) and were usually working on homework or preparing for an exam. On one such occasion, I helped a student with a homework problem that require the student to determine how far away a sound could be heard. The student and I solved the problem and arrived an answer. While I continued to stare at the board, she thanked me and started to pack her things to leave. "Wait a minute" I said, "that can't be right." She looked puzzled. "No one should hear you from that far away. That distance is larger than the radius of the earth" I said. The student did not seem to share my concern, but she waited while we went through the set-up of the problem again. It turned out that I had used the natural logarithm instead of base 10 logarithm to calculate the sound level in the earlier part of the calculation. However, even at the end of the session, my student did not seem to expect the result of her calculations to make real-world sense or have real-life implications.

1.2 General Introduction

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 dimensional 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 one of the examples of what it means to "think like a physicist" [3]. In both physics and mathematics, evaluation is considered an important step in problem solving [4]–[7]. Evaluation is also an important step in the mathematical modeling process [8], [9], as well as a step in models of the use of mathematics (mathematical reasoning) in physics [10]–[12]. From the perspective of metacognition, the use of evaluation is an expression of control/regulation, self-evaluation, and beliefs about knowledge [6], [13]–[15].

The aim of this project is to probe students' use and understanding of evaluation strategies. This project explores evaluation strategies as an avenue for students to find connections between mathematical operations, physics concepts, intuition, and lived experience. This project also joins other studies found at the confluence of mathematics and physics as it probes how student ground their use of mathematics in physics in the context of using evaluation strategies. The integration of mathematics and physics in a way that makes sense physically is a mark of expertise in physics. Consequently, our project also studies how students use of evaluation strategies evolves as students gain expertise on physics.

Despite its relevance to problem solving, mathematical modeling, mathematical reasoning, and metacognition, there have only been a few studies focused on the use of evaluation strategies [2], [16]–[19]. Furthermore, prior research on evaluation in physics has largely focused on teaching and learning of the strategies of *special case analysis*, *unit analysis,* and *use of reasonable numbers*. As a result, in this project, we refer to these three strategies as canonical evaluation strategies. Most prior studies have not addressed the fundamental questions of when, and how, students choose to use evaluation strategies, and what skills they use when doing so. To add to the body of knowledge on the use of evaluation strategies in physics, we aim to answer the following questions:

1. To what extent, and in what ways, do students evaluate the validity of derived expressions or solutions when prompted?
2. To what extent are existing frameworks for problem solving and reasoning consistent with students' use of evaluation strategies?
3. How do students' use of evaluation strategies compare at different levels the physics curriculum?

To address these questions, we will be using a combination of interviews and written free responses to tasks in which students were provided students with the context and solution of a physics problem and asked how they would go about checking to see if the solution was reasonable. One of the administered tasks is shown in Figure 1.1. There were eight tasks in total. However, in this dissertation, we focus on three: the inclined plane task, the point charge task, and the bubble skating task (Figure 1.1, 1.2, and 1.3).

1.3 Project design and development

The aim of our project is to explore students' use and understanding of evaluation strategies. To this end, we designed 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.

The general layout of our research task is as follows: First, we posed a posed a hypothetical physical scenario that is simplified into a physics problem. Next, we provided an equation that was the solution to the problem posed in the scenario. Lastly, we asked students how they would go about checking if the solution was reasonable. The three tasks that we administered are shown in figure 1.1, 1.2, and 1.3, and summarized in figure 1.4.

The choice to provide students the solution to the physics problem as part of the prompt was intentional. First, providing the students with the solution of the physics problem allowed us to focus on how students evaluate expressions. Studies in PER that have shown that generally, students do not evaluate solutions spontaneously. Any study in which students are students are not explicitly prompted will give rise to the question: do students not evaluate because they don't know how to, or do they not choose to evaluate despite being able to do so when prompted? Providing students with the solution of the problem allowed us to skip these questions and hone in on how students can evaluate solutions when they are prompted. Studies where students were explicitly asked to evaluate the solution of a problem [17-19] were published after this study was designed.

Secondly, providing the students with the solution of the physics problem makes the time to compete the task shorter since students do not have to spend time solving the problem. Essentially, we designed our task so that students focused on the step of evaluating the solution, rather than solving the physics problem outlined in the task. However, one consequence of this choice is that the prevalence of students evaluating solutions in our study is perhaps not representative of the frequency of students who would evaluate their solutions to a problem in a natural problem-solving setting.

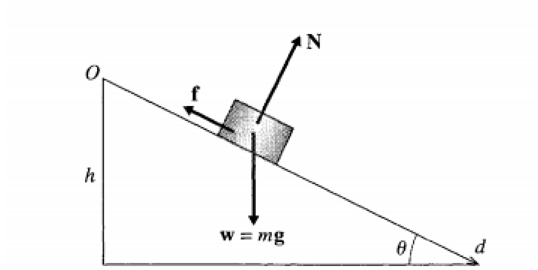
PHY 122 Physics for Engineers and Scientists Name:____________________

Fall 2018 November 13, 2018

You were asked to solve for the velocity of a block sliding down an incline just when it reaches the bottom of the slope and you obtained the following result:

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

where d is the length of the incline, θ is the angle of the incline, and μ is the coefficient of kinetic friction between the block and incline surfaces. The block starts from rest at the top of the incline (marked by O).



a. Without knowing the correct answer, how would you go about checking if your solution is reasonable?

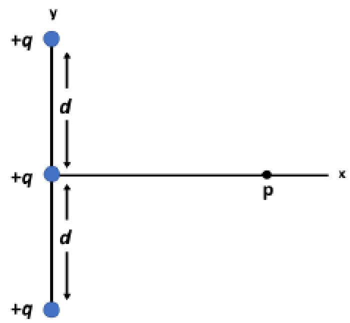
b. Using the approach(es) you described in part a, determine whether or not this is likely the correct result.

Figure 1.1: An example of the inclined plane task

The physics problems in the posed scenarios were taken from back-of-the-chapter questions in the assigned textbooks of the courses associated with each task. Specifically, the questions discussed in this dissertation were taken from the 4th edition of Randall Knight's *Physics for Engineers and Scientists*. We chose physics problems with symbolic solutions that could be easily evaluated using special case analysis, and unit analysis.

PHY 122 Physics for Engineers and Scientists Name:__________________
Fall 2018 September 13, 2018

You and your friends are working on a homework problem where you have been asked to derive an expression for the electric field (magnitude and direction) at a point p on the x-axis due to three equal positive point charges located on the y-axis at $y = 0$, and at points $y = \pm$d.



Your friend obtained the following result and claims that it is correct:

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

a. Without knowing the correct answer, how would you go about checking if your friend's solution is reasonable?

b. Using the approach(es) you described in part a, determine whether or not this is likely the correct result.

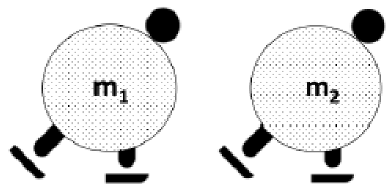
Figure 1.2: An example of the point charge task

There are two consequences of the choice to use symbolic solutions. First, most solutions to the problems given to first year students are arithmetic (i.e., the solution is a number). As a result, the symbolic solution we provided lends it itself to that first-year students would normally not use. Secondly, the symbolic nature of the tasks makes them more adaptable to expert-like strategies like unit analysis and special case analysis. Thus, the nature of our task might skew responses so that the number

PHY 121 Physics for Engineers and Scientists Name:____________________

Fall 2018 September 24, 2018

You and your friends are preparing for a game of bubble skating (skating while in a human hamster ball). Being a great physicist, you explain to your friends how the game is going to be a series of elastic collisions. You proceed to approximate the players as balls, and solve the final velocities of two balls involved in a one-dimensional elastic collision.



You obtain the following results:

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

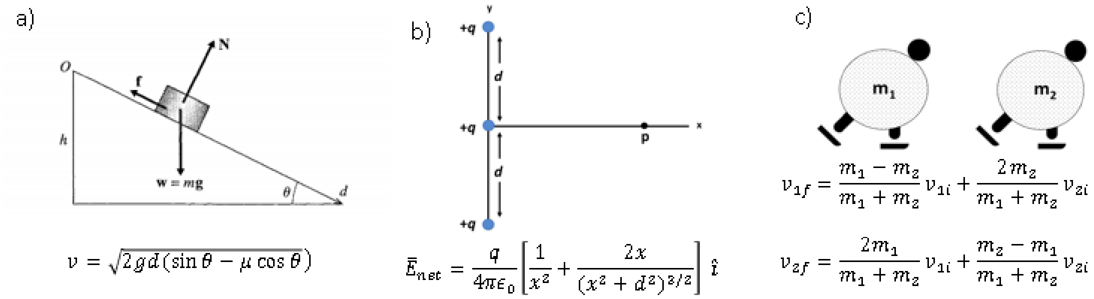
where m_1 and m_2 are the masses of the balls, v_{1i} and v_{2i} are the velocities of the balls before the collision, and v_{1f} and v_{2f} are the velocities of the balls after collision, respectively.

a. Without knowing the correct answer, how would you go about checking if your solution is reasonable?

b. Using the approach(es) you described in part a, determine whether or not this is likely the correct result.

Figure 1.3: An example of the bubble skating task

of students who use unit analysis and special case analysis is more than it would be had we provided an arithmetic solution to the given physics problems. Furthermore, symbolic solutions are the norm at the upper level of the physics curriculum where evaluating solutions is more common. Catering to the problem-solving practices of upper-division courses allow for continuity of tasks across levels.



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

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

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

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

Figure 1.4: 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

While choosing questions for the tasks, we were careful to choose questions with complex or multi-step solutions. This decision was made for two reasons. First, we wanted to dissuade students from solving for the expression again, so we chose expressions that they could not re-derive quickly. Secondly, we did not want students to evaluate based on recalling the "correct answer". Consequently, we chose complex solutions to minimize instances in which students said an answer is reasonable because they remember the equation from class, homework, or any other part of coursework.

The choice to ask students to check whether the given solution was "reasonable" was purposeful. We chose to not ask the students to check whether the given expression was "correct" in order to avoid cueing students to solve for/ rederive the provided solution. 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 the tasks so that they would be easily evaluated using the expert strategies of unit analysis, special case analysis and using reasonable numbers.

1.4 Data analysis

The analysis of our data was a multistep and iterative process. First, all written responses on the first task we administered (inclined plane) were coded openly 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. We coded responses that were both attempts and suggestions of how to evaluate the given expression. 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.

Secondly, we compared our codes and classifications to Bing and Redish's epistemic frames. This comparison helped define the highest-level categories of our codes [20][21]. Thirdly, interviews on the inclined plane task helped further flesh out codes. After the first set of (13) interviews on the point charge task, we went back and recoded the written responses. The insight from the interviews helped clarify some students' responses, decreasing the number of responses initially coded as zeros and ones. Insight from the interviews also helped us reassign a few responses to code categories to which they were better suited.

Next, we compared our results with previous work on evaluation strategies. This comparison to helped us collapse some codes into bigger categories based on the overall goal of a student's work. For instance, responses that suggested checking whether the solution was reasonable and responses where student plugged numbers into the given expression were then coded as *using reasonable numbers* in accordance with previous work on teaching students how to evaluate [16], [18].

As our project progressed, student responses from interview and written responses on the other tasks helped distinguish between similar codes and introduced new ones. Whenever a new code was created during the first coding of a newly administered task, the responses on earlier tasks were analyzed again for instances of responses that fit the description of the new code.

Finally, at every step of data analysis, code categories were decided upon through discussions between the three researchers. One of the many products of these discussions was the binning of our codes into large intermediate categories for ease of analysis and presentation.

1.5 Layout of dissertation

This dissertation was intended to be structured as three journal manuscripts framed by a common introduction and conclusion. However, due to the SARS-CoV-2 pandemic, the manuscripts that were planned to be individual and mostly independent chapters were not completed when the time came for the dissertation to be completed. As a result, this document is a hybrid between a standard-format dissertation and a multiple-manuscript format: the project literature review and background are primarily in Chapter 2, while chapters 3, 4, and 5 each have their own introduction, methodology, research questions, results, and discussion sections. The methodology described in Chapter 3 is valid for the remaining work; the only difference is the inclusion of more advanced students in later chapters, which is explicitly discussed in those chapters and in the final, concluding chapter.

In chapter 2, we explore evaluation and situate it in the fields of physics, mathematics, and cognitive science. We then delve deeper into the phenomenon of evaluation using the theoretical frameworks of epistemological frames, proofs and justification in mathematics education research, and metacognition. In chapter 3, we outline and categorize the evaluation strategies that we observed first-year students use, responding primarily to the first and second research questions. In chapter 4, we focus on an evaluation strategy called grouping, and explore the phenomenon from the perspective of mathematical reasoning, symbolic forms, and chunking; this is primarily responding to the second research question. In chapter 5, we focus on how the use of evaluation strategies differs over the physics curriculum using quantitative data and case studies at the introductory and junior/senior levels, thus responding to the third research question. In chapter 6, we summarize and reflect of the results of our project and discuss future extensions and applications of our work.

CHAPTER 2

2. LITERATURE REVIEW

In this section, we explore the significance of evaluation from the perspectives of problem solving in physics, the use of mathematics in physics, and mathematical modeling. We then delve deeper into the phenomenon of evaluation using the theoretical frameworks of epistemological frames, proofs and justification in mathematics education research, and metacognition.

2.1 Problem Solving in Physics and Mathematics

2.1.1 Evaluation as a part of critical thinking, and problem solving in physics and mathematics

Evaluation is an important aspect of critical thinking, and consequently problem solving, in mathematics and physics. One widely acknowledged goal of undergraduate science, technology, engineering and mathematics (STEM) education is the development of critical thinking [22]. Critical thinking can be defined as "the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action" [23]. Consequently, it comes as no surprise that one goal of every science department at any university is to help students develop the skill of thinking critically. In physics and mathematics, one vehicle for teaching critical thinking is thorough problem solving.

Problem solving in mathematics involves critical thinking because it entails settling a conjecture using the logical consequences of information from mathematical definitions, assumptions, and theorems. Consequently, many models of problem solving in mathematics education research explicitly include evaluation [5], [6]. For instance, according to Polya, 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 [5].

Similarly, problem solving in physics teaches critical thinking as it entails actively assessing the situation put forth in a problem statement, deciding on what details to consider or ignore in the problem based on the given context, and applying the appropriate physics concepts to arrive at a solution to the problem. Consequently, many models of problem solving in physics explicitly include evaluation [4], [7]. In physics, evaluation entails checking to make sure the solution of a problem obeys the laws of physics, is reasonable, and satisfies the constraints relevant to the context of the problem[1]. For instance, according to Wright and Williams, problem solving involves describing what is happening in the problem, isolating the unknown, substituting in the knowns, and evaluating the derived solution [4]. Even when evaluation is not listed in a problem solving rubric, it is acknowledged as an important step in problem solving [24]–[26]. However, one finding from the research in problem solving is that students do not spontaneously evaluate their results while solving physics problems [26].

In PER, a few studies have described the range of sophistication of students' problem solving skills and approaches [20], [26]–[29]. Some of these studies have been phenomenological while others have described the spectrum of students' work using frameworks including epistemic games and frames. For instance, in a phenomenological study, Walsh and colleagues observed students' approaches to problem solving ranged between a *scientific approach*, *structured plug and chug approach*, *unstructured structured plug and chug approach*, *memory based approach,* and *no clear approach* [28]. For instance, a *scientific approach* is characterized by a qualitative analysis of the problem scenario, a plan for a solution, implementation of a plan based on prior qualitative analysis, and use of physics concepts to guide the solution, and evaluation of the result. On the other hand, a structured plug and chug approach is characterised by a qualitative analysis of the problem scenario based on required formulars, plans based on a solution based on variables, a systemic implantation of the plan, a reference to physics concepts that guide the solution, and evaluation of a solution. We will discuss problem solving from the perspective of epistemic frames and games in section 2.3.

2.1.2 **Evaluation as a part of mathematical modeling**

Evaluation is also an important part of the mathematical modeling process [8], [9], [30], [31]. Mathematical modeling involves using mathematical representations to symbolize and describe the behavior of a system. Perrenet and Zwaneveld [9] studied many models of mathematical modeling and concluded that overall, models of mathematical modeling include translating between a mathematical and non-mathematical world in both directions. They also found that all models of mathematical modeling included evaluation; validating a mathematical result to make sure it fulfils the needs of the non-mathematical world being modelled. For instance, according to Blum and Leiß (figure 2.1), modeling entails interpreting a mathematical result, and validating it in the context of the real world situation being modeled [8].

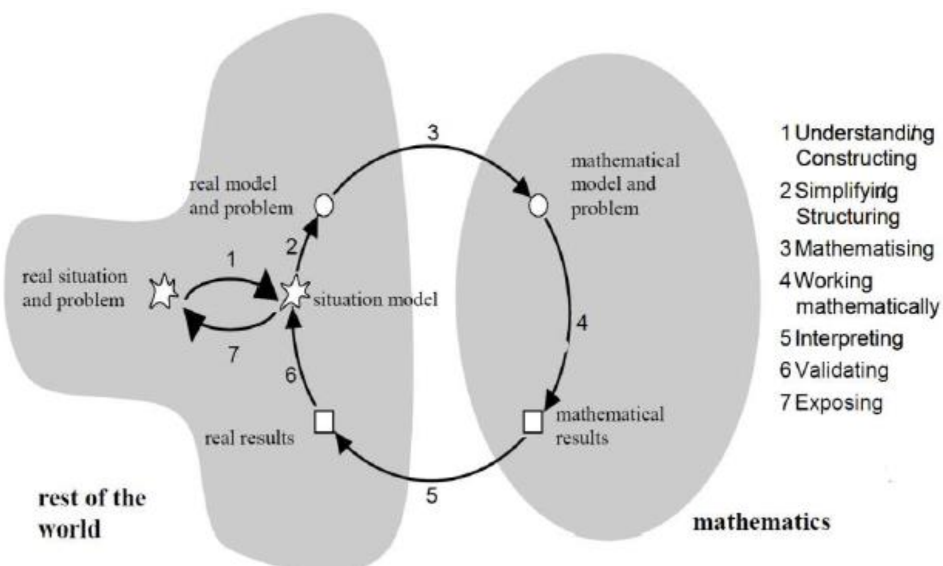


Figure 2.1: The modelling cycle according to Blum and Leiß (2007)

There are many ways to model the behavior of a system. For instance, the relationship between two quantities in a system can be mathematically described using covariational reasoning. In this paper, we use the definition of Carlson and colleagues; covariational reasoning is the cognitive activity that entails coordinating two varying quantities while considering the ways in which they change in relation to each other [32]. Carlson's covariation framework includes five developmental stages of covariation that are increasingly sophisticated and complex [33]. These stages are called mental actions and they are summarized in Table 2.1.

Table 2.1: Mental Actions of the covariation framework (Carlson, 1998)

Mental Action (MA)	**Description**	**Behaviors**
MA1	Coordinating the value of one variable with changes in the other	• Labeling the axes with verbal indications of coordinating the two variables (e.g., *y* changes with changes in *x*)
MA2	Coordinating the direction of change of one variable with changes in the other variable	• Constructing an increasing straight line • Verbalizing an awareness of the direction of change of the output while considering changes in the input
MA3	Coordinating the amount of change of one variable with changes in the other variable	• Plotting points/constructing secant lines • Verbalizing an awareness of the amount of change of the output while considering changes in the input
MA4	Coordinating the average rate-of-change of the function with uniform	• Constructing contiguous secant lines for the domain • Verbalizing an awareness of the rate of change of the output (with respect to the input) while considering uniform increments of the input
MA5	Coordinating the instantaneous rate-of-change of the function with continuous changes in the independent variable for the entire domain of the function	• Constructing a smooth curve with clear indications of concavity changes • Verbalizing an awareness of the instantaneous changes in the rate-of-change for the entire domain of the function (direction of concavities and inflection points are correct)

2.1.3 **Mathematics in mathematics vs. Mathematics in physics**

The use of mathematics (mathematical reasoning) in physics is a widely studied subject in PER [10]–[12], [34]. Competence in the use of mathematics in physics includes knowledge of both computational procedures and mathematical concepts. Problem solving in physics is unavoidably mathematics intensive so the use of mathematics (mathematical reasoning) in physics is a heavily studied subject in PER [10]–[12], [34]. Proficiency in mathematics is essential for success in physics, and a goal of almost every physics course. Competence in mathematics includes knowledge of both computational procedures and mathematical concepts. In the rest of this paper, we define computation as the act of performing "mathematical moves", procedures and operations such as algebraic manipulation, taking derivatives, taking cross products, and multiplying matrices.

Studies of the use of mathematics in physics demonstrate that mathematics and physics are interconnected in a strong, productive, and multifaceted manner. For instance, Uhden and colleagues claim that the use of mathematics in physics has three aspects: it serves as a tool (pragmatic perspective), it acts as a language (communicative function), and it provides a means for logical deductive reasoning (structural function) [12]. The authors assert that mathematics in physics goes beyond the structural function of establishing quantitative relationships between physical quantities. For example, sometimes theoretical explanations in physics are enabled by the deductive nature of mathematical formalism. Consequently, the use of mathematics in physics includes but is not limited to problem solving in physics.

Redish and Kuo assert that the use and meaning of mathematics is different for mathematicians and physicists [10]. They claim physicists load physical meaning onto symbols and equations while mathematicians do not. Another source of this difference is that physicists and mathematicians have different goals for the role of mathematics. While mathematicians tend to use mathematics to explore mathematical formalisms, physicists tend to use mathematics to model physical systems. Consequently, mathematical reasoning in physics includes the ability to describe the laws of the physical world using mathematical representations (sometimes referred to as "mathematization"[12]) and to determine the physical consequence of mathematical manipulations of these laws.

Consequently, one mark of a physicist is the ability to integrate physics and mathematics in a way that makes sense physically. For instance, when calculating the work done by a force on a charge moving along a path, a student must recognize that the equation $W_{ab} = \int_b^a \bar{F} \cdot \overline{ds}$ is symbolic representation of the relationships between the work done and the projection of the force along the direction of motion. The student must know how to successfully take an integral as well as a dot product. The student must also know that conceptually, taking a dot product is selecting the component of the force applied that is in the same direction as the displacement, while integration is "summing up" the work done over infinitesimal distances along the path.

2.1.4 **Evaluation in models of using mathematics in physics.**

In PER, there have been a few efforts to model mathematical reasoning in physics [10]–[12]. In the "zeroth order" model, from Redish and Kuo [10] (Fig. 2.2), in order to use mathematical reasoning in physics, one must first select the physical system one wants to describe, and then decide what characteristics of the system one should focus on or ignore. Next, one maps the physical structures into mathematical ones, creating a mathematical model. After modeling the system, one then must *process,* or use one's knowledge of mathematics to transform the initial description of the system. In this step, one might

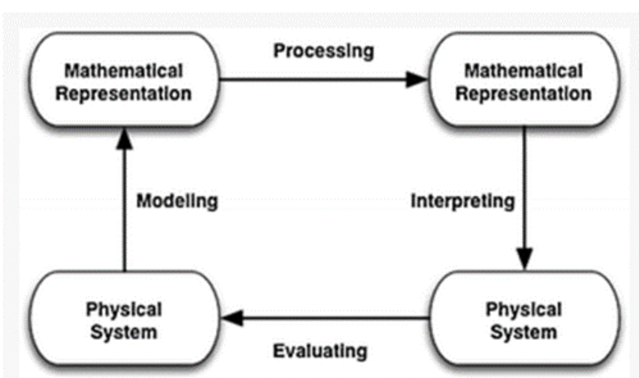


Figure 2.2: Modelling mathematical reasoning according to Redish and Kuo (2015)

solve an equation or derive an equation. Next, one interprets what the results say about the physical system, and finally one evaluates whether the results were derived correctly and whether they adequately describe the physical system at hand. If the derived behavior is found to be inadequate or incorrect, the model is then refined.

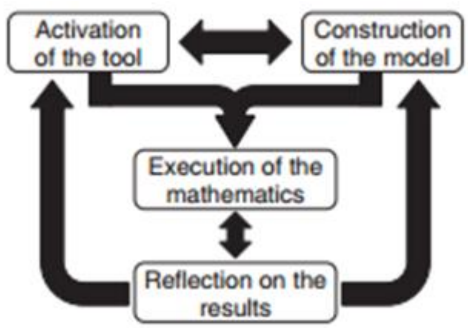


Figure 2.3: Visual representation of the ACER framework. Wilcox, Caballero, Rehn, and Pollock (2013)

Another model that describes mathematical reasoning in physics is the ACER framework (Fig. 2.3) by Wilcox and colleagues [11]. The ACER framework was specifically designed to guide and structure investigations of student difficulties with the sophisticated mathematical tools used in upper-division physics courses. The steps of the ACER model include activation of the [mathematical] tool, construction of the [mathematical] model, execution of the mathematics, and reflection on the results. Similar to the Redish and Kuo model, this model is aimed at helping physics students develop the skill of grounding mathematics in a physical system. However, unlike the Redish model, which focuses on end products and states, the ACER model focuses on the process of getting from one state to another. Also, the steps aren't so linear and cyclic – they could be done in different orders.

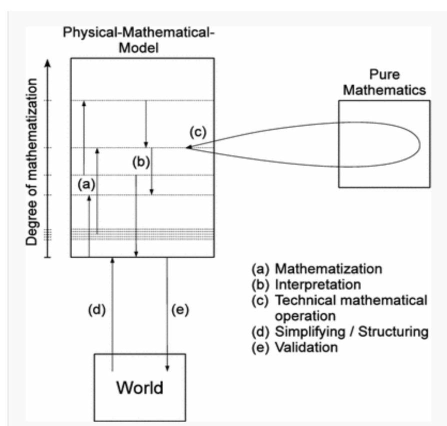


Figure 2.4: A model of the use of mathematics in physics according to Uhden, Karam, Pietrocola, and Pospiech (2012)

Lastly, another model that describes mathematical reasoning in physics is from Uhden and colleagues [12](Fig. 2.4). Like Redish's model, this representation involves moving up or down a physical-mathematical "ladder" with consecutively higher steps representing higher levels of mathematization. Notably, this model intertwines physics and mathematical reasoning in mathematization in its physical-mathematical model, and explicitly states that mathematization is done in numerous little steps rather than one big step of processing or modeling as seen in the Redish model. Unlike the Redish and ACER models, computation or other "pure mathematics" is represented as a horizontal diversion, which returns to the same level; Uhden and colleagues' model thus separate pure mathematical manipulations from the other steps involved in using math in physics.

In the first model from Redish (Fig. 2.2), mathematical reasoning in physics includes evaluating a result in the context of the physical system it is supposed to describe [10]. Similarly, the ACER framework (Fig. 2.3) entails reflecting on the results of solving a physics problem [11]. Lastly, in the Uhden et al. model (Fig. 2.4), mathematical reasoning in physics constitutes validating, checking that a physical-mathematical model is consistent with the real world [12].

All three models emphasize and illustrate the grounding of mathematics in a physical system. All three models of using mathematics also include a step of evaluation; *evaluating* in the Redish square, *reflection* on the results in the ACER framework and *validation* in the Uhden et al. model. All of the models also show that evaluation strategies involve integrating physics and mathematics in ways that make sense physically [2].

2.2 Previous research on evaluation strategies

A few PER studies have explored student use of validity checks/evaluation strategies [2], [16]–[19]. In a study that explored students' ability to reason mathematically, Loverude [2] asked physics students in an upper-level mathematical methods in physics course to evaluate whether expressions for the acceleration of masses on an Atwood machine were correct and why. He reported that when asked to evaluate an expression for the acceleration of masses on an Atwood machine, about 10-15% of the students attempted to solve the problem, while 20-50% of the students 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% of the students employed limiting case analysis. Loverude reported that this task was challenging for the students even though solution evaluation is typical of a physicist's practice.

In a study showing the impact of teaching evaluation strategies on student achievement, Warren found that learning evaluation strategies made students better self-evaluators and problem solvers [16]. Warren also asserts that checks may help establish global and local coherence of physics knowledge. In their study, Sikorski, White, and Landay [18] 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 often, a finding also replicated by Warren [16] as well as Burkholder, Blackmon, and Weiman [19], followed by checking limiting cases, and then using reasonable values. Also, they found that while performing limiting cases, most students gave a mathematical form of the limits without providing a reason why they expected the physical system to act the way it did (or claim it did). They also found that students had a hard time with reasonable values because they had not developed intuition for what values were reasonable for different contexts. Lenz, Emigh, and Gire suggest that students need explicit instruction on how to perform special case analysis if they are expected to use it to evaluate solutions [17].

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 as they force students to tie their use of mathematics to the physical context of the problems they encounter [2].

To show how evaluation strategies are used and embody mathematical reasoning, consider the masses on a frictionless pulley in an Atwood machine. While examining the validity of an expression for the acceleration of the masses, first, a student could also ensure that the units of the expression reduce to that of the quantity of acceleration (m/s^2). The student could perform a special case analysis. For instance, the student could ensure that if the two masses on the pulley are equal, the expression does not explode 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 acceleration due to gravity (g).

In summary, evaluation is an important aspect of critical thinking, and consequently problem solving, in mathematics and physics. Models of problem solving in both mathematics and physics education research include evaluation. Furthermore, evaluation is part of the mathematical modeling process. Evaluation is also an essential aspect of mathematical reasoning in physics. Prior studies on the use of evaluation strategies in PER have shown that students do not spontaneously evaluate solutions to problems but can learn how to evaluate. Studies have also shown that learning to evaluate [16] and teaching via the lens of mathematization can improve students' performance in problem solving [34]. The position of evaluation in models of problem solving in physics, and mathematical reasoning make evaluation strategies a great avenue for integrating physics and mathematics in ways that make sense physically.

2.3 Frameworks in physics, mathematics, and cognitive science

To further explore the phenomenon of evaluation, we examine evaluation through the lens of frameworks in physics education, mathematics education, and cognitive science. From physics education research, we adopt the framework of epistemological games and frames; from mathematics education research, we explore evaluation through the lens of proofs/justification; from cognitive science, we explore evaluation from the perspective of metacognition.

2.3.1 **Frameworks in PER**

2.3.1.1 **Epistemological games and frames**

In PER, studies of the use of mathematics in physics have examined the integration of mathematics and physics by implementing and building upon resources from the fields of education and psychology. Two of such borrowed tools are epistemological games and frames.

Framing as we will use it was first introduced by social psychologist Erving Goffman [35]. Frames are structures of expectations that dictate how individuals interpret situations or events [36]. A frame answers the question "what sort of activity is this?" [35]. For instance, when one visits a restaurant, the type of restaurant dictates the process to get food. At a fast-food restaurant, one would go up to the counter to order food while at a high-end restaurant, one would wait to be seated. In these scenarios, the type of restaurant is what frames the activity. Epistemological frames were imported into PER to describe problem-solving strategies in physics, initially at the introductory level, but eventually at more advanced levels as well [20], [21], [29], [37].

Bing and Redish used the lens of epistemological frames to analyze students' justification of their approaches to solve physics problems [20]. According to the authors, there are four epistemological frames students can activate when confronted with a problem: *invoking authority*, *calculation, physical mapping*, and *mathematical consistency*. In the *invoking authority* frame, students justify their problem-solving approach by citing that information that comes from an authoritative source, e.g., a physics professor or textbook, is accurate. This frame is also characterized by recalling equations, facts, and properties of physical quantities without conceptual justification, and the absence of extended chains of reasoning. In the *calculation* frame, students justify their problem-solving approach by arguing that algorithmically following a set of established computational steps lead to a dependable result. This frame is also characterized by a focus on technical correctness and mathematical formalism. In the *physical mapping* frame, students justify their problem-solving approach by arguing 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, in the *mathematical consistency* frame, students justify their problem-solving approach, by citing that certain mathematical concepts can be the underlying structure of different physical scenarios. This frame is also characterized by analogies with mathematical ideas.

Epistemic games are the set of rules and strategies that guide inquiry [38]. First put forth by Collins and Ferguson, epistemic games describe how to carry out investigations of phenomena in different disciplines. The authors called these courses of action "games" because they were not just inquiry strategies or methods, but they also involved a complex system of rules, strategies, and moves associated with particular representations. The authors identified four components of an epistemic game: epistemic form, entry conditions, moves, and exit conditions. Epistemic forms are target structures that guide a scientific query, while the entry conditions of an epistemic game determine when it is appropriate to play that game. The moves in an epistemic game are the actions that can be taken at different stages of the game. Finally, exit conditions determine when it is appropriate to stop playing the game.

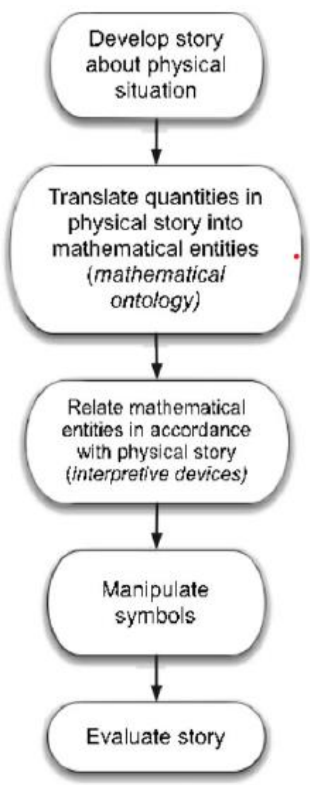


Figure 2.5: Schematic diagram of the mapping mathematics to meaning game. Tuminaro and Redish (2007)

Tuminaro used epistemic games as a cognitive framework for analyzing and describing introductory students' use and understanding of mathematics in physics [29]. He analysed videos of students working on regular back-of-the-chapter homework problems in an algebra-based introductory physics course. He identified six epistemic games with varying levels of sophistication of mathematical sensemaking with the physical context: *Mapping meaning to mathematics, mapping mathematics to meaning, physical mechanism, pictorial analysis, recursive plug and chug,* and *transliteration to mathematics*. These epistemic games both govern and limit what knowledge students think is appropriate to apply at a given time. They also shed light on the specific differences between the problem-solving abilities of novices and experts. The most relevant game here is *mapping mathematics to meaning* (Fig. 2.5), which involves solving physics problems by using equations to describe physical scenarios and processes. Tuminaro further categorized the epistemic game into three epistemic frames. The *mapping meaning to mathematics* and *mapping mathematics to meaning* games were catalogued under the *quantitative sense making* frame. The *physical mechanism* and *pictorial analysis* games were catalogued under the *qualitative sense making* frame. Finally, the *recursive plug and chug*, and *transliteration to mathematics* games were filed under the *rote equation-chasing* frame.

While Tuminaro's epistemic frames address how students solve a problem, Bing's frames address why students justify a problem solving approach. However, there are also similarities between the frames in both studies. For instance, Tuminaro's rote equation chasing frame could be likened to Bing's invoking authority and calculation frames, as all three frames depend heavily on trusting an authority (professor, textbook, etc.), recall, and calculations. Similarly, Tuminaro's qualitative sense making frame could be likened to Bing's physical mapping frames as both frames depend heavily on the physical context of the given physics problem.

Furthermore, there are similarities between the observations of Walsh and colleagues, Tuminaro and Redish, and Bing and Redish. For instance, Walsh and colleagues' unstructured plug and chug approach and Tuminaro and Redish's *rote equation chasing* frame, and Bing and Redish's *calculation* frames, are similar as all three are involve calculations without sensemaking. On the other hand, Walsh and colleagues' scientific approach and structured plug and chug approach are similar to Tuminaro and Redish's *qualitative sense making* frame and Bing and Redish's *physical mapping* frames, as all three groups involve grounding the use of mathematics in the physical context of a physics problem.

These three studies touch on expertise in physics problem solving. For instance, there is a hierarchy of approaches observed by Walsh and colleagues such that the scientific approach is more sophisticated than the structured plug and chug approach, the unstructured plug and chug approach, memory-based approach, and no clear approach, in that order. Similarly, there is a hierarchy of approaches observed by Tuminaro and Redish such that approaches in the quantitative sense making frame are more sophisticated than approaches in the rote equation chasing frame. Redish and Kuo assert that while novices tend to activate the c*alculation* and *invoking authority* epistemological frameworks, experts tend to activate the *physical mapping* and *mathematical consistency* epistemological frameworks[10].

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 [20], [21]: the more sophisticated and expert-like students become, the more flexible students are in their framing and the more likely they are to develop hybrid frames. The authors suggest that students should be instructed in such a way that their physical intuition is first activated, then the physical system is modelled with mathematics and finally the result of modeling is checked for mathematical consistency. Redish and Kuo also assert that the epistemological shift from invoking authority and calculation frames to *physical mapping* and *mathematical consistency* epistemological frames is often implied, part of the hidden curriculum for upper-level and graduate courses.

2.3.1.2 **Understanding physics equations: Symbolic forms**

Another aspect of mathematical reasoning in physics is making sense of equations. To delve into how students' make sense of equations, we focus on the epistemic complexity of equations, grouping, symbolic forms, and chunking. However, for the sake of organization, chunking will be discussed in section 2.3.3.2, as it is a framework in the cognitive sciences.

According to Bing and Redish, equations are epistemologically complex as they embed a lot of information[20]. Specifically, an equation provides a concise system for recalling encoded rules and previously derived results, encodes a calculational scheme, a physical relation among measurements, and fits within a large web of mathematical ideas. First, an expression encodes a calculational scheme. For instance, if an object starts at $x_i = 2m$ and maintains an average velocity of 6m/s for 3 seconds, $x_f = x_i + \langle v \rangle \Delta t$ tells one how to combine values to obtain the final position $x_f = 2m + \left(6\frac{m}{s}\right)(3s) = 20m$. Secondly, an expression encodes physical relation among measurements so that average ve locity tells how far an object travels per given length of time. $\langle v \rangle \Delta t$ represents how far you move in a given time interval. Adding that to where the obje ct started, x_i, must yield the final position x_f. Thirdly, an expression provides a concise system for recalling encoded rules and previously derived results. No one starts all physics problems from first principles every time. A physicist sees $x_f = x_i + \langle v \rangle \Delta t$ and thinks "that is what the final position is." An equation also fits with a large web of mathematical ideas. For instance, $x_f = x_i + \langle v \rangle \Delta t$ can be derived from the definition of average velocity by simple algebraic manipulation. It is the area under the curve of a velocity-time graph during the Δt interval. It also is recognizable as an example of the base-plus-change symbolic form [39].

Symbolic forms [39] are descriptions of students' conceptualization of a physics equations. While this work does not directly describe what we are referring to as grouping, it is one of the few bodies of work explicitly relating to physics student sensemaking with mathematical symbols. A symbolic form is made up a *symbol pattern* and a *conceptual schema*. A symbol pattern is a recognizable abstract template of an equation. For instance, the template $[\quad] = [\quad]$ is the pattern for an equation where two expressions are equal, while the template $[\quad] + [\quad] + [\quad]$ shows many terms added together. Each $[\quad]$ can be filled in with one or more terms. A conceptual schema is the internalized knowledge of mathematics that is associated with a symbol pattern. However, there can be more than one conceptual schema associated with a mathematical operation. For instance, one conceptual schema that can be associated with addition of terms, which could be represented by the template $[\quad] + [\quad] + [\quad]$, is that a whole is composed of two or more components. So, a conceptual schema is the concept that is expressed in an equation, while the symbol pattern is how this concept is represented symbolically. The symbolic form combining the template and schema above is called *parts of a whole*; examples of this would be an expression for the total energy of a system or for the components of a vector [40]. Another relevant form is *opposition,* represented by the template $[\quad] - [\quad]$ and with the schema "…influences that work against each other." Symbolic forms allow students to "(a) construct expressions, (b) reconstruct partly remembered expressions, (c) judge the reasonableness of a derived expression, and (d) extract implications from a derived expression" [39].

Equation interpretation is also aided by reasoning strategies called interpretative devices (originally called representational devices by Sherin) [36], [39], [41]. According to Sherin, there are 3 classes of interpretative devices: *narrative*, *static*, and *special case*. In the *narrative class*, equations are interpreted as telling a story that describe a changing situation. The narrative class is made up of 3 interpretative devices: changing parameters, physical change, and changing situation. While using changing parameters, an equation is interpreted by allowing some terms or variables to vary while other terms/variables are held fixed. For instance, a student might assume that some terms in an equation are held fixed while others are allowed to change. For instance, while working with the equation $a = F/m$, a student can infer that if the mass decreases, but force applied is constant, the acceleration must increase. While using physical change, an equation is interpreted by allowing some terms or variables to vary while other terms/variables are held fixed. However, in this case, the parameters that are allowed to vary are those that actually vary during the motion/scenario that the equation describes e.g., the velocity in the expression $\sum F = mg - kv$. Finally, while using changing situation, an equation is interpreted by comparing the situation the equation describes with another situation that is very different.

In the static class, equations are treated like a snapshot of the moment, describing a moment in a motion e.g., an equation, may be seen as only applicable when at the apex of the trajectory of a projectile. Interpretative devices in this task include *specific moment, generic moment, steady state, static forces, conservation,* and *accounting.* In specific moment interpretations, an equation is viewed as describing one peculiar moment in a motion, e.g., when two forces are in balance. In generic moment, an equation is viewed as describing any moment in a motion or statements that are true at any time during a motion e.g., a free-body diagram. In steady state moment, an equation is viewed as describing a system where no parameters vary with time. A static forces interpretation is a specific case of a generic moment where an equation is projected into a free body diagram rather than a motion or physical situation. In conservation, an equation is viewed as describing an application of conservation principles, where each side of the equation is associated with a different moment in the motion or physical situation. In accounting, an equation is viewed as systematically accounting for all a quantity. Here, the job of the equation is to describe how much or how many.

Finally, in the special case class, an equation is restricted to certain cases or sets of cases. Interpretative devices in this task include *restricted value, specific value, limiting case,* and *relative value.* In restricted value device, an equation is interpreted by restricting the range of some quantities in the expression. In specific value device, an equation is interpreted by restricting a quantity in the expression to a particular value. Compared to specific moment where an equation is valid for describing a particular moment in motion, in special case, the behaviour of an expression is considered in a range that is narrower than its window of validity. A limiting case is a specific version of the specific value where a quantity is assigned an extreme or limiting value. One drawback of the limiting case device is that it does not allow for detailed examination of the validity of an expression over a large range. Finally, in relative values, two quantities in an expression are compared to each other, and the value of one is restricted relative to the value of the other.

2.3.2 **Frameworks in Mathematics Education Research**

In mathematics education research, evaluation has been studied from the perspective of proofs or justifications. According to Sowder and Harel, there are three types of student's justification in mathematics: *externally based proof schemes*, *empirical proof schemes*, and *analytical proof schemes* (Fig. 2.6) [42]. *Externally based proof schemes* are further broken into *authoritarian*, *ritual,* and *symbolic proofs*. A*uthoritarian proofs* entail justifications that are based on sources like a textbook, teacher's statements, and a more knowledgeable peer. *Ritual proofs* entail justifications based on form rather than correctness, for instance, believing a proof just because it is arranged in a two-column format. S*ymbolic proofs* entail justifications based on manipulating symbols in a mathematical expression without attaching contextual meaning to them.

E*mpirical proofs* entail justifications based on empirical evidence including perceptual proofs (drawings) and examples that involve repeating patterns. Lastly, *analytical proof schemes* are either

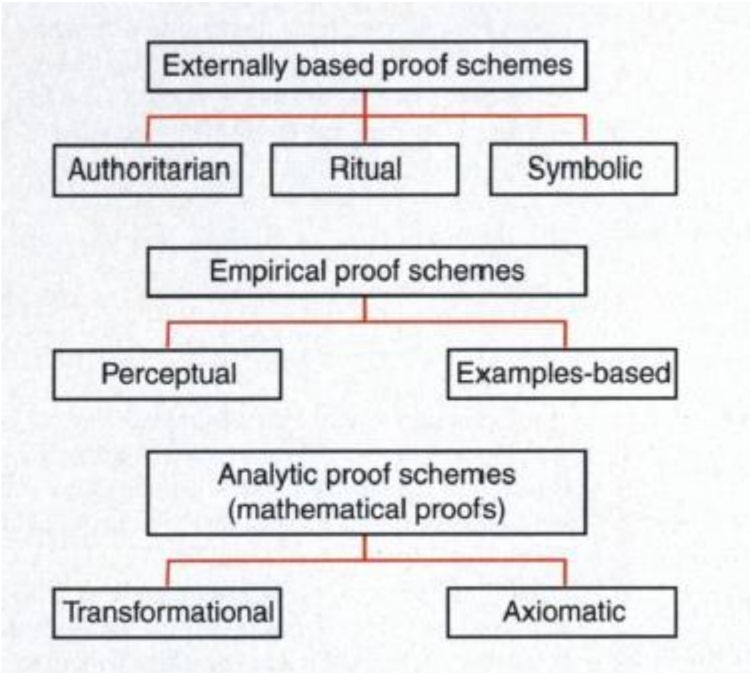


Figure 2.6: Types of students' mathematical proof justifications (Sowder and Harel 1998)

transformational or *axiomatic*. T*ransformational proofs* entail justifications that are general, perceiving the underlying structure behind patterns, include unpacking the contextual meaning of symbols in a mathematical expression and involve reasoning aimed at settling the conjecture put forth. On the other hand, *axiomatic proof schemes* entail justifications based on following logical sequences and consequences of pervious results. They also involve careful application of definitions, assumption and theorems. Sowder and Harel also claim that transformational proof scheme is a necessary precedent to the axiomatic proof schemes.

Sowder and Harel's proof schemes are similar to the epistemic frames we have discussed. For instance, Bing and Redish's epistemic frames are similar to Sowder and Harel's classification of students' justifications. Sowder and Harel's externally based proof schemes – particularly the authoritarian proof scheme – is similar to Bing and Redish's invoking authority frame. Sowder and Harel's empirical proof schemes are similar to Bing and Redish's calculation frame. While the transformational and axiomatic proof schemes are similar to Bing and Redish's mathematical consistency and physical mapping frames. These frameworks will be compared in more detail in section 3.5.

2.3.3 **Frameworks in cognitive science**

From the perspective of cognitive psychology, grouping can be examined from the frameworks of metacognition and chunking.

2.3.3.1 **Metacognition**

From the perspective of metacognition, evaluation strategies are part of *self-regulation* and implementation of evaluation strategies is an indicator of epistemological beliefs about knowledge. Self-regulation is the ability to plan, implement, and use feedback in the moment while carrying out the plan. It involves planning, monitoring, assessment, decision-making and other conscious metacognitive acts [14]. In the context of problem solving, the core of self-regulation is keeping track of one's actions while solving a problem and using the input from those observations to direct problem-solving actions. Evaluating the solution of a problem can be considered self-regulation as it involves using information about the problem to get feedback about the quality of the result in terms of its fit with the problem context, and overall plan of action for solving the problem. Evaluating a solution can be considered a pause to answer the question "does this make sense?"

Furthermore, according to Vygotsky, self-regulation is a higher-order cognitive skill and thus should be encouraged in students [13]. When students evaluate their own work, they get the opportunity to learn how to identify and correct their mistakes on their own [16]. Self-evaluation has been identified as a necessary component of self-regulated learning and has been shown to promote self-regulated learning in young students [43], [44].

Metacognition also involves beliefs and intuitions about knowledge. For instance, in the field of mathematics, Schoenfeld asserts that metacognition deals with the ideas about mathematics that students bring to work in that mathematics, and how these ideas shapes the way they work in mathematics [14]. Belief systems shape cognition – even when the beliefs are held unconsciously [6].

Consequently, students' problem-solving performance is not simply the product of what the students know but, it is also a function of their perceptions of that knowledge, derived from their experience with mathematics. During problem-solving, epistemological beliefs can determine which techniques will be used or avoided, and how long and hard one will work on a problem, and the notions one has about how mathematics knowledge is created and evaluated.

Furthermore, beliefs systems are at play when students do not perceive their mathematical knowledge as being useful to them in certain situations. Students think about mathematics problems in different "*modes*" like *discovery mode*, *proof mode*, *formal computation* mode, and *confirmation mode*. To illustrate the effect of epistemological beliefs on student performance, Schoenfeld shared the story of 45,000 students' performance on a nationwide NAEP secondary mathematics exam. The students were asked the question "*An army bus holds 36 soldiers. If 1128 soldiers are being bused to their training site. How many buses are needed?*". In response, 29% of the students responded "*31 remainder 12*", 18% responded "*31*", 23% said "*32*", and 30% did not do the computation correctly. The students that said "31 remainder 12" did the computation without considering the physical context of the problem, and treated it as requiring formal computation [14] .

Schoenfeld's ideas can be adapted to physics education such that students' problem-solving performance is also a function of their perception of physics based on their experience with physics. Many students treat school physics like it is completely divorced from real life [29] . Furthermore, the *modes* that Schoenfeld mentions are consistent with epistemological frames in PER. Students think about physics problems in different frames including *invoking authority*, *plug and chug*, and *physical mapping* [20], [29]. For instance, responding to the army bus question with the answer "*31 remainder 12*" is similar to working in *a plug and chug* frame where a student plugs numbers into a physics equation and derives a numerical answer without considering the physical context of the problem, or interpreting the meaning of the number. Also, a scenario where students accept procedures at face value and think that knowledge passed on "from above" is reminiscent of Bing and Redish's *invoking authority* frame.

2.3.3.2 **Chunking**

Another way that students make sense of equation is consistent with a phenomenon in psychology and cognitive science called chunking. Chunking was first documented in 1965 by de Groot while studying expertise in chess players [45]. However, the term chunking was coined by Simon and Chase, who repeated and modified de Groot's chess experiment in 1973 [46], [47]. Since then, chunking theory has been studied by many cognitive scientists and has undergone several refinements. A widely adopted definition of a chunk, as a collection of elements that have stronger associations with one another and weaker associations with elements within other chunks, was provided by Gobet and colleagues [48]. Chunking is studied in different fields including psychology, cognitive science, computer science, linguistics, and education, and has different meanings in different fields. Here we will limit our conversation to chunking as it pertains to memory and perception in psychology and cognitive science.

In the context of memory, chunking is the process whereby familiarity with a class of objects or events leads to the creation of a pattern of recurring networks of features or components [49]. Chunking also refers to the process by which the amount of information that can be stored in short-term memory is increased by finding patterns within a set of items to be remembered [50].

According to chunking theory, chunks are single storage units of both meaning and perception that are retrievable from long term memory in a single act of recognition [46], [47]. Chunks are accessed through a process that probes for critical features of the representation or perceived stimuli and compares those features to those of chunks in the long-term memory. This allows the perceptual stimuli (e.g., circuit diagram, chess board, equation) to be easily recognized and categorized. Chunks are also linked to other information such as useful concepts, and what actions to carry out or plans to implement given the specific patterns that are recognized in the perceptual stimuli. In this way, a chunk acts like a condition that can be satisfied by the recognition of a pattern in the perceptual stimuli [51]. Once the condition is satisfied, then the concepts, moves, and rules that are associated with the chunk or recognized pattern are evoked from long-term memory.

According to the latest revision of chunking theory called template theory, large, frequently used chunks develop into complex schematic retrieval structures called templates [52]. Templates are automatically created during pattern recognition and can be referenced within short term memory as a single chunk [53]. A template is made up of a *core* and a *slot*. A core is made up of a stable information. A template also imposes a condition that must be satisfied for the template to be used. On the other hand, a slot contains specific chunks/information that occur often but with some variation. The contents of a slot can change rapidly with new perceptual stimuli. Slots also contain information related to the domain of the stimuli including rules, facts, moves, problem solving strategies, procedure s, and processes that might have produced the perceptual stimuli.

For instance, in the template of a room, the core consists of a wall, floor and ceiling, while the slot consists of the number of doors and windows. Templates are deliberately acquired sequ ences in long term memory that are used to store identifiers for different perceptual stimuli so that related information can be retrieved from long term memory. Templates are also easily modified. The fact that templates are easily and rapidly modifiable allow for rapid recall, accounting for the superior memory skills of experts [54].

Deliberate vs. Automatic Chunking

From the perspective of memory, there are broadly two types of chunking: deliberate and automatic [48]. Deliberate chunking is conscious, explicit, intermittent, goal oriented and strategically intended to structure the material for memorization. Deliberate chunking produces chunks that are quite easy to identify as they are explicitly defined by the individual who is chunking and can be justified. On the other hand, automatic chunking is implicit, unconscious, and continuous during perception. Automatic chunking usually occurs in long-term memory and when developing expertise in a domain [48].

There are several forms of deliberate chunking. For instance, deliberate chunking can be sequential by similar terms, e.g., (aaabbbccc) is chunked as (aaa)(bbb)(ccc), or it can refer to categorizing items, e.g., (apple car plane lemon boat banana) becomes (apple lemon banana) and (car, plane, boat). Deliberate chunking can also involve recoding items, so that 11110110001 is chunked as 1969. Finally, it involves using prior knowledge to memorize material, e.g., 1969 can be memorized as the year of the moon landing.

Automatic and deliberate chunks can be used together or separately. First, both types of chunking can be used independently of each other. For instance, automatic chunking can be used alone, e.g., when learning a first language and deliberate chunking can be used alone e.g. when using a mnemonic to briefly memorize a phone number without committing it to long term memory. Both types of chunking can occur together e.g., when using mnemonics where the information is consciously chunked to access long term memory. Finally, there is memory in the absence of either type of chunking, for instance, in the mechanical rehearsal of a phone number many times without any long-term memory encoding.

Chunking and expertise

In psychology and cognitive science, chunking is positively correlated with expertise. Expertise in problem solving is associated with a large repertoire of chunks that are relevant to the problem at hand [55], [56]. While chunking theory alone does not explain expertise from the perspective of memory, it provides insight into observed expert behaviors. There are a few interesting consequences of chunking in memory as pertains to expertise. First, chunking allows experts to encode information with a smaller number of units (chunks) than novices as their units encode more information. Secondly, chunking allows experts to swiftly recognize patterns in perceived stimuli and automatically access the information linked to these patterns such as potential plans or moves. To delve further or explore these claims, we will discuss this relationship between chunking and expertise in the context of chess and interpretations of electric circuit diagrams.

Chunking is strongly associated with expertise in chess [45], [46], [52], [57], [58]. Chess is a domain that has been studied in the context of problem solving and expertise because of its complexity. In the seminal experiment on expertise, De Groot found that experts chess players were better at recalling positions of previously seen chess boards than novice chess players [45]. While a grandmaster could recall the positions of almost all the pieces, a strong amateur struggled to recreate half of the chess board. Furthermore, experts perceived the chess boards not as individual pieces but in large complexes that included information such as threat, potential moves and move sequences [54]. These large complexes were later called *chunks* by Chase and Simon, who repeated and modified De Groot's chess experiment[46], [47]. Eye-tracking experiments by de Groot and Gobet also showed that expert chess players looked at the board in groups of pieces rather than pieces individually [45], [52], and that experts have more, and larger, chunks than novices. Lastly, Chase and Simon [46], [47] found that for all participants, performance on recall tasks was better with board with chess positions that are possible in an actual game than boards with random positions. However, even on the recall task for the randomized positions, experts still outperformed novices. Experts have better recall than novices for random perceptual stimuli in their domain of expertise. This skill difference in random material is explained by chunking because since experts have more chunks than novices, they are more likely to recognize these chunks even in random positions/configurations [52], [54].

Chunking is also strongly associated with expertise in the interpretation of circuit diagrams. In an experiment similar to De Groot's chess experiment, Egan and Schwartz asked expert and novice subjects to reconstruct from memory circuit diagrams that they were previously shown [59]. There were a few notable results from this study. First, like in the chess experiment, participants in the circuits experiment recalled elements of the circuit diagrams in groups (chunks) and experts (electricians) recalled larger chunks than the unskilled subjects. Secondly, the expert participants recalled the drawings systematically while unskilled subjects did not. As predicted by chunking theory, Egan and Schwartz found that experts were able to quickly identify a concept that characterized an entire repre sentation and relates many groups of elements together. Consequently, the technicians grouped elements of the circuits into functional units and not just spatial proximity; they grouped the elements into overarching components such as a filter and an amplifier. The expert subjects' knowledge and understanding of relationships within the functional units helped them to link spatially segregated groups of symbols in recall, and enabled them to retrieve symbols systematically, because functional units are conceptually related to categories of the displayed representation in long-term memory. For instance, skilled technicians know that a power supply is likely to include a source, rectifier, filter, and regulator.

The systematic recall of the electric circuit drawings by experts in this study is also consistent with the definition of a chunk as a condition. Chunking observed in this study seemed to involve systematically retrieving elements of the representation by a generate-and-test process that examines representations to verify local details suggested by the overarching concepts relevant to th e display. Once this condition is satisfied, then the concepts, moves, and rules that are associated with the chunk or recognized pattern are evoked from long-term memory. For instance, once an expert figured out that the electric circuit was a power supply, they searched the circuit for elements needed for a power supply.

Lastly, while the average number of chunks recalled did not increase with increased study time, the average size of the circuit chunks that were recalled increased systematically. The authors explain this by the fact that if expert subjects know the conceptual category of a drawing, it seems reasonable that they would expound the details of the drawing, rather than remember entirely different sections of it, when given additional study time.

Chunk Decomposition

Other research shows that the reverse process exists as a phenomenon, labelled chunk decomposition. Chunk decomposition is the act of breaking chunks into constituent chunks. According to Knoblich, Ohlsson, Haider, and Rhenius [49], chunk decomposition is the mind's response to failure in problem solving. According to Ohlsson [60], when solving a problem, if the available chunk does not parse the problem situation in a way that is helpful towards finding a situation, decomposing the inappropriate chunks into their component features might pave the way for finding alternate routes to solving the problem.

The authors posited that the probability that a chunk will be decomposed is indirectly proportional to the "tightness" of the chunk, which is a measure of the perceptual divisibility of a chunk into other chunks, i.e., the extent to which the components of a chunk are each a meaningful perceptual pattern or chunk themselves. Tight chunks have the lowest probability of being decomposed into smaller chunks.

To test their theory, the authors gave participants a matchstick problem: a false statement written with Roman numerals, arithmetic operators, and equal signs, all constructed out of matchsticks (Fig 2.7). The goal of the problem is to move only one matchstick in such a way that the original false statement becomes true mathematically. A move consists of moving, sliding or rotating a matchstick.

In terms of chunk tightness, composite numerals such as II and IV were categorized as loose chunks because they can perceived as containing the chunks I,I, and I,V respectively. In contrast, the numerals I, and V are tight chunks because they are perceived as being a single unit. One could argue that the numeral V can be broken into the symbols \ and /, however, this division is unusual and not meaningful in this context and therefore not a chunk. Finally, the plus and equal signs have features of both loose and tight chunks. First, they decompose into potentially meaningful components. Specifically, the plus sign can be decomposed into the components "– "and "I", while the equal sign can be decomposed into two "–", all useful symbols in this context. On the other hand, both plus and equal signs are hardly ever decomposed in this way in prior experience. As a result, both operators are classified as intermediate chunks. To solve a matchstick problem, different chunks need to be decomposed. For instance, moving a sick from a symbol like VI requires that the chunk for VI is decomposed into its components V and I.

The authors found that as postulated, the probability of a chunk being decomposed decreased with the tightness of the chunk. On all versions of the research task, problems that required decomposition of tight chunks were solved less frequently and took more time to solve than problems that requires only decomposition of loose chunks.

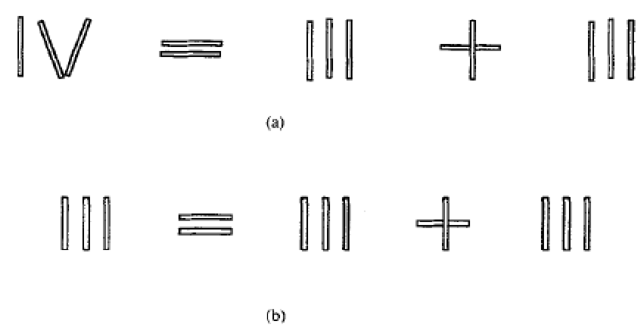


Figure 2.7: Two examples of matchstick arithmetic problems by Knoblich, Ohlsson, Haider, and Rhenius (1999)

In summary, chunking is a phenomenon in psychology and cognitive science that is consistent with grouping. Chunks are storage units of both meaning and perception that are retrievable from long term memory. They are also linked to other information such as useful concepts, plans, and rules. From the perspective of memory, chunking can be deliberate and/or automatic. Chunking is strongly associated with expertise. Chunking allows an expert to recall and encode relevant information. Chunks can also be decomposed into constituent chunks. The probability that a chunk will be decomposed is indirectly proportional to the extent to which the components of a chunk are each a meaningful perceptual pattern or chunk.

2.4 Summary of literature review

To deeply examine evaluation strategies, we adopt the frameworks of epistemological frames, proofs/justifications in mathematics, and metacognition. While there are overlaps in the arguments from these perspectives, the common denominator is the insight they provide into evaluation strategies. Like the studies in our literature review, these theoretical frameworks also advocate evaluation as a crucial skill that should be developed and encouraged in students. From the epistemological framing perspective, we want our students to be in a frame where mathematics and physics is blended in productive manner, e.g., in the *physical mapping* and *mathematical consistency* frames. Similarly, from the mathematical proof perspective, ideal evaluation strategies are in the *analytical* proof schemes because they entail logical reasoning and attaching contextual meaning to mathematical symbols and representations. Finally, from the metacognition perspective, evaluation is a great tool because it is part of self-regulation and the ideal evaluation strategies are rooted in a belief that physics is not divorced from real life. As we will show later, in the context of problem solving in physics, evaluation entails making sense of equations. Sensemaking of equations can be described by chunking theory, symbolic forms, and interpretative devices.

CHAPTER 3

3. SURVEY OF EVALUATION STRATEGIES USED BY FIRST-YEAR STUDENTS

3.1 Introduction

Despite the important role of using evaluation strategies in thinking like a physicist, metacognition, expert problem solving, and modeling, there has been little research focused on evaluation. Models of problem solving in both physics and mathematics claim that evaluation is important. However, most research in problem solving is focused on deriving the right result or 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 [11], [24], [27], [29], [61], [62]. Models of modeling and mathematization also include evaluation. Research on the use of mathematics in physics has focused on helping students attach the correct physical processes to corresponding mathematical tools but such studies do not focus on evaluation. Previous research in problem solving shows that students do not spontaneously evaluate their results while solving physics problems but can adopt the practice [6][17]. However, there is little research on what students do when prompted to evaluate solutions, and what they do if not using expert evaluation strategies.

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

1. To what extent do students use evaluation strategies when prompted?
2. To what extent are existing frameworks for problem solving and reasoning consistent with students' use of evaluation strategies?
3. How does students' use of evaluation strategies fit current models of problem solving, justifications, and reasoning from PER and adjoining fields?

The primary purpose of our research 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.

3.2 Research design and methods

To answer these questions, we designed tasks that prompted students to evaluate solutions to physics problems. The provided solutions were in form of mathematical expressions that described the physical quantity that was being sought or calculated in the problem statement. These tasks were given in both interview and written form and administered at different levels of the curriculum as well as with different problem contexts. However, for the scope of this paper, we focus on three introductory-level tasks (see Fig. 3.1). In each of these tasks, students were given a correct expression for a quantity: the velocity of a block at the bottom of an incline with friction; the electric field at a point some distance from three point charges of equal magnitude; or the final velocities of two masses in an elastic collision. The students were first prompted to describe how they would go about checking whether the expression was reasonable and then asked to use their suggested approaches to determine whether the expression was likely to be correct.

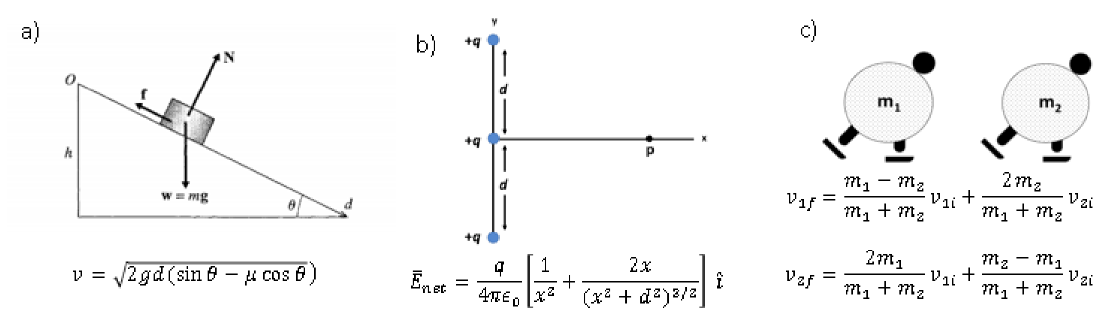


Figure 3.1: Figures and given expressions for the assigned tasks: (a) the velocity of a block at the bottom of an incline with friction; (b) the electric field at a point some distance from three point charges of equal magnitude; (c) the final velocities of two masses involved in an elastic collision

In the context of the Blum and Leiß modeling cycle [8], one could say that our research tasks ask students to pick up from the end of arrow 4 (mathematical results). In our prompt, we pose a real situation and problem and then simplify it as a problem in terms of the laws of physics. Next, we mathematize the problem so that it is in a mathematical form of variables and symbols and use this mathematical form of the problem to get a mathematical result. We expected our students to continue from the end of arrow through step 7 i.e., interpret and validate the results in the context of the real world.

The written tasks were administered in the calculus-based introductory physics sequence for engineers at a public research university in New England. The textbook used for the courses was *Physics for Scientists and Engineers: A Strategic Approach* by Knight [63]. 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 the inclined plane and point charge data were collected were taught by the same instructor. Both courses had both lecture and recitation component, but the weekly homework was almost completely quantitative. On the other hand, the course in which the conservation of momentum task data was collected had two sections co-taught by different instructors so that students received similar instruction and assessment. The courses had both lecture and recitation components and weekly homework had both quantitative and conceptual components. The written data collection depended on the 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 ($5). Some of the interviews were individual, while others were conducted with pairs of students.

While it is not possible to eliminate all potential variables, the phenomena described appeared in our data across variation in our approach, format, and level. The interviewees were first year students, and juniors. The first year interviewees were enrolled in the calculus based introductory physics sequence courses while the juniors were all physics majors.

3.3 Data analysis

Written data were analyzed using modified grounded theory/phenomenography [64] as the analysis was in part based on previous literature and there were some expectations of certain categories. For instance, interviews were conducted after the tasks had been conducted in written form, thus data acquired from interviews were analyzed with some expectation of certain categories. Also, data analysis was done with previous work like Loverude's study and Bing's epistemological frames in mind. We hoped to be able to identify recurring themes in student responses/reasoning. Our research design and data analysis have focused on emergent patterns in the data. Written data were open-coded, with phrases in a response categorized based on an overall theme. For instance, on the inclined plane task, responses in which students suggested plugging in numbers to check a velocity value were coded as "plug in numbers." To analyze interview data, we transcribed the videos and coded for approaches that were also present in the written data, then for new ones that emerged in the interview. Like the written responses, the interview codes were not based on the presence or absence of certain words or phrases but in the overall approach with which the student seemed to tackle the prompt.

On both the written and interview formats of the task, there were many different kinds of responses given, and most students suggested and/or used more than one approach. Furthermore, several (written) responses were not clear in describing what the student would do e.g., illegible handwriting or incoherent sentence. In order to account for this, we rated written responses from 0 to 3 based on clarity of explanation (3 being the clearest). After performing interviews in an attempt to clarify and shed light on the written responses, we re-analyzed the written responses for clarity; some of the response ratings were changed when deemed appropriate.

3.4 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 tasks. As we will report in a subsequent chapter, we did not observe any additional strategies in upper-division courses. At the introductory level, we collected 215, 174, and 191 responses on the inclined plane, point charge and bubble skating task respectively.

3.4.1 **Evaluation strategies observed**

We broadly classify the evaluation strategies observed in our data into three categories: *comparing to the physical world, checking through computation,* and *consulting external sources*. Strategies in the *comparing to the physical world* category involve evaluating the given expression by checking whether it is consistent with prior 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 source* category involve evaluating given expressions by checking with a trusted external source. Figure 7 shows the different strategies and the classification scheme.

For each category, we present the strategies roughly in reverse order of sophistication, from most to least sophisticated. For each strategy observed, we describe defining attributes and key features of corresponding responses. While most of the evaluation strategies observed in the data cut across the tasks at the introductory level, a few of them are specific to only one task. Thus, except when explicitly stated, the discussed evaluation strategy is present in all three task responses. However, to explain each code category, we primarily use examples from responses to the inclined plane task, for easier

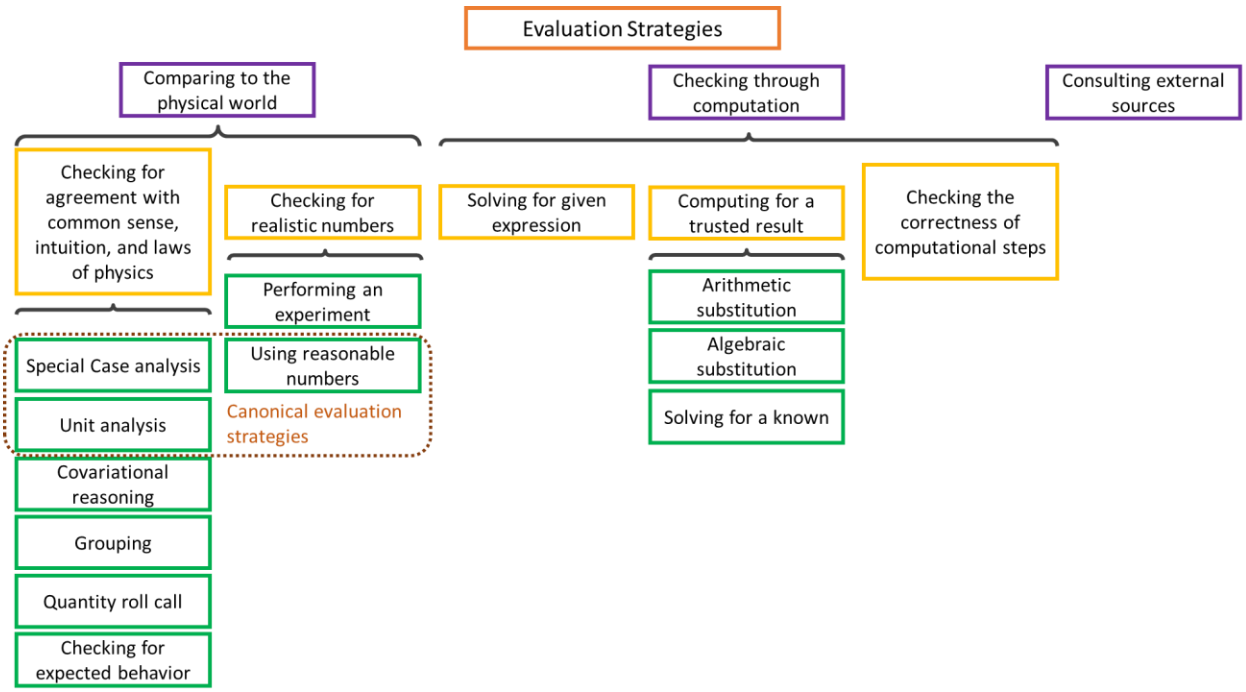


Figure 3.2: Breakdown of categories of evaluation strategy

comparisons between categories. A comprehensive list of codes and corresponding sample responses can be found in the appendix.

3.4.1.1 **Comparing to the physical world**

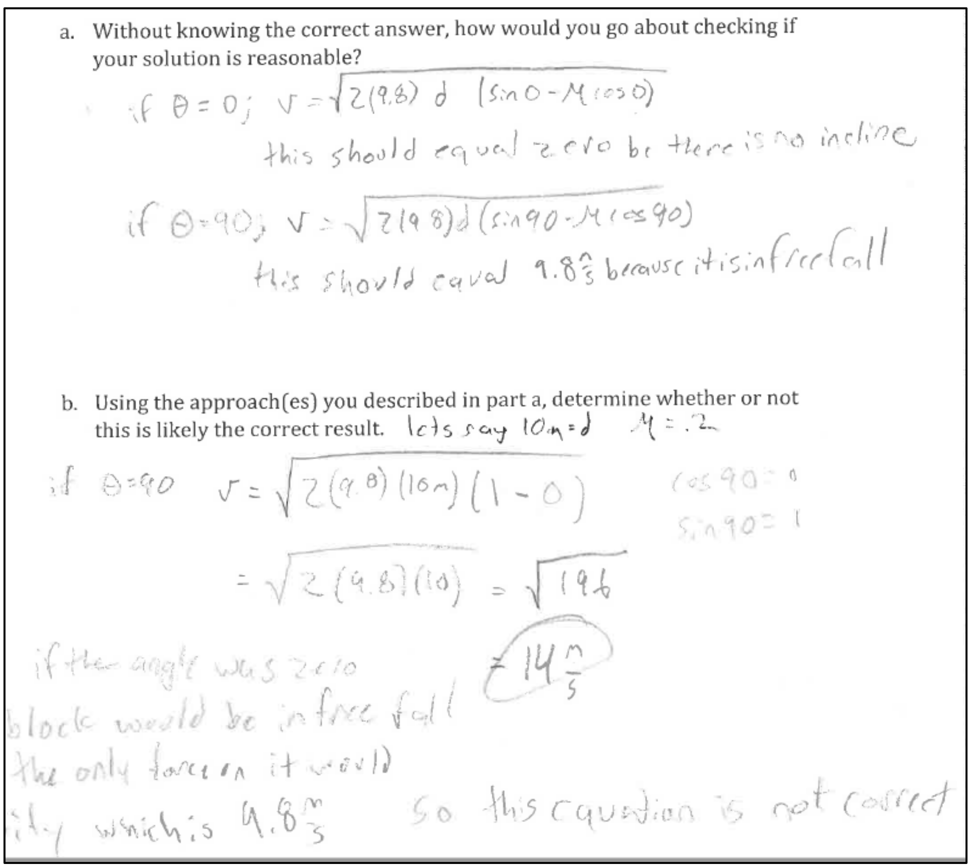
The evaluation strategies in this category involve investigating the ability of the expression to describe the physical world. All the strategies in this category also explicitly or implicitly involve the use of knowledge of mathematics, mathematical computation, knowledge of physics, and familiarity with/intuition about the physical world. During the implementation of these strategies, students may use variables/symbols or use numeric values of physical quantities. Evaluation strategies using numerical values involve attaching or extracting physical meaning from a numerical answer. This category of responses is broad. As a result, we divide this category into two sub-categories: *checking for agreement with common sense, intuition, and laws of physics,* and *checking for realistic numbers*.

a. Checking for agreement with common sense, intuition, and laws of physics

Strategies in this sub-category focus on checking for an agreement between the given expression and common sense, intuition, and laws of physics. This subcategory includes six strategies of *special case analysis*, *unit analysis*, *covariational reasoning*, *grouping*, *variable roll call,* and *checking for expected behavior.* On the inclined plane task, about 37% of students gave responses coded in this sub-category.

i. Special case analysis

This strategy involves checking whether the given expression is consistent with real world behavior and the laws of physics under certain physical or corresponding mathematical conditions, or "special cases." This group includes both students who merely suggest this evaluation strategy, and those who also say what they expect


a. Without knowing the correct answer, how would you go about checking if your solution is reasonable?

if $\theta = 0$; $v = \sqrt{2(9.8)\, d\, (\sin 0 - \mu \cos 0)}$

this should equal zero bc there is no incline

if $\theta = 90$; $v = \sqrt{2(9.8) d (\sin 90 - \mu \cos 90)}$

this should equal $9.8 \frac{m}{s}$ because it is in freefall

b. Using the approach(es) you described in part a, determine whether or not this is likely the correct result. lets say $10m = d$ $\mu = .2$

if $\theta = 90$ $v = \sqrt{2(9.8)(10m)(1-0)}$ $\cos 90 = 0$ $\sin 90 = 1$

$= \sqrt{2(9.8)(10)} = \sqrt{196}$

$\neq 14 \frac{m}{s}$

if the angle was zero block would be in free fall the only force on it would be g which is $9.8 \frac{m}{s}$ so this equation is not correct

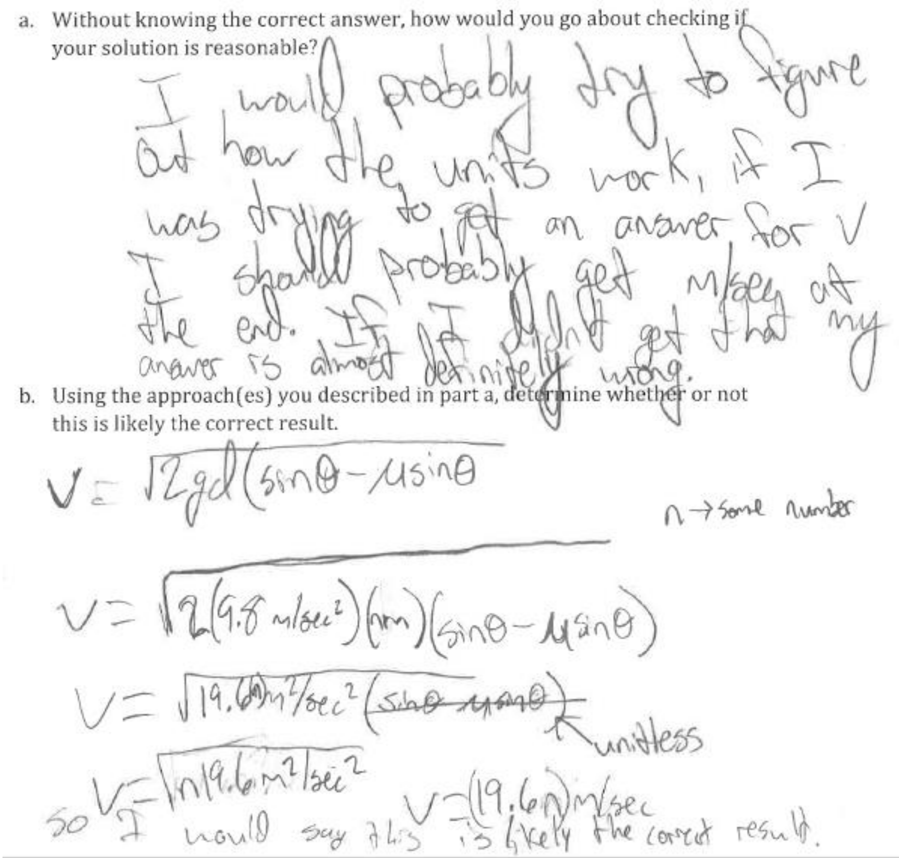
Figure 3.3: An example of a student's response coded as special case analysis

to physically happen at the chosen conditions or limits. We observed both arithmetic and algebraic versions of this strategy. In the arithmetic version of the strategy, students use numbers in place of the variables in the expression and interpret the numeric answer in the physical context of the special case being probed. An example of a response that involves the arithmetic version of the special case analysis strategy is shown in Figure 3.3. This example was coded as using *special case analysis* because the student checks whether the given expression is consistent with the law of physics under the conditions $\theta = 90°$ and $\theta = 0°$. The student stated that when $\theta = 0°$, the velocity should be zero because there is no incline, and when $\theta = 90°$, the velocity should be $9.8\frac{m}{s}$ (this is incorrect). The students then implemented the condition $\theta = 90°$, but arrived at the result $v = 14\frac{m}{s}$. This result led the student to conclude that the equation was incorrect. This example was also coded as arithmetic *special case analysis* because the student substituted numbers for variables in the expression while implementing the special case conditions.

ii. Unit analysis

This strategy involves evaluating the given expression by checking whether the expression has expected dimensions or units. An example of a response in this group is "*I would check to see if the units were reasonable, as velocity is m/s and in this case, it is* $\sqrt{\frac{m^2}{s^2}} = m/s$." This example was coded as using *unit analysis* because the student checked if the units of the expression was that of velocity $({}^{m}/_{s})$. As in the special case analysis, there were arithmetic versions of the strategy in which students substituted numbers in place of the variables in the expression while paying attention to the units of the final numerical result (See Fig. 3.4.).

The example in figure 3.4 was coded as using *unit analysis* because the student checked whether the units of the given expression came out to $\frac{m}{s}$. This response was also a coded as using arithmetic unit analysis because the student substituted numbers for variables in the expression while determining the units of the equation.


a. Without knowing the correct answer, how would you go about checking if your solution is reasonable?

I would probably try to figure out how the units work, if I was trying to get an answer for V I should probably get m/sec at the end. If I didn't get that my answer is almost definitely wrong.

b. Using the approach(es) you described in part a, determine whether or not this is likely the correct result.

$V = \sqrt{2gd(\sin\theta - \mu\sin\theta}$

$n \to$ some number

$V = \sqrt{2(9.8\ m/sec^2)(nm)(\sin\theta - \mu\sin\theta)}$

$V = \sqrt{19.6n\ m^2/sec^2\ \cancel{(\sin\theta - \mu\sin\theta)}}$ unitless

$V = \sqrt{n19.6\ m^2/sec^2}$ $V = (19.6n)\ m/sec$

so I would say this is likely the correct result.

Figure 3.4: An example of a student's response coded as using unit analysis

iii. Covariational reasoning

This strategy involves evaluating the given expression by using covariational reasoning to check whether the expression behaves as expected, i.e., by citing an expected variation between the given variable and another variable in the given expression. Examples of responses that use covariational reasoning include:

"*the velocity increases with the θ increasing which makes sense*".

and

"*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*."

The first example was coded as using *covariational reasoning* because the student evaluated the expression by citing that it has the expected covariation between the velocity and the angle of the incline θ. Likewise the second example was coded as using covariational reason because the stated that the velocities corresponding to similar values of θ should be close. Note that this student does not state the direction (increase or decrease) of expected change in either the velocity or angle of incline. This type of covariational reasoning will be further discussed in section 3.5.3. Nonetheless, the student expected a change in θ to lead to a coordinated change in the velocity. Sometimes, students employed an arithmetic version of this strategy, plugging in numbers for the variables to check if the expression produces the expected covariational behavior. An example of this is:

"*a higher velocity should result from a smaller μ and a larger θ as compared to a higher μ and lower ϑ.* $v = \sqrt{\left(9.8\frac{m}{s^2}\right)(5m)\left((sin 45) - 0.2\, cos\, 45\right)}$ *should give a higher velocity than* $v = \sqrt{\left(9.8\frac{m}{s^2}\right)(5m)\left((sin 20) - 0.7\, cos\, 20\right)}$ *which has both higher θ and lower μ decreasing the velocity*."

The above example was coded as using *covariational reasoning* because the student stated that velocity should increase with increasing θ and deceasing μ. However, it is also an example of arithmetic *covariational reasoning* because the student also substituted numerical values for variables to check for the expected covariation between the velocity, the angle of the incline θ, and coefficient of friction μ.

iv. Grouping

This strategy involves making sense of the given expression by identifying a group of symbols within an expression and describing the physical significance of the group. We describe this as *grouping*, which we define as identifying a subset of a mathematical expression that is bigger than one mathematical symbol and associating it with some physical significance. It is not uncommon for groups to be separated by mathematical operators or the equal sign. An example of a grouping response 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 example was coded as using *grouping* because 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 $(\mu\, cos\,\theta)$.

v. Quantity roll call

This strategy involves evaluating the given expression by checking whether expected quantities are present in the given expression as symbols. The quantities the students focus on are those they expect to come into play in the context described by the task or the physics described by the given mathematical expression. Examples of responses that uses the strategy of quantity roll call are:

> "*I would also look at the equation I derived and 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*."

and

> "*Does it have* μ*,* θ*,* g*, and* d *in it*?"

The first example was coded as using *quantity roll call* because the student expected the given equation to include quantities that he thinks are needed to solve for the velocity. Specifically, the student expected the expression to include mass, and cites the absence of mass in the equation as a reason why the given expression might not be reasonable. The second example was coded as using *quantity roll call* because the student listed the quantities they expect the given expression to contain. *Quantity roll call* is distinguished from covariational reasoning because it accounts for mere presence of a quantity and not for if or how the quantity affects another quantity in the given expression.

vi. Checking for general expected behavior

This strategy involves evaluating the given expression by checking that the expression generally describes or agrees with prior knowledge of physics and intuition/common sense but do not fall into any of our prior sub-categories. At the introductory level, many responses in this category are incomplete and focus on a single aspect of the expression like its sign or direction. An example of a student response that uses this strategy is "Make sure the solution is negative since the block is sliding in the –y direction." In the point charge task, responses citing the direction of the electric field are filed under this category. Responses that state that momentum or energy should be conserved are filed under this category for the bubble skating task.

b. Checking for realistic numbers

Strategies in this sub-category focus on checking numerical values of the given physics term or variable. Since the expression is supposed to describe the real world, the result of the expression should be "life-like". As seen below, not all responses provide operational definitions of terms like "feasible" or "reasonable". This subcategory includes the strategies of *using reasonable numbers* and *performing an* *experiment.* On the inclined plane task, about 25% of students gave responses coded in this sub-category.

i. Performing an experiment

This strategy involves suggesting that the given expression can be evaluated by performing an experiment to see whether it is reasonable for the physical context depicted in the task. An example of a student response that uses this strategy 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 example is coded as using performing an experiment because the response included a suggestion to perform an experiment, and for the result of the experiment to be compared to the value of velocity calculated using the given expression.

ii. Using reasonable numbers

This strategy involves 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 suggest checking the magnitude of the numerical value of the given quantity to see whether it is feasible given the context of the task. Responses in this category vary from those that just say that the given expression should be "reasonable" (interpreted as feasible) to those that explain what "reasonable" means. Examples of responses in this sub-category are:

> "*see if answer feels possible*"
>
> "*you should be able to tell if v is too slow or too fast for example, if d =10m, a v of 100,000m/s wouldn't make sense*"

and

"If I had values for d and θ, and μ I would plug these (along with the value for g) into the equation to see if the resultant velocity seemed reasonable. Otherwise, I would make up reasonable values for d and θ, and μ. $d = 5m, \theta = 45^{\text{o}}, \mu = 0.30$. $v = \sqrt{\left[2\left(9.8\frac{m}{s^2}\right)(5m)\right][sin\ 45\ - 0.3cos\ 45]} = 48.5\frac{m}{s}$. This seems like a relatively high value, so I'd say this is unreasonable."

The first example was coded as showing the *using reasonable numbers* strategy because the student suggested checking to see if the result "feels possible". The second result stated that the value of the velocity of the box should be "neither too fast nor too slow". Finally, the last response explicitly outlines reasonable values for all the quantities in the given expression.

3.4.1.2 **Checking through computation**

The evaluation strategies in this category involve evaluating the given expression using approaches that emphasize mathematical computation. In the context of this paper, we define "computation" to mean algebraic manipulation, calculation, and mathematical operation. Checking through computation is a broad category covering a spectrum of responses. As a result, we divide this category into three sub-categories: *solving for the given equation*, *computing for a trusted result,* and *checking the correctness of computational steps.*

c. Solving for given expression

Responses in this sub-category involve not evaluating the expression at all but rather directly solving the original problem from the beginning or first principles. At the introductory level, this category of response is present in all tasks and is the most frequent response. This group includes responses in which students suggest and or attempt to compare the solutions of solving for the given expression using two different methods. On the inclined plane task, about 54% of students gave responses coded in this sub-category. On the inclined plane task, 9 of the 13 interviewed students considered starting over to be a way of checking if their solutions were correct. 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 of responses coded as solving for given expression are

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

and

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

The first 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"). The second example involved solving for the velocity from first principles: first using Newton's second law to get acceleration, and then using kinematics to solve for velocity using the already calculated acceleration.

Finally, a few responses in this category involve integrating the given expression to get an "original equation". At first, it was not clear what students with such responses were trying to do. However, interview results showed that some students interpreted the word "derived" in the task prompt as "took the derivative of," i.e., differentiated. An example of such a response is:

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

The above example was coded as interpreting the word derived as "took derivative" because it includes taking an integral to get the original (given) expression for velocity.

d. Computing for a trusted result

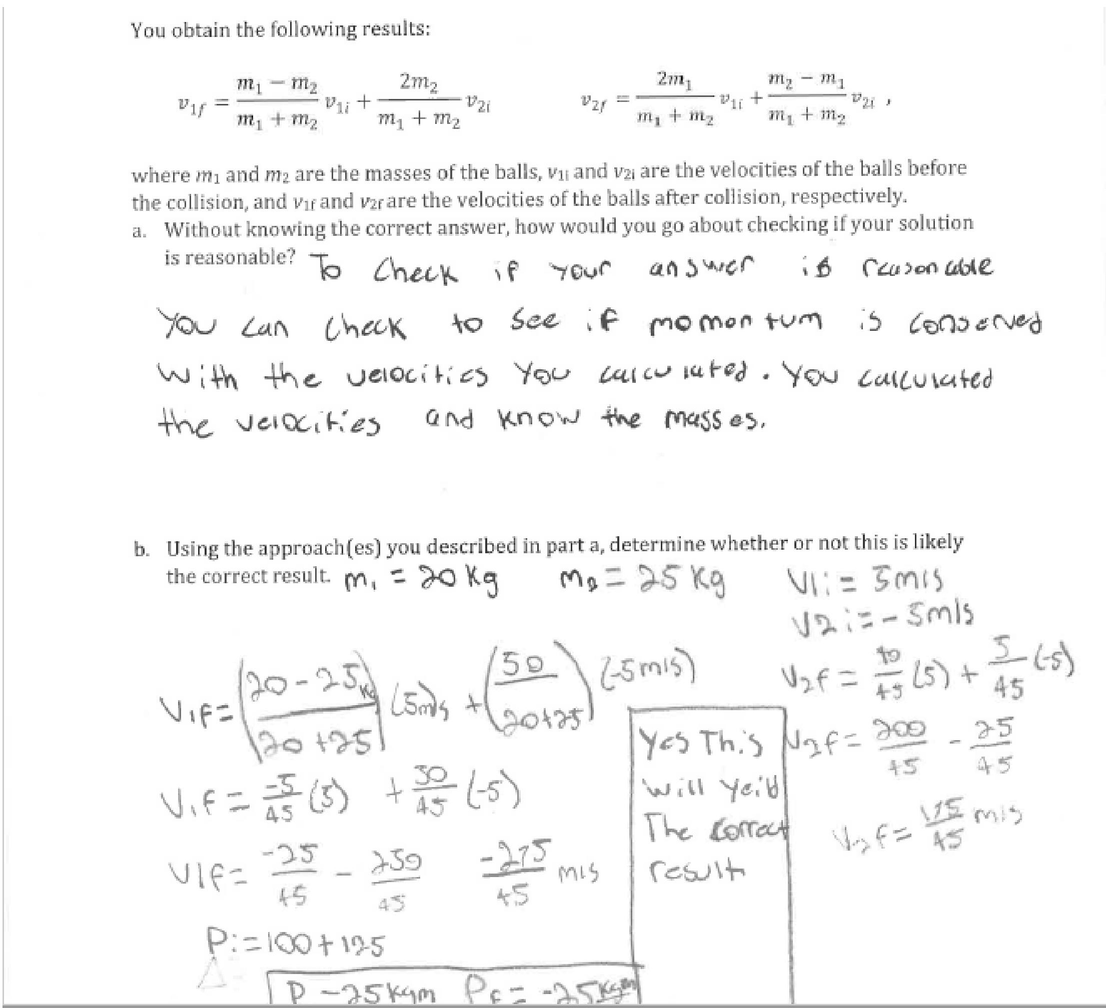
Responses in this sub-category involve using mathematical moves to produce a known arithmetic or algebraic result. All responses in this category entail a sizeable amount of mathematical computation aimed at confirming that a law of physics (e.g., conservation of energy) is obeyed or that the value of a constant or known quantity (e.g., g) is numerically correct. These are distinct from the strategies above because responses in this category generally tested for computing whether two sides of an equation were indeed equal, rather than, say, comparing a numerical result to known quantities. On the inclined plane task, about 8% of students gave responses coded in this sub-category. There are three strategies in this sub-category: *arithmetic substitution*, *algebraic substitution*, and *solving for a known*.

i. Arithmetic substitution

This strategy involves evaluating the given expression by assigning numbers to quantities in the given expression and use them to check that, numerically, a physics rule or concept holds for the given expression. Here, the trusted rule or physics concept being probed depends on the context of the task. In the inclined plane task, students checked for conservation of energy, while in the bubble skating task students checked for conservation of momentum, conservation of energy, and adherence to the laws of 1-dimensional kinematics. In context of the bubble skating task context, an example of a response that uses the strategy of *arithmetic substitution* to verify that the expression is consistent with conservation of momentum is shown in figure 3.5. This example was coded as using *arithmetic substitution* because the student chose values for the masses, and initial velocities of the skater, and plugged the numbers into the given expression to find the final velocities of the skaters. The student then uses the masses, initial and final velocities of the skaters to find the initial and momentum of the skater. The student concluded that the given expression is correct because the initial and final momentum 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 is followed. As with the arithmetic version of this strategy, students generally check against a trusted physics rule or principle relevant to



You obtain the following results:

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

where m_1 and m_2 are the masses of the balls, v_{1i} and v_{2i} are the velocities of the balls before the collision, and v_{1f} and v_{2f} are the velocities of the balls after collision, respectively.

a. Without knowing the correct answer, how would you go about checking if your solution is reasonable? To Check if your answer is reasonable You can Check to see if momentum is Conserved with the velocities You calculated. You calculated the velocities and know the masses.

b. Using the approach(es) you described in part a, determine whether or not this is likely the correct result. $m_1 = 20$ Kg $m_2 = 25$ Kg $V_{1i} = 5$ m/s $V_{2i} = -5$ m/s

$V_{1f} = \left(\frac{20-25}{20+25}\right)(5\text{m/s}) + \left(\frac{50}{20+25}\right)(-5\text{m/s})$

$V_{1f} = \frac{-5}{45}(5) + \frac{50}{45}(-5)$

$V_{1f} = \frac{-25}{45} - \frac{250}{45} \quad \frac{-275}{45}$ m/s

$V_{2f} = \frac{40}{45}(5) + \frac{5}{45}(-5)$

$V_{2f} = \frac{200}{45} - \frac{25}{45}$

$V_{2f} = \frac{175}{45}$ m/s

Yes This will Yeild The Correct result

$P_i = 100 + 125$

$\boxed{P = 25\text{kgm} \quad P_f = -25\text{kgm}}$

Figure 3.5: An example of a student's response categorised as using arithmetic substitution

the context. Usually, the equations representing these physics rules include the quantity/term that is being evaluated. For instance, in the inclined plane context, any law of physics into which the equation is substituted must contain a "v". Likewise, in the bubble skating context, the expression to be substituted into must contain a "v_f". If the given expression is consistent with a known law or equation of physics, then it is reasonable. In the bubble skating task, an example of a response that uses the strategy of *algebraic substitution* is shown in figure 3.6. This example was coded as using *algebraic substitution* because 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 or assumed-given quantity. To determine whether the original expression is reasonable, the student compares the value or expression of the calculated quantity to its given or known value. An excerpt from an interview on the inclined plane task showing the use of this strategy is:

> *"Well if I solve for v [...] you can also isolate like a variable other than v and make sure that you have that [value of chosen variable] number like a constant. So, if I were to isolate for d or isolate for g, then I'd know that answer [value of g or d]. So, if I got around the same answer then it should be the same. [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."*

This response was coded as using *solving for a known* because the student suggests using the calculated value of the velocity to solve for quantities g (acceleration due to gravity) or d (the length of the incline).

You obtain the following results:

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

where m_1 and m_2 are the masses of the balls, v_{1i} and v_{2i} are the velocities of the balls before the collision, and v_{1f} and v_{2f} are the velocities of the balls after collision, respectively.

a. Without knowing the correct answer, how would you go about checking if your solution is reasonable?

Check to see if Kinetic energy is conserved.

$$\frac{1}{2} m_1 V_{1i}^2 + \frac{1}{2} m_2 V_{2i}^2 = \frac{1}{2} m_1 V_{1f}^2 + \frac{1}{2} m_2 V_{2f}^2$$

b. Using the approach(es) you described in part a, determine whether or not this is likely the correct result.

$$m_1 V_{1i}^2 + m_2 V_{2i}^2 = m_1 V_{1f}^2 + m_2 V_{2f}^2$$

$$m_1 V_{1i}^2 + m_2 V_{2i}^2 = m_1 \left(\frac{m_1 - m_2}{m_1 + m_2} V_{1i} + \frac{2m_2}{m_1 + m_2} V_{2i} \right)^2 + m_2 \left(\frac{2m_1}{m_1 + m_2} V_{1i} + \frac{m_2 - m_1}{m_1 + m_2} V_{2i} \right)^2$$

Figure 3.6: An example of a student's response categorised as using algebraic substitution

e. Checking the correctness of computational steps

This strategy involves evaluating the given expression by checking the correctness of the computation algorithm or steps taken to compute the given mathematical expression. Examples of actions performed in the responses in this category include checking to see that the correct numbers are plugged into the calculator, making sure numbers have the correct sign and checking that mathematical operations are carried out correctly. On the inclined plane task, about 3% of students gave responses coded in this sub-category. In the point charge task, this sub-category also includes responses involving looking over a friend's work; this may reflect the fact that the task prompt is framed such that the equation to be validated is the result of a friend's work. An example of response a using the strategy of checking the correctness of computational steps is:

"*I would check to see if the values of each variable were put into the equation correctly*".

and in the point charge context, an example of a response in this category is:

"*…ask to see friend's steps throughout his attempt to solve the equation*".

The first example was coded as using *checking the correctness of computational steps* strategy because the student suggested checking to see whether the values of the quantities in the expression were substituted in correctly. The second example involves checking the steps of the work of a friend (who had solved for the given expression). In the point charge task, the given expression was presented as the result of a friend's work.

3.4.1.3 **Consulting external sources**

The third major category of strategy involves evaluating the given expression by consulting an external source (e.g., class notes, textbook, Google) or person (e.g., T.A., professor) as an authority on correctness. This category represents a small percentage of responses and is sometimes absent in data sets. However, taking into account both written and interview data, this response category is present in all three tasks. On the point charge task, about 13% of students gave responses coded in this category.

Examples of responses in this category include:

"*I would ask my TA or go to the PLC [Physics Learning Center] for help if I couldn't figure it out*"

"*…in reality, I'd probably check online*"

"*I would check my notes to see if the derived expression matched what I learned in class*."

The first example was coded as using consulting external sources because the student suggested asking her T.A. or going to the Physics Learning Center. In the second example, the student mentions that she would check online, and in the third example, the students states that she would consult her class notes. In the interview version of the inclined plane task, when students were asked how they normally check their work, 6 out of 10 students mentioned checking with the professor and comparing answers with friends. When asked how a physicist would respond to the prompt, 4 out of 7 stude nts said that a physicist would check the textbook or consult other physicists.

3.4.2 **Summary of observed categories**

When asked to evaluate the solution of physics problems in three different contexts, first-year students used a variety of strategies which we classified into 3 groups: *consulting external sources, checking through computation,* and *comparing to the physical world*.

The prevalence of the observed strategies for each task, at the introductory level, is shown in Figure 3.7. Note that multiple codes could be attributed to the same student even in one task, so the totals for a given task often add to more than 100%.

As mentioned earlier, the most prevalent strategies are *solving for the given expression,* which is in the *checking through computation* group, and *checking for agreement with common sense, intuition, and physical laws,* which is part of the more sophisticated *comparing to the physical world* group. 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 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. See the Discussion for more details. Checking through computation was a very common category of response, particularly the sub-category of *solving for given expression*. Over half the student responses in the point charge and incline questions were coded as this sub-category. A similar fraction of responses were categorized as comparing to the physical world, though some caution is appropriate as the more common responses in this category were less sophisticated responses from the checking for expected behavior group. The 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 data set, with under ten percent of responses identified with each category for most of the tasks.

Relatively few students in our sample were categorized as *consulting external sources*, and 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.

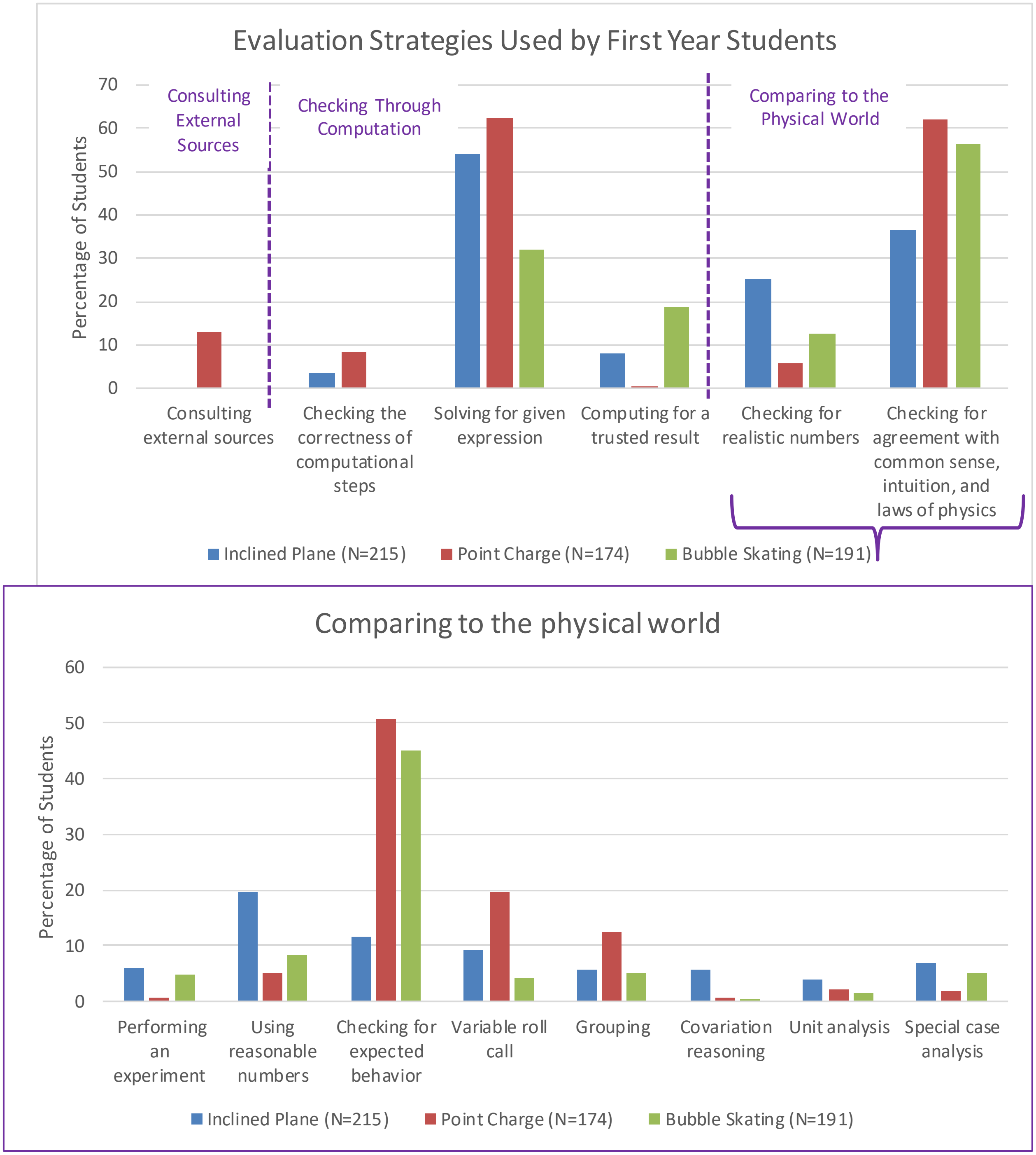


Figure 3.7: Prevalence of evaluation strategies used by first year students

3.5 Discussion

To discuss our results, first, we compare the strategies we coded in our data to strategies observed in prior study of evaluation strategies. Many of the evaluation strategies observed in our data; including *special case analysis, unit analysis*, *using reasonable numbers*, *solving for the given expression*, *covariational reasoning,* and *quantity roll call;* have been reported in prior research on evaluation [2], [16]–[19], albeit sometimes with different names. However, compared to the results of Loverude's study [2], a greater fraction of our students attempted solving for the given expression. Unlike Warren [16] or Sikorski and colleagues [15], we did not find any systematic trend in the prevalence of the use of unit analysis, limiting cases, or using reasonable numbers. The strategies of *arithmetic substitution, algebraic substitution,* and *solving for a known*, *grouping*, *consulting external sources,* and *performing an experiment* have not been reported in any prior study to our knowledge.

As described above, previous literature tends to focus on a subset of the strategies we identified: *special case analysis, unit analysis*, and *using reasonable numbers*. As these are mentioned as strategies taught to physics students [16]–[18], we classify these as canonical evaluation strategies. The other strategies we observed are thus categories as non-canonical.

To delve further into our results, first, we present some of general notes on the findings, and connect our observations to mathematical modeling, and prior research on evaluation strategies. We also examine our results from the perspective of using mathematics in physics. Next, we dissect our results using the frameworks of epistemological frames in PER, proofs/justifications in mathematics education research, and metacognition. For each of these theoretical frameworks, we discuss all 3 groups of evaluation strategies. We end with some insights from classroom practices and implications for teaching.

3.5.1 **General Observations**

There are three major general observations in 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 of the strategy, many students were unable to complete the evaluation task successfully. Third, many students used evaluation strategies that are not canonical but nonetheless useful.

3.5.1.1 **Most introductory students did not use expert-like evaluation strategies**

On every task at the introductory level, only a few students used the canonical evaluation strategies, as seen in Fig. 3.7 (b). The most prevalent was *using reasonable numbers,* used mainly on the inclined plane task, by about 20% of students. Special case analysis and suggestions to perform an experiment were seen at the 0%-10% level, and primarily on the inclined plane and bubble skating tasks.

Even when students did use canonical evaluation strategies, their implementation was not always expert-like. For instance, within *using reasonable numbers* responses, there was a range of sophistication in terms of how specific students are about what "reasonable" means. The response "see if answer feels possible" is not specific about what reasonable means. However, this response is clearer: "*you should be able to tell if v is too slow or too fast for example, if d =10m, a v of 100,000m/s wouldn't make sense."* The 3rd response provided in section *3.4.1.1.a.iii* (on page 50) is even more explicit about what "reasonable" means, because the student gave specific "reasonable" values for the mass of the block, the coefficient of friction, the incline length, and the angle of incline. Thus, while *using reasonable numbers* is a potentially productive direction, its utility is diminished when students don't know what is considered "realistic" or if they lack the physical intuition in some cases.

Previous studies showed that students do not check their work spontaneously [26]. Our study instead focused on how, and indeed whether, students evaluated the solution when explicitly prompted to do so. We observed that 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.

3.5.1.2 **Many students used evaluation strategies that emphasized computation**

Our second observation was that instead of using *canonical evaluation strategies*, many students attempted evaluating the given expression using strategies that were based on checking the correctness or accuracy of computational procedures. This is shown in the frequency of the use of evaluation strategies that are based on computation (*checking through computation* strategies), e.g., *solving for the given expression, checking the correctness of computational steps,* and *solving for a trusted result*. Essentially, these strategies reflect the notion that if the given expression is correct, it must be reasonable, and if it is reproducible then it is correct.

For instance, on every introductory-level 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. *Solving for the given expression* is based on trust in the computational steps used to obtain the given expression. In the inclined plane task for instance, the use of this strategy implies a belief that the known methods of solving for the velocity of the block (Newton's second law, kinematics, and conservation of energy) are always accurate and so, if the computational steps of these methods are correctly followed, the resulting expression should be correct.

Similar to *solving for the expression*, *checking the correctness of computational steps* entails being punctilious about following the computational steps that are relevant to solving the problem statement posed on the task. Specifically, *checking the correctness of computational steps* entails checking details such as ensuring that signs of quantities have been correctly carried through, the appropriate equations have been used, and numbers have been correctly plugged into the equation or calculator. *Solving for a known result* also involves performing a substantial amount of mathematical operation to prove the given equation is correct.

Students' preference for computation is further shown by the existence of arithmetical versions of strategies like *covariational reasoning*, *unit analysis*, and *special case analysis*. In the arithmetical versions of these strategies, students make decisions based on arithmetical computation instead of more qualitative processes. For instance, in the arithmetical version of covariational reasoning, a few students plugged in numbers for quantities to check whether the expression produces the expected covariational behavior. The preference for arithmetical versions of these strategies may reflect the lack of qualitative tools available to students and the familiarity with arithmetical substitutions in most course assessments. This is consistent with other literature on novice understanding of problem solving; novices are less comfortable manipulating terms in multi-symbol expressions, and thus more likely to plug in numbers sooner in a problem than experts would [26].

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 bubble skating task, none of the students who attempted *algebraic substitution* finished the computation at all. While the evaluation strategies in *checking with computation* involve little to no physical interpretation of the given expressions, both *algebraic* and *arithmetic substitution* strategies include recognizing the 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 bubble skating task.

3.5.1.3 **Students used evaluation strategies that are not canonical but nonetheless useful**

Our third observation was that some of the observed non-canonical strategies are sophisticated and useful (*covariational reasoning* and *grouping*), while others are unsophisticated but nonetheless an effort toward expert evaluation, e.g., *quantity roll call* and *checking for general expected physics behavior.*

a. Non-canonical, sophisticated strategies

While *covariational reasoning* is not a strategy that a physics expert would necessarily consider an evaluation strategy, we believe that it should be considered a sophisticated strategy because it involves interpreting the expression as a dynamic mathematical description of what happens physically, consistent with the covariational reasoning literature [32]. 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 examples "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 student responses were classified as a type 2 mental action (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". We believe that this strategy, while often not discussed in physics, is nevertheless expert-like; when students use covariational reasoning, not only do they load meaning onto the symbols and equation, they also leverage the quantitative relationships between physical quantities [10], [12].

For the sake of comparison, Carlson et al. reported on a study with high-achieving second semester calculus students; they stated that while the majority of students were able to use categories MA1 through MA3, the higher levels were used inconsistently by most students [32]. Their sample had difficulty with L5 reasoning, prompting them to recommend that instructors 'take into account the complexity of acquiring L5 (instantaneous rate) reasoning.'

Another sophisticated but not canonical strategy was *grouping*. Grouping involves students extracting the story that the equation tells about the physical scenario it describes. Grouping also involves 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. It is worthwhile to distinguish this version of grouping from the procedural resource defined by Wittmann and Black [65]. 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. We believe this strategy is sophisticated and 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 chapter 4, including its relationship to constructs from cognitive science, such as "chunking" and "chunk decomposition".

b. Non-canonical, unsophisticated yet useful strategies

Quantity roll call and *checking for expected behavior* are non-canonical, 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, they are useful attempts at expert evaluation. *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-covariational reasoning*" or a less sophisticated form of *covariational reasoning*, as the given quantity is a function of other physical quantities represented in the expression as variables. However, it does not go as far as *covariational reasoning* to explicitly claim that a change in one variable affects another, or mathematically describe how the given quantity would change if there is a change in another variable in the equation.

Similarly, at the introductory level, *checking for general expected physics behavior* entails 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 general expected physics 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 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 (field decreases with *r*), but the specific dependence is incorrect. This would require a unit analysis to discover. However, both *quantity roll call* and *checking for general expected physics behavior* strategies are valuable because they involve making meaning with mathematics and physics.

3.5.1.4 **Context dependence of strategy prevalence**

A key observation from the data is that the strategies chosen by students seem to be highly dependent on the task. We see some strategies appearing at much greater rates in certain tasks than other. For example, grouping is much more prevalent in the point charge and inclined plane tasks than it is in the bubble skating task. Checking for realistic numbers is a somewhat common strategy in the inclined plane and bubble skating task, but fairly uncommon in the point charge task. This phenomenon suggests a few observations.

First, prior discussions of evaluation (and of problem solving in general) appear to treat skills as generically applicable but this is likely to be an oversimplification. Checking for realistic 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 useful for tasks in later parts of the introductory physics curriculum with unfamiliar and esoteric units. The assumption that evaluation is a set of skills that transfers readily across the physics curriculum may need to be examined.

Second, there is some reason to believe that students are being more selective in their choice of task-specific strategies. Strategies that are not productive in a given task context are largely not present in the responses from higher level courses. For instance, in the bubble skating task, *grouping* and *quantity roll call* are not present beyond the first-year level. This may reflect that more experienced students recognize that *grouping* and *quantity roll call* are not productive on that task. On the flip side, because of the salience of the terms in the electric field expressions, *grouping* can be a productive strategy in the point charge task context; indeed, it is more prevalent in the more advanced sample.

Finally, this trend is in keeping with many prior PER results suggesting the importance of context. This task and context dependence also supports the need for multiple instruments using different physics content to fully explore the variety 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.

3.5.2 **Connecting results to previous research on mathematical modeling and the use of mathematics in physics**

Here we discuss how our findings fit into the models of mathematical modeling and mathematical reasoning in physics that were described in section II. Overall, we see that, when evaluating expressions, most introductory students either get stalled in the modeling process or repeat a step that was already taken rather than continuing along the model trajectory.

Using the Blum and Leiß modeling cycle [8], our task prompts students to begin at the "mathematical results" point, after "working mathematically," which refers to computation, and continue through steps 5, 6, and 7, i.e., interpret and validate the results in the context of the real world. However, our results show that this is not the typical process taken by students. Instead of moving between and connecting the mathematics world and the rest of the world, the many students who used evaluation strategies in the *checking through computation* category remained in the mathematics world. For instance, students who solved for or rederived the given expression were completing step 4; they used the mathematical model of the problem to solve for the mathematical solution. Similarly, students who solved for a known result repeated step 3 and 4, but this time mathematically reformulated the problem as solving for a different quantity or known mathematical result. Thus, these students focused on the mathematical result and did not relate it to the physical world. On the other hand, students who used strategies in the *comparing to the physical world* category did as they were prompted and performed steps 5 through 7 of the Blum and Leiß modeling cycle [8], connecting "*mathematics"* to the "*rest of the world*".

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 mathematization and mathematical reasoning because they involve interpreting the equation as a mathematical representation of the laws of physical world and determining the physical consequences of mathematical manipulations of these laws. In this way, these evaluation strategies are consistent with Redish and Kuo [10] (Fig. 2.2) and with Uhden and colleagues [12] (Fig. 2.4), connecting the post-computational physical-mathematical model (the given expression) to the real world (the physical scenario).

While using these strategies, students load physical meaning onto symbols and equations, a characteristic that differentiates physicists from mathematicians [10]. 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 expert-like 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 expert-like 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 [2],[10]–[12]. All the strategies in this category also 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.

To show how evaluation is consistent with mathematical modeling and models of using mathematics in physics, consider checking limiting cases in the context of the inclined plane task. First, the student has to assume that the equation is a mathematical description of what happens in real life; the velocity of a block is influenced by gravity and is a function of the angle of the incline and the friction between the block and ramp. Next, to check that the expression accurately predicts the real world, we consider what is expected to happen under certain conditions, e.g., when the ramp is upright and when it is completely flat. Next, we mathematize the vertical and horizontal cases as $\theta = 90^{\circ}$ and $\theta = 0^{\circ}$, respectively. Next, one evaluates the expression at these limits with technical mathematical operations including multiplication and subtraction. Then one interprets the result of the evaluation of the given expression at the two special-case angles in the context of the inclined plane problem and checks whether the results are valid and representative of the real world when the incline is upright ($\theta = 90°$) and when it is flat ($\theta = 0°$). Special case analysis uses knowledge of mathematics including algebra and trigonometry. It also requires the knowledge of physics including Newton's first law and motion under the influence of gravity. Essentially, performing validity checks forces students to tie their knowledge of mathematics, mathematical computation, and mathematical formalism to the physical context of problems with which they are confronted [2].

3.5.3 **Connecting results to frameworks in mathematics and physics education**

To delve even deeper into our results, we examine them using the frameworks of mathematical proofs and justification in mathematics education, epistemological frames in PER, and metacognition.

3.5.3.1 **Mathematical proof and justification**

From the perspective of the theoretical framework of mathematical proof and justification, strategies in the *comparing to the physical world* category are consistent with Sowder and Harel's *analytical proof schemes* [42]. This group of strategies emphasize unpacking the contextual meaning of symbols in the given mathematical expression. The category involves following logical sequences, and careful application of definitions, assumptions, and theorems, e.g., knowing the underlying physics concepts that apply in certain scenarios, mathematical representations of physical situations, and the corresponding physical consequences of mathematical operations. In the context of our prompt, these ideas translate to invoking and applying the laws of physics under certain assumptions (e.g., special cases). On the other hand, evaluation strategies in the *checking through computation* category are consistent with the *symbolic proof scheme* because they involve treating variables and numbers as though they are devoid of physical meaning. Finally, strategies in the *consulting with external sources* strategy are consistent with Sowder and Harel's *authoritarian proof scheme,* as they refer to situations where students rely on sources like a textbook, teacher's statements, and a more knowledgeable classmate to justify or validate a result.

From the perspective of epistemological frames in physics, evaluation strategies in the *comparing to the physical world* category are consistent with Bing's *physical mapping* frame [20]. 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. For instance, both *perform an experiment* and *check answer magnitude* are based on the belief that the since the expression is supposed to describe the real world, the resulting value of v should be realistic even if an operational definition of realistic is not offered.

On the other hand, strategies in the *checking with computation category* are generally consistent with Bing and Redish's *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 attention to mathematical formalism and include minimal connection between variables and the physical quantities they represent.

Lastly, the *consulting with external sources* category is consistent with Bing and Redish's *invoking authority* frame, as it cites the information from these external sources are accurate. A common feature of responses is the absence of extended chains of mathematical reasoning.

Table 3.1: Summary of comparison between results and existing frameworks

Framework → / Category ↓	Mathematical Modeling (Blum& Leiß, 2007)	Proofs/Justification (Harel and Sowder,1998)	Metacognition	Epistemic Games (Bing &Redish, 2009)
Comparing to the physical world	Performed step 5 through 7.	Analytical proof scheme	1. Developed self-regulation and self-evaluation skills. 2.Physics is not divorced from real-life.	Physical Mapping frame
Checking through computation	Repeated steps 3 and 4.	Symbolic proof scheme	1. Mathematical correctness is an evaluation tool. 2. Mathematical formalism is trustworthy.	Calculation frame
Consulting external sources	N/A	Authoritarian proof scheme	1. Undeveloped self-evaluation skills. 2.External sources are trustworthy sources of knowledge.	Invoking Authority frame

3.5.3.2 **Metacognition**

In terms of the more cognitive aspects of metacognition [14], students who use evaluation strategies in the *comparing to the physical world* category exhibit self-regulation, as they are able to step back and use input from other observations, such as intuition and formal mathematics and physics knowledge, to guide inquiries about the validity of a given expression. However, students who use strategies in the *checking with computation* category use input from the correctness or mathematical accuracy of their calculations as a guide to inquire about the validity of the given expression. Even though this input is not obtained from the physical world, at least, the power to check is self-centered rather than externally validated. Lastly, students who use the strategy of *consulting with external* *source*s have not developed the skill of self-evaluation and so depend on external sources to make decisions regarding the validity of the expression.

In terms of epistemological beliefs about physics [6], [14] , the use of evaluation strategies in the *comparing to the physical world* category reflects a belief that physics is not divorced from real life, and that the mathematical expression is a dynamic description of the laws of the physical world. The students who use these strategies seem not to hold a belief that instructors are the only authority on knowledge. They also perceive physics and physics knowledge to be useful in a real-world context. On the other hand, the use of strategies in the *checking through computation* category seem to reflect a belief in mathematical formalism and prescribed regimen for solving physics problems with little to no emphasis or importance ascribed to interpreting the mathematical operations performed or the physics principles invoked while solving the problem. Finally, the use of *consulting with external sources* reflects a belief that textbooks and physics professors are reliable sources of physics knowledge and can be invoked as authorities to check the validity of a mathematical expression that describes a physics scenario. This is further shown by instances when students said a physicist would check the textbook or consult other physicists to evaluate the solution of a problem.

3.5.4 Implications about and for instruction

Even though evaluation is one of many components of the problem-solving process, evaluation strategies incorporate several expert-like skills: meld knowledge of mathematics and physics, develop critical thinking, develop self-evaluation skills, and improve general understanding of physics.

The aim of instruction in physics is not just to increase students' knowledge of mathematics and physics, but to also teach students how to decide what types of knowledge counts as valid proof of a new result. The use of evaluation strategies forces students to tie their knowledge of mathematics, mathematical computation, and mathematical formalism to the physical contexts of problems with which they are confronted. This suggests that instructional activities that have students learn to use these checks can develop that knowledge integration between worlds. Prior research shows that teaching from the perspective of mathematization, i.e., explicitly emphasizing the coherence between physical meaning and mathematical formalism, can improve student achievement [34].

In addition to integrating mathematics and physics knowledge, using evaluation strategies gives students an opportunity to see that physics is self-consistent [16]. For instance, when students use *special 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 expert-like 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 incline plane) [66].

Some of the evaluation strategies we observed in our data and filed under the *comparing to the physical world* category are not what an expert would suggest, e.g., *covariational reasoning, grouping, and quantity roll call*. However, these strategies are productive attempts at evaluating as they are in agreement with mathematical modeling and models of using mathematics in physics. Consequently, as instructors, when we ask students to evaluate a solution, we should look for these strategies in student work and give credit to and applaud students who use these unusual evaluation strategies.

Strategies in the *checking through computation* category and arithmetic versions of strategies in the *comparing to the physical world* category make up a significant chunk of students' responses at the introductory level. This result is not surprising as it reflects the culture of the traditional physics classroom. We teach our students – perhaps unconsciously – that physics is following a set of procedures and algorithms and so that is all they know. Consequently, when prompted to evaluate a solution, students perform mathematical operations and follow procedures of solving physics problems because that is all they know and have been taught to do. Competence in mathematical computation is not wrong, it is a skill that is necessary to success in physics. However, a physicist also needs to be able to mathematize by blending 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.

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 [43] and a possible catalyst for the development of authentic scientific reasoning [22], a teaching goal should include explicit instruction on how to evaluate one's own work [67]. Instruction should also include dissuasion from relying on the instructor and other sources as the sole source of knowledge or authority on evaluation and encouragement to consider alternate methods of evaluation.

3.6 Conclusion

In summary, to investigate students' use and understanding of evaluation strategies, we asked first-year students to evaluate the solution to a physics problem in three contexts. We found that students used a slew of evaluation strategies to evaluate the given expressions, including a few strategies such as *grouping* and *performing an experiment* that have not been documented before. We divided the observed evaluation strategies into three groups: *consulting external sources*, *checking through computation* and *comparing to the physical world*. We found that most introductory students did not evaluate solutions to physics problems using expert-like strategies, many students used evaluation strategies that emphasized computation, and students used evaluation strategies that are not canonical but nonetheless useful. The analysis of our results showed our characterizations of these groups to be consistent with prior research in mathematical modeling, the use of mathematics in physics, proof/justifications in mathematics education, and control and beliefs about knowledge in metacognition.

CHAPTER 4

4. GROUPING AS AN EVALUATION STRATEGY

4.1 Introduction

One widely acknowledged goal of undergraduate science, technology, engineering and mathematics (STEM) education is the development of critical thinking [22]. In physics, one vehicle for teaching and assessing critical thinking is problem solving. Evaluation is an important aspect of critical thinking and consequently problem solving in physics [23]. In physics, evaluation entails checking to make sure 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 dimensional 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 one of the unspoken examples of what it means to "think like a physicist" [3].

Evaluation is also a component of the use of mathematics (mathematical reasoning) in physics, a heavily studied subject in PER [10]–[12], [34]. Studies of the use of mathematics in physics demonstrate that mathematics and physics are interconnected in a strong, productive, and multifaceted manner. Uhden and colleagues claim that the use of mathematics in physics has three aspects: it serves as a tool (pragmatic perspective), it acts as a language (communicative function), and it provides a means for logical deductive reasoning (structural function) [12]. The authors assert that mathematics in physics goes beyond the structural function of establishing quantitative relationships between physical quantities. For example, sometimes theoretical explanations in physics are enabled by the deductive nature of mathematical formalism. Similarly, Redish and Kuo assert that the use and meaning of mathematics is different for mathematicians and physicists: in particular, physicists load physical meaning onto symbols and equations while mathematicians do not [10].

Models of mathematical reasoning in physics emphasize and illustrate the grounding of mathematics in a physical system [10]–[12]. These models also include a step of evaluation, and show that evaluation strategies involve integrating physics and mathematics in ways that make sense physically [2]. Similarly, evaluation is a step in models of mathematical modeling [8], [9]. Evaluation is also an acknowledged step of problem solving in both physics and mathematics, and it is usually listed in problem solving rubrics [4]–[7], [24]–[26].

A few PER studies have explored student use of validity checks or evaluation strategies [2], [16]–[19], [68]. These studies have reported and described students' use of the commonly taught evaluation strategies of unit analysis, limiting cases, and using reasonable numbers. They have also documented students' use of non-traditional evaluation strategies including solving the problem, citing the presence or absence of a variable in the expression, describing the physical mechanism at play in the physical situation, and performing a limiting case analysis.

In the previous chapter, we presented a comprehensive list of evaluation strategies observed at the introductory level [68] when students were asked to check the validity of mathematical expressions describing physical scenarios in different contexts. We identified 3 broad categories of responses: *consulting external sources, checking through computation,* and *comparing to the physical world*. The responses in the *comparing to the physical world* category include a subcategory that we labelled "grouping." We define grouping as identifying a subset of a mathematical expression, or 'group,' that is bigger than one mathematical symbol, and associating it with some physical significance in the given physical scenario or context. It is not uncommon for groups to be separated by mathematical operators or the equal sign. Grouping involves making sense of terms in an expression and explaining its significance using the physics at play.

One example of a response that we have classified as using grouping comes from a student response to the written version of the inclined plane task (see Fig. 4.1). The student responded:

> "${v_f}^2 = 2gd(\sin\theta - \mu\cos\theta)$. 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".

We classify this response as grouping because the student identifies a group of symbols $g \bullet (\sin\theta - \mu\cos\theta)$ and associates them with the acceleration of the box on the incline. The student also identifies $\sin\theta \bullet g$ as the acceleration of the box without the friction between the box and the incline, and $\mu\cos\theta$ as the force of friction. These and other types of grouping will be described further in section IV.

We first noticed grouping in introductory interviews when students were asked to evaluate an expression for the electric field due to three point charges at some distance from the charges. During the interviews, 4 out of 5 pairs of students employed grouping to evaluate the given expression. The frequency of grouping during interviews on the point charge task prompted us to look through written results on the task as well as interview responses in other contexts. We believe that the process of grouping is driven by physics as the students do not just focus on mathematical operations in the expressions but on their significance of the situation at hand.

The goal of this chapter is to define and describe the phenomenon of *grouping*. To do this we will show and classify instances of grouping in the three task we administered. We will also describe an compare grouping to the phenomenon of chunking in cognitive science, and through framework of symbolic forms in PER.

4.2 Research Design / Methods

4.2.1 **Research Questions**

Despite the importance of evaluation as part of critical thinking, problem solving, mathematization and metacognition, there has been little to no research focused on how, where, and when physics students develop and employ the skills of evaluation. For instance, even though evaluation is an important aspect of problem solving, However, most research in problem solving is focused on deriving the right result or 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 [11], [24], [27], [29], [61], [62]. Research on the use of mathematics in physics has focused on helping students attach the correct physical processes to corresponding mathematical tools but such studies do not focus on evaluation. Similarly, studies involving metacognition in PER tend to focus on student reasoning during problem solving. To focus on students' understanding and use of evaluation strategies, we seek to answer the following research questions:

1. How do students make sense of an expression when they check its validity?
2. To what extent do frameworks in education research and psychology describe grouping?
3. What determines the prevalence of grouping as an evaluation strategy?

The primary purpose of our research is to explore the use of evaluation strategies as a tool for helping students meld their knowledge of mathematics and physics in a way that is both useful and profitable. Our goal is to add to the effort towards 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.

4.2.2 **Task design and administration**

To answer these questions, we designed tasks that prompted students to evaluate solutions to physics problems. The provided solutions were in form of mathematical expressions that described the physical quantity that was being sought or calculated in the problem statement. These tasks were given in both interview and written form and administered at different levels of the curriculum as well as with different problem contexts. However, for the scope of this paper, we focus on three introductory-level tasks (see Fig. 3). In each of these tasks, students were given a correct expression for a quantity: the velocity of a block at the bottom of an incline with friction; the electric field at a point some distance from three point charges of equal magnitude; or the final velocities of two masses in an elastic collision. The students were first prompted to describe how they would go about checking whether the expression was reasonable and then asked to use their suggested approaches to determine whether the expression was likely to be correct.

The written tasks were administered in the calculus-based introductory physics sequence, primarily taken by engineering majors, at a public research university in New England. The textbook used for the courses was *Physics for Scientists and Engineers: A Strategic Approach* by Knight[63]. By the time the tasks were administered in both interview and written formats, all participants had covered the relevant physics content in class. All the students received instruction through 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 the inclined plane and point charge data were collected were taught by the same instructor. Both courses had both lecture and recitation component but, weekly homework was almost completely quantitative. On the other hand, the course in which the conservation of momentum task data were collected had two lecture sections taught by different instructors, but instruction was coordinated between the instructors so that students received similar instruction and assessment. The courses had lecture , recitation, and laboratory components; weekly homework had both quantitative and conceptual components. The written data collection depended on the 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 ($5). Some interviews were individual, while others were paired. While it is not possible to eliminate all potential variables, the phenomena described appeared in our data across variation in our approach, format, and level.

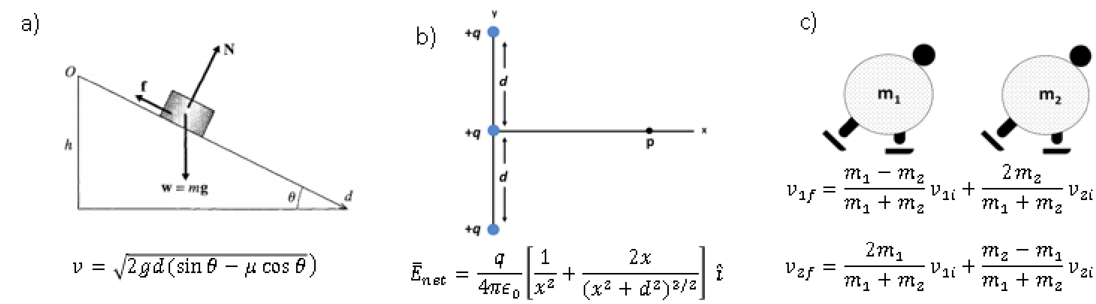


Figure 4.1: Figures and given expressions for the assigned tasks: (a) the velocity of a block at the bottom of an incline with friction; (b) the electric field at a point some distance from three point charges of equal magnitude; (c) the final velocities of two masses involved in an elastic collision

4.3 Results

We first noticed grouping in 4 out of 5 pair interviews at the introductory level on the point charge task. This prompted us to reexamine the written data in search of similar responses that we may have overlooked or coded differently earlier. This was challenging because written responses do not give the level of insight into in-the-moment thinking as interviews. It is possible that a student's expression might have also grouped terms implicitly while working on the task. As it turned out, upon reexamination, we found instances of grouping in both written and interview responses on all three tasks.

In this section, we demonstrate how we apply the definition of grouping by identifying and describing specific occurrences in our data. We will present examples from the inclined plane, point charge, and bubble skating tasks respectively. Each set of responses is described as "grouping for ____", in which the expressed physical interpretation of the group is identified explicitly. At the end of the section, we discuss the frequency of the different types of grouping in different tasks, and at different level of the physics curriculum.

4.3.1 **Data analysis**

In this section, we will first discuss how we analyzed our general data set. Then, we will focus on how we analyzed the data set specific to the strategy or phenomenon of grouping. The strategy of grouping is subset of the strategies observed in our dataset. Consequently, all the general design and data analysis processes apply to our data set specific to grouping. However, we also performed some data analysis steps exclusively on the responses coded as showing grouping.

Our overall research design and data analysis have focused on emergent patterns in the data. Written and interview data were analyzed using modified grounded theory/phenomenography, as the analysis was in part based on previous literature and there were some expectations of certain categories. We hoped to be able to identify recurring themes in student responses/reasoning. Written data were open-coded, with phrases in a response categorized based on an overall theme. For instance, on the inclined plane task, responses where students suggested plugging in numbers to check a velocity value were coded as "plug in numbers."

Interviews were conducted after the corresponding written data was collected. Consequently, data acquired from interviews were analyzed with some expectation of certain categories observed in the written data. To analyze interview data, we transcribed the videos and coded for approaches that were also present in the written data, then for new ones that emerged in the interview. Like the written responses, the interview codes were not based on the presence or absence of certain words or phrases but in the overall approach with which the student seemed to tackle the prompt.

On both the written and interview formats of the task, there were many different kinds of responses given, and most students suggested and/or used more than one approach. Furthermore, several (written) responses were not clear in describing what the student would do. In order to account for this, we rated written responses from 0 to 3 based on clarity of explanation (3 being the clearest). After running interviews in an attempt to clarify and shed light on the written responses, we re-analyzed the written responses for clarity and some of the response ratings were changed when deemed appropriate.

We found many evaluation strategies in students' responses, and these strategies were discussed in chapter 3. However, in this chapter, we will focus on the strategy of *grouping.* As a result, we will only focus on students' responses that were coded as using *grouping*. Responses were coded as using grouping when they included an explicit connection between segments of and/or the entirety of the given expression and its corresponding physical meaning as determined by the student. In addition, we also looked out for segments of the interviews where students explicitly referred to portions of the given expression or pointed to equations they had written on the board while they talked about the scenario the equation described. Furthermore, as we will show in the show in section 4.3, we characterized responses coded for grouping in the form "grouping for *x*" where *x* represents the physical significance students' associate with parts of the equation. For instance, in the inclined plane task, we further coded students' responses as "grouping for forces", and "grouping for energy".

4.3.2 **Grouping in the Point Charge Task**

The first set of examples of grouping we show are in the context of the point charge task, in which students were asked to evaluate $\frac{q}{4\pi\epsilon_0}\left[\frac{1}{x^2}+\frac{2x}{(x^2+d^2)^{\frac{3}{2}}}\right]\hat{\imath}$ (Fig. 4.1). In this task, we observed three types of grouping: *grouping for distance*, *grouping for projection*, and *grouping for electric field*. In the point charge context, responses were coded as showing grouping at the introductory, sophomore, and junior levels.

4.3.2.1 **Grouping for distance**

This type of grouping involves associating parts of the expression with the distance between the charges and the point at which the electric field is being evaluated. In the following excerpt from an interview, two first-year students evaluate the given expression:

> *Ev*: So $1/x^2$, that's the distance for the charge in the middle.
>
> *Em*: Charge in the middle, yep. So, we are just going to make sure that this [line earlier drawn from the bottom charge to point p] is the right one.
>
> *Ev*: So, that [$1/x^2$] makes sense. And then $\frac{2x}{(x^2+d^2)}$...
>
> *Em*: So that's the distance, if this [x-axis] is what we're calling x, that's $x^2 + d^2$ is r so that would make sense. And then there's two of them, that is why it is raised to the three halves.

The students were classified as grouping for distance because they connected a group of symbols to a distance on the diagram. First, they identified that $\frac{1}{x^2}$ represents "the distance for the charge in the middle." They also concluded (incorrectly) that the $(x^2 + d^2)^{\frac{3}{2}}$ makes sense because $x^2 + d^2$ is the [square root of the] distance from an off-axis charge to point *p*. The students called this distance "*r*." They also rationalized that the three-halves power is because there are two off-axis charges.

From the written data at the first-year level, examples of responses that were categorized as grouping for distance include:

"I will look @ the equation $E = \frac{kq_1q_1}{r^2}$ & compare what is written to the equation. q is the charge, k is $\frac{1}{4\pi\varepsilon_0}$, x is the radius @zero, $\sqrt{x^2+d^2}$ is the radius when +q is @d or -d. I think the result is correct because there is a k & q. The it is separated by the 2 different radi[i]."

and

"...the algebra looks alright as well, with the distance being 2 hypotenuses & x."

4.3.2.2 **Grouping for projection**

This type of grouping is more complicated than the previous examples, as it involves associating segments of the expression with the projections (or "components," in the words of some students) of the electric field vector in the *x* and *y* directions. In the following interview excerpt, two first-year students evaluate the given expression:

> *Martin*: Off the top of my head I don't know why we add it [$2x$ in numerator] on but if it's because of the components of the electric field caused by these [circles top and bottom charges] point charges and the formula says you have to add it on then I would agree with that.

Nate responded to this, drawing *x* and *y* components of the electric field due to the top and bottom charges. He then continued:

> *Nate*: This [$\frac{2x}{(x^2+d^2)^{\frac{3}{2}}}$] term is, umm, due to this [*x* component of top charge E-field] and this [*x*-component of bottom charge E-field] combined, so that is where you get the *2x* in the numerator.

In the above conversation, Martin expressed uncertainty about the physical meaning of the $2x$ in the expression and Nate proceeded to explain. He broke down the electric field vector into x and y components, showing that the y components of the E-field cancel, so that the x components of the electric field of the top and bottom charges add up to yield the $2x$ in the given expression. In the rest of the interview, Nate and Martin did not explicitly connect the terms in the expression (as a whole) to the electric fields due to the charges. However, another set of first-year students (Dom and Jake) did so.

We also show two examples of written response that included grouping for components:

"I do not know there is the "$\frac{q}{4\pi\varepsilon_0}$" outside the bracket but everything within the bracket looks reasonable for the fact that "$(x^2+d^2)^{\frac{3}{2}}$" would calculate for the y-component of the force."

"check to make sure everything in the brackets is the $\frac{1}{r^2}$ of each charge in the $\hat{\imath}$ direction. $\left[\frac{1}{x^2}+\frac{x}{(x^2+d^2)^{\frac{3}{2}}}+\frac{x}{(x^2+d^2)^{\frac{3}{2}}}\right]$. The distance from the point to $\pm d$ is $\sqrt{x^2+d^2}$, so the field would be $\frac{kq}{\left(\sqrt{x^2+d^2}\right)^2}$ the x component is found using the fact that $\cos\theta = \frac{x}{\sqrt{(x^2+d^2)}}$ when multiplied by the field, $\frac{kqx}{(x^2+d^2)^{3/2}}$, likely right."

Written responses were coded in this category if they included explicit references to components or projection, or statements like "in the i-hat direction". Another example of a response coded as showing grouping for distance and projection from the written data at the sophomore level is shown in Figure 4.2. In this example, the student connects x^2 and x^2+d^2 to the distances from the middle charge, and off axis charges to point p respectively. The student also connects the term $\frac{x}{\sqrt{x^2+d^2}}$ to the x-axis projection of the electric field.

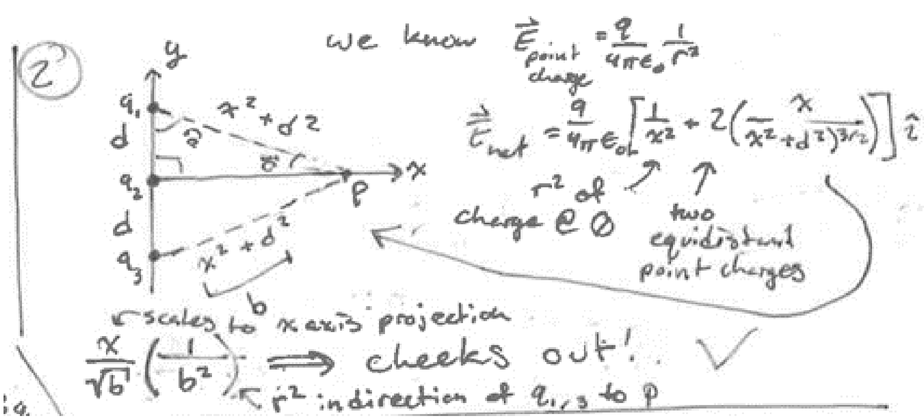


Figure 4.2: A written response showing (successful) grouping for distance and projection at the sophomore level

4.3.2.3 **Grouping for electric field**

This type of grouping involves associating segments of the expression with the expression for electric field due to a point charge. In the following excerpt, Jake and Dom make sense of and evaluate the given expression.

> *Jake*: … if you are to multiply this [$\frac{q}{4\pi\epsilon_0}$] in here [$\frac{1}{x^2} + \frac{2x}{(x^2+d^2)^{\frac{3}{2}}}$] you will get that [$\frac{q}{4\pi\epsilon_0 x^2}$] back so we know at least this [$\frac{q}{4\pi\epsilon_0}\left(\frac{1}{x^2}\right)$] part of the equation is accounting for the first [middle] charge, and without checking this [$\frac{2x}{(x^2+d^2)^{\frac{3}{2}}}$], we know it looks like we are going to be using this same, this [draws a box around $\frac{q}{4\pi\epsilon_0}$ in $\frac{q}{4\pi\epsilon_0 x^2}$] bit right here, which would be assuming the q's are the same, which they are according to this diagram. Then this [$\frac{q}{4\pi\epsilon_0}$] would just be constant.

In distributing $\frac{q}{4\pi\epsilon_0}$ into the rest of the expression, Jake compared the given expression to the equation for electric field ($E = \frac{q}{4\pi\epsilon_0 r^2}$). He recognized that the $\frac{q}{4\pi\epsilon_0 x^2}$ term accounted for the electric field of the "first" (middle) charge, while $\frac{2x}{(x^2+d^2)^{\frac{3}{2}}}$ was connected to the other two charges and accounted for the distance to each charge using trigonometry.

From the written data, an example of a response that was coded as showing grouping for projection is shown in figure 4.3. This student explicitly connects the term $\frac{kq_2}{L^2}$ to the electric field of the middle charge, and $2\frac{kq_1}{\sqrt{d^2+l^2}}\cos\frac{d}{l}$ to the x components of the top and bottom charges.

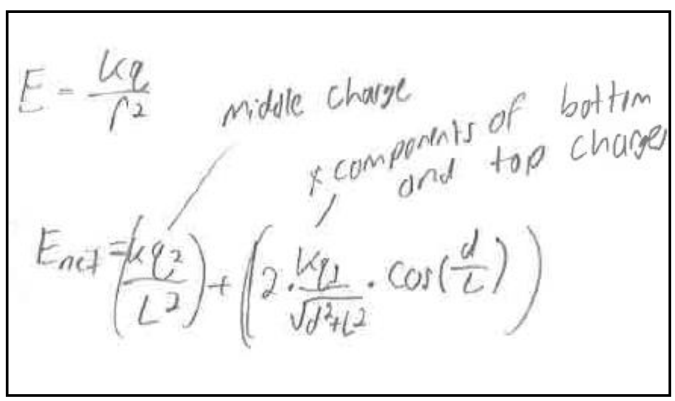


Figure 4.3: A written response coded as showing for grouping for E-field and projection at the first year level

4.3.3 **Grouping in the Inclined plane task**

The second set of examples we show are in the context of the inclined task in which students were asked to evaluate $v = \sqrt{2gd\,(sin\,\theta\, - \mu cos\,\theta\,)}$, (figure 4.1). In this task, we observed five types of grouping: *grouping for projection, kinematic quantities, forces, energy, and distances.* Inclined plane responses were coded as showing grouping at the introductory, sophomore, and junior levels.

4.3.3.1 **Grouping for projection**

This type of grouping involves associating the trigonometric portions of the expression with projections in the *x* and *y* directions. It is not unusual for this type of grouping to be associated with the orientation of the coordinate system assumed while the given expression was derived.

In the following excerpt from an interview, a first-year student attributes part of the given expression to the rotated coordinate system of the inclined plane:

> "And I think that that's there [points at$(sin\,\theta\, - \mu cos\,\theta\,)$] because um I assume that your axis is your traditional x y [draws axis parallel and perpendicular to floor], but if you were to turn it [gestures tilt and points to coordinate system perpendicular and parallel to incline], I don't think you'd have to use to those angles."

In the excerpt above, the student attributed the presence of the trigonometric terms to the fact that the given expression was derived with reference to a "traditional" coordinate system. She expressed that the expression would not contain both terms if it had been solved for using an incline-oriented coordinate system.

From the written data at the introductory level, examples of responses that were categorized as grouping for components are:

> "$2\mathrm{gd}\,\sin\theta$ y component, $\mu\cos\theta$ x component…"

"Cos is x, sin is y...I think if you drew the [coordinate system] at angle like [inclined coordinate system], your x component would be the only part to have any friction, but that might not be the case here."

4.3.3.2 **Grouping for kinematic quantities**

This type of grouping involves associating part of the expression with the direction of motion, velocity, and acceleration of the box on the incline. Responses coded as grouping for motion usually connected the trigonometric terms in the equation to the direction of motion, velocity, and acceleration of the box sliding down the incline.

In the following excerpt from an interview, one first-year student associated part of the expression to the velocity of (and force of friction on) the box sliding down the incline:

> "...it would make sense because you'll have friction going this way [draws an arrow labelled μ and pointing up slope of incline] so slowing down the velocity which is going to be equal to the acceleration times the distance. '*2*' I am not too sure about... So, then velocity would be slowing down as it is accelerating that way [gesticulates going down the incline]."

In the excerpt above, the student attributed the presence of the group gd to the velocity of the box which is slowed down by the friction which is acting in the direction up and parallel to the incline.

From the written data at the introductory level, an example of a response that was categorized as grouping for acceleration is:

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

In this response, the student associates the term $g(\sin\theta - \mu\cos\theta)$ with the acceleration down the incline. The student further associates the term $\sin\theta\, g$ as the acceleration for a frictionless incline and the term $\mu\cos\theta$ as accounting for friction.

4.3.3.3 **Grouping for forces**

This type of grouping involves associating part of the expression with forces at play in the box sliding down an inclined plane scenario. Like the previous two types of grouping in this task, it is not uncommon for responses coded as grouping for forces to refer to trigonometric terms in the equation.

Examples of written, introductory level responses categorized as grouping for forces are:

> "… $\mu\cos\theta$ associated with normal force."
>
> "…Because there is friction, the force of friction can be found to be $\mu\cos\theta$ (because $\cos\theta = N$ in this case)."

From the written data at the sophomore level, a response categorized as grouping for forces is:

> "I would check if it makes sense (each part) such as the signs and make a story of it so to speak. $2gd$ makes sense, the velocity of the block is in proportion with gravity and distance, to see how far it has travelled. $\sin\theta$ would be the y component of the force, $\cos\theta$ would be the x-component (with μ for friction). Makes sense except d. d should be an x value but not the total distance, that would mean the block at the top has the same velocity as the bottom. Incorrect."

4.3.3.4 **Grouping for energy**

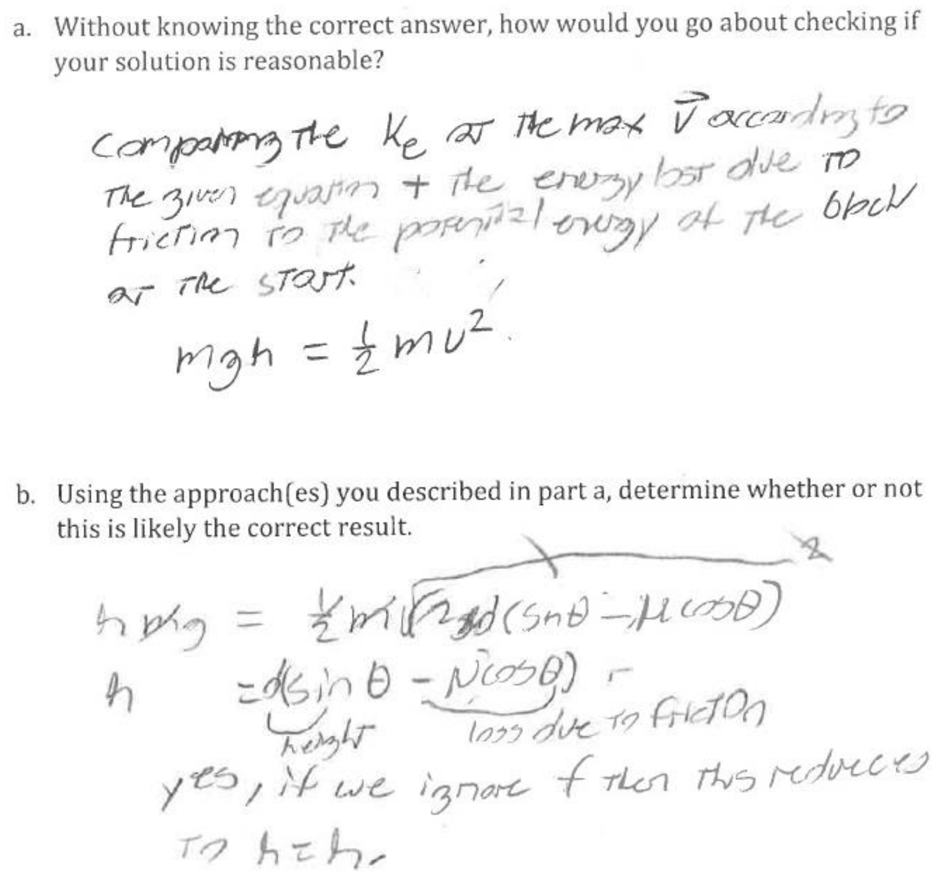
This type of grouping involves associating terms of the given expression with energies at play in the box sliding down an inclined plane scenario. In the following excerpt from an interview, one first-year student associated part of the expression to the potential energy and energy lost to friction of the box sliding down the incline:

"it [$2\mathrm{gd}\sin\theta$] is almost like a potential energy function if you had mass in it and then this [$2\mathrm{gd}\mu\cos\theta$] is like a friction… so it's the potential energy minus the friction like so you are losing the friction… it makes sense because you have energy minus energy…your work is going to be your potential energy you lose here (points at top and bottom of incline) minus this [$2\mathrm{dg}\mu\cos\theta$] so minus what you have to fight as the friction resists it trying to go down… your work is opposed by the friction so they will not be additive…"

Here the student connected the segment $2gd\,sin\,\theta$ to the potential energy without the mass included. He also connected the second term $(2dg\mu\,cos\,\theta)$ to the energy lost due to friction resisting the motion of the block. In a prior segment of the interview, he had attempted to solve for the velocity of the block using conservation of energy. In the process, calculated $v^2 = \sqrt{2gh}$ using the expression $U = hgm$. So, this might have made him perceive the segment *[*$2gd\,sin\,\theta$*]* as the potential energy without the mass.

From the written data at the introductory level, an example of a response that was categorized as grouping for energy (and distance) is shown in figure 4.4. This response was coded as grouping for energy and distance because the student explicitly connected $d\sin\theta$ to the height of the incline, and $\mu\cos\theta$ to the energy lost to friction.

From the written data at the junior level, an example of a response that was categorized as grouping for forces and energy is shown in figure 4.5. in this response, the student explicitly connected $mg\sin\theta$ to the force pushing the box down the ramp, and $mg\mu\cos\theta$ to the force of friction pushing up the ramp. The student also connected $mgd(\sin\theta - \mu\cos\theta)$ to the work done on a block as it goes down the ramp.



a. Without knowing the correct answer, how would you go about checking if your solution is reasonable?

Comparing the K_e at the max $\vec{V}$ according to the given equation + the energy lost due to friction to the potential energy of the block at the start.

$$mgh = \frac{1}{2}mv^2$$

b. Using the approach(es) you described in part a, determine whether or not this is likely the correct result.

$$h\cancel{m}g = \frac{1}{2}\cancel{m}\left(\sqrt{2gd(\sin\theta - \mu\cos\theta)}\right)^2$$

$$h = \underbrace{d(\sin\theta}_{\text{height}} - \underbrace{\mu\cos\theta)}_{\text{loss due to friction}}$$

yes, if we ignore f then this reduces to h=h.

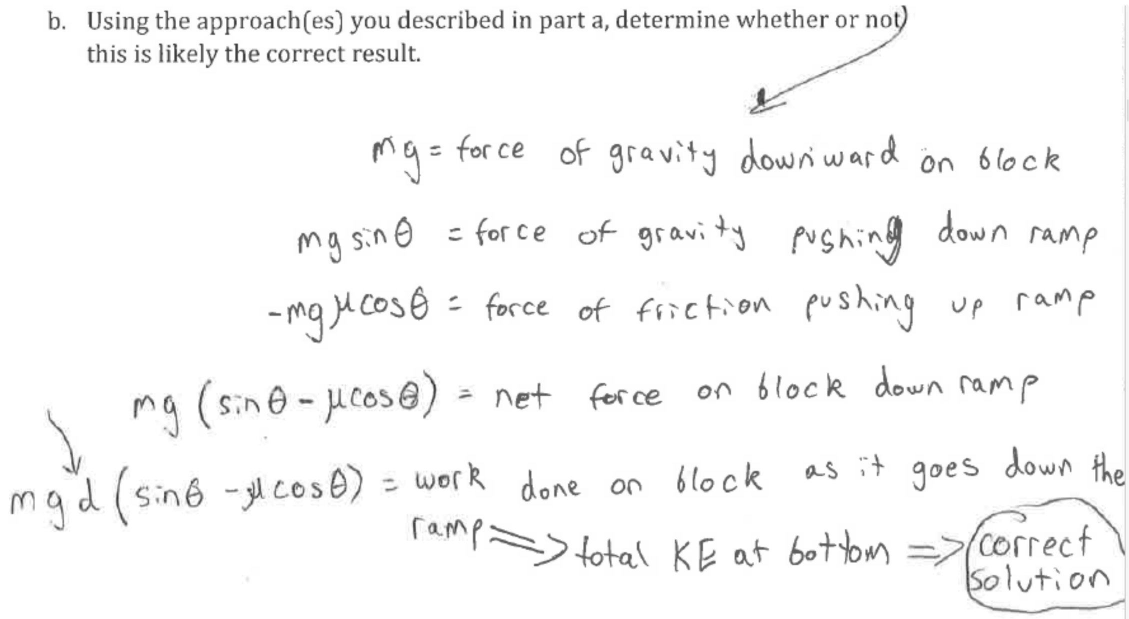
Figure 4.4: An example of a student's response at the first-year level showing grouping for distance and energy

4.3.3.5 **Grouping for distances**

This type of grouping involves connecting parts of the expression with the physical dimensions of the incline including its height, length, and hypotenuse. In the following excerpt from an interview, one first-year student associated part of the expression to the height of the incline:

> "...like why is that here? I can go back and ask myself like why does that make sense as an answer. For me, I am going to write this [writes $v^2 = 2gd\ sin\ \theta - 2gd\mu cos\ \theta$]. That makes sense to me ...To me if $2g$ is pulled out then d times $sin\ \theta$ would give me this [Draws an incline labelled with height h and length d], well would give me the side here [points at side h] and then $d\mu cos\ \theta$...I would be given this side [adjacent] which makes sense in my head because

there is a coefficient of friction...to me that makes sense because it has both components but its moving in the x-direction due to friction"



b. Using the approach(es) you described in part a, determine whether or not this is likely the correct result.

mg = force of gravity downward on block

$mg \sin\theta$ = force of gravity pushing down ramp

$-mg\mu\cos\theta$ = force of friction pushing up ramp

$mg(\sin\theta - \mu\cos\theta)$ = net force on block down ramp

$mgd(\sin\theta - \mu\cos\theta)$ = work done on block as it goes down the ramp $\Rightarrow$ total KE at bottom $\Rightarrow$ correct solution

Figure 4.5: A Junior's response coded as grouping for forces and energy

In the above excerpt, the student connected portions of the expression to the height of the incline. First, she reparsed the equation into $v^2 = 2gd\ sin\ \theta - 2gd\mu cos\ \theta$, a form she claimed made more sense to her. She then rewrote the first part of the reparsed expression as $2g(d \sin\theta)$ and finally grouped $dsin\ \theta$ as the height of the incline. She also associates the term $d\mu cos\ \theta$ as the adjacent side of the inclined plane.

In the following excerpt from an interview, another first-year student associated part of the expression with height of the incline:

"...like you would apply the variables to each thing you know that like this is the 2gd is the two times gravity and the distance [...] and sin θ at least to me will represent where its starting at the incline at zero [points at height of incline]… so I am using I'm like breaking apart each thing that I have and applying it to make sure that it actually goes with the equation like it makes sense. That's how I would check it."

In the above excerpt, the student connected $\sin\theta$ to the height of the incline. Perhaps more importantly, she also gave her definition of what we would later call grouping; breaking apart the expression and making sure each segment makes sense with the physical scenario the expression is supposed to describe.

From the written data at the introductory level, examples of responses that were categorized as grouping for incline dimensions include:

"$\sin\theta$ associated w/ height (h). $\mu\cos\theta$ associated with normal force."

"It makes sense that the velocity is equal to $g.d$ because the block must travel that distance, while accelerating. This is multiplied by the $\sin\theta$ minus the coefficient of friction of the $\cos\theta$, or adjacent, part of the triangle that the block travels down."

4.3.4 **Grouping in the Bubble Skating Task**

The third set of examples is in the context of the conservation of momentum task, where students were asked to evaluate $v_{1f} = \frac{m_1-m_2}{m_1+m_2}v_{1i} + \frac{2m_2}{m_1+m_2}v_{2i}$, $v_{2f} = \frac{2m_1}{m_1+m_2}v_{1i} + \frac{m_2-m_1}{m_1+m_2}v_{2i}$ (Fig. 4.1). In this task, we observed four types of grouping: *grouping for mass combinations, grouping for momentum, grouping for velocity, and grouping for mass ratios.* Unlike the other tasks, responses code as showing grouping in the bubble skating task were only present in the written data at the introductory level. Above the introductory level, there was only one instance of a response coded as grouping: a junior during an interview.

4.3.4.1 **Grouping for mass combinations**

This type of grouping involves associating terms of the given expression with behaviors/ combinations of the masses involved in the collision. In this type of grouping, students connect mass terms in the equation to mass changing, bouncing off, or sticking to each other. Examples of responses that were categorized as grouping for mass combination are:

"Neither velocity equations should have $2m_2$ or $2m_1$ since this is an elastic collision, they are never one combined mass".

"The above method seem to be more inelastic collisions involving masses added together."

4.3.4.2 **Grouping for momentum**

This type of grouping involves associating terms of the given expression with the momentum of the masses involved in the collision. At the introductory level, examples of responses that were categorized as grouping for momentum include:

"Looks correct as you'll have a sum of velocities fractionally related by mass to the velocities, so momentum".

"They are finding the averages of the masses in the equations (not exactly but similar process) and multiplying it by velocities to get momentum, the format looks similar to finding uncertainties."

Outside the introductory level, the response that was categorized as grouping for momentum is:

"We have this weird mass equation $[\frac{m_1-m_2}{m_1+m_2}]$ multiplied by a velocity plus this weird mass $[\frac{2m_2}{m_1+m_2}]$ equation multiplied by a velocity. That looks like a momentum."

4.3.4.3 **Grouping for velocity**

This type of grouping involves associating terms of the given expression with the velocities of the masses involved in the collision in the given scenario. Examples of responses that were categorized as grouping for velocity include:

> "… The equations will not give the correct answer because it is giving the combined velocities for 1 and 2"
>
> "…you can't add factors of the initial velocities together to get the final velocities in the way that is presented here".

4.3.4.4 **Grouping for mass ratios**

This type of grouping involves associating terms of the given expression with mass ratios in the given scenario. Examples of responses that were categorized as grouping for mass ratios include:

> "This is likely not the correct result. They are using fractional masses for some reason."
>
> "They are finding the averages of the masses in the equations (not exactly but similar process) and multiplying it by velocities to get momentum, the format looks similar to finding uncertainties."

4.3.5 **Prevalence of grouping responses**

The prevalence of responses coded as showing grouping in the written data is summarized in Table 4.1 and Figure 4.6. In both the inclined plane and point charge tasks, the prevalence is higher in the non-introductory levels than the introductory level. In the bubble skating task, the use of grouping disappears in the sophomores. (The task was not administered in writing to the junior/senior cohort.)

4.3.6 Summary of results

In summary, there are three important points from the results presented above. First, although there are no previous reports on grouping, we find that it is a strategy that is widely used to evaluate physics equations. We have described examples of grouping observed in all of the contexts on which we have collected data. Grouping involves explicitly tying segments of an equation to a physical quantity or process. Grouping involves students saying the story that the equation tells about the physical scenario it describes.

Secondly, the prevalence of grouping as a strategy in sensemaking and validity checking seems to vary with context. For example, we identified more examples of grouping in the point charge task than in the bubble skating task. However, it is possible that grouping is happening, but we are unable to identify it due to limitations of our data set.

Table 4.1: Summary of the prevalence of grouping in written versions of tasks

Task/grouping type	First-year	Sophomore	Junior/Senior
Point charge	**11% (n=170)**	**32% (n=22)**	**50% (n=18)**
For distance	11/18	6/7	5/9
For projection	6/18	3/7	2/9
For electric field	9/18	3/7	8/9
Inclined plane	**6% (n=211)**	**27% (n=11)**	**30% (n=20)**
For projection	4/12	1/3	0/6
For force	5/12	2/3	2/6
For energy	1/12	0/3	2/6
For distance	5/12	0/3	2/6
For kinematic quantities	4/12	0/3	3/6
Bubble Skating	**5% (n=190)**	**0% (n=22)**	**N/A**
For mass combinations	6/10	0	
For momentum	3/10	0	
For velocity	2/10	0	
For fractions	2/10	0	

Thirdly, we claim that grouping is a sophisticated strategy that reflects mathematical sense-making. Because grouping seeks to explicitly associate mathematical symbols with a physical interpretation, it implicitly stems from the belief that physics equations are mathematical representations of the laws of the physical world and therefore should accurately describe the physical scenario painted in the problem statement. As a result, students might evaluate an expression by checking to see if it reflects expected aspects of the physical situation, e.g., the distance from the charge to the point where electric field is being evaluated or the force of friction. Grouping can also involve reformatting a given expression) as in the distribution for the electric field example) to produce a form that can be interpreted in the context of the task /problem statement.

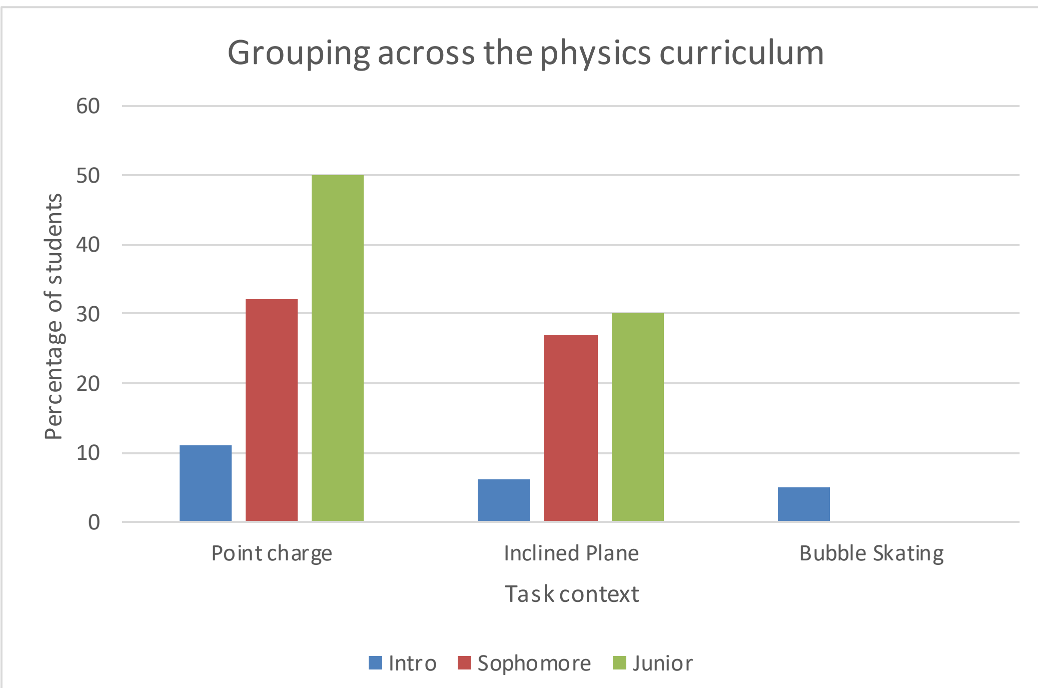


Figure 4.6: Summary of the prevalence of grouping in written responses

Specifically, grouping is a demonstration of the communicative and structural functions of mathematics in physics [12] by loading meaning onto symbols and expressions, a behavior of physicists, but not mathematicians [12]. Specifically, when students use grouping, the given mathematical expression is used to tell the story of the physical scenario it describes, and the equation statement facilitates the use of logical deductive reasoning.

Consequently, grouping is a productive strategy because it is students using mathematics in physics as physicists desire [10], [12]. The students were making sense of the physical world/context of the problem using the given mathematical equation. In this way, the students were exhibiting the desired physics learning outcome of coherence between physical meaning and mathematical formalism.

4.4 Discussion

4.4.1 Connecting our results to cognitive psychology

To delve further into the phenomenon of grouping, we compare it to the phenomenon of chunking in cognitive science. We illustrate how chunking matches grouping by discussing the characteristics that grouping shares with chunking including the definition and nature of chunks (automatic vs deliberate), the nature of characterization, the structure of chunk templates, the process of chunk decomposition and correlation with expertise.

First, grouping is consistent with the definition of chunking and the nature of chunks. Like chunking, grouping involves gathering of elements that have strong associations with one another. In responses that we categorized as showing grouping, symbols of the equation were gathered together. Like chunks, these groups were also assigned meaning. Furthermore, like chunks, groups seem to be familiar as the students state the significance of an identified group. Consistent with chunking, these statements of the physical significance of identified groups often include useful physics concepts relevant to the physical contexts of each task including the vector nature of the electric field, conservation of energy, the work-energy theorem, conservation of momentum, and the nature of elastic and inelastic collisions.

Secondly, grouping shares characteristics with both automatic and deliberate chunking and may reflect a combination of these behaviors. For instance, grouping has characteristics of automatic chunking as it involves instantly recognizing the forms of the given expression. It is hard to determine whether grouping is implicit or explicit as the nature of our research task requires students to explicitly explain their reasoning. It is not clear, however, whether grouping is associated with long term memory as is automatic chunking. On the other hand, grouping also has elements of deliberate chunking. For instance, by the nature of our research task, grouping is conscious and explicit since students are asked to explain their reasoning. During interviews, one student described what we later coded as grouping as "so I am using, I'm like, breaking apart each thing that I have and applying it to make sure that it actually goes with the equation like it makes sense. That's how I would check it." Like deliberate chunking, grouping is goal-oriented, the goal being to determine whether the given expression is reasonable by connecting it to the physical context. Like deliberate chunks, groups are explicitly defined by the student and are readily explained, described, and justified.

Specifically, grouping is very similar to characterization since it involves grouping terms for meaning, utility, or significance. Grouping involves gathering symbols in equations based on their physical significance and not just spatial proximity. Thus, for instance, in the point charge scenario, the sequence of variables x^2, and $\frac{1}{x^2}$, are associated with the distance from the charge to point p while $1/4\pi\varepsilon_0$ is chunked as *k,* Coulomb's constant. Similarly, in the inclined plane context, $\mu\cos\theta$ is chunked as the energy lost to friction. Finally, in the conservation of momentum context, $2m_2$ and $2m_1$ are characterized as the masses being stuck together.

Some of the responses categorized as grouping are consistent with chunking template theory. Specifically, in the point charge task, it seems that some students are working with the template $E = {kq}/{r^2}$. First, this is a pattern that the students recognize from seeing the equation in class and working with it in homework and lab. The core of this template specifies the presence and placement of the symbols in the expression, e.g., k and q in the numerator and r^2 in the denominator. This template also contains *slots* that contain variables that can be altered: the charges that produce the electric field, and the distance between the charges and the point at which the electric field is being evaluated.

A quick look at the results of the inclined plane tasks suggests that there are many associations students have with the given expression for velocity, including force and energy. Consequently, grouping for forces, energy, and distances are consistent with force, energy, and distance templates, respectively. The cores of the templates define how the problem context gives rise to certain forces, energies, and distances, e.g., the presence of a box, coefficient of friction, and an inclined plane. The corresponding slots are different kinds of forces, energy, and physical dimensions at play in the context. For instance, the slot of forces can be filled with forces of friction and gravity. The slots for energy can be filled with kinetic energy, potential energy, and energy lost to friction. The slots for dimensions are filled with height, base, or length of the incline. Students' familiarity with the templates or slots make them salient and thus easy to recognize and group.

In the bubble skating task, responses coded as grouping for momentum are consistent with the template $m * v$ The core of the template includes the mathematical relationship between, and the presence and placement of the symbols in the equation. The slots are the masses and velocities (initial and final) of the bodies involved in the collision. Even though the given expression is one of velocity, this momentum template may be salient because the equation has the perceptual form $[v] = [M][V] + [M][V]$ as highlighted by the student in section [4.x.x.x]. Furthermore, the salience of this template may be due to the students' experience with the canonical equation for momentum: $p = mv$ and $mv_{1i} + mv_{2i} = mv_{1f} + mv_{2f}$.

In our observations, salience of terms is aided by reparsing the given expression to familiar or recognizable form. For instance, in the point charge task, grouping for electric field entails implicitly or explicitly multiplying $\frac{q}{4\pi\epsilon_0}$ through into $[\frac{1}{x^2} + \frac{2x}{(x^2+d^2)^{3/2}}]$ to get $[\frac{q}{4\pi\epsilon_0}\frac{1}{x^2} + \frac{q}{4\pi\epsilon_0}\frac{2x}{(x^2+d^2)^{3/2}}]$. Then $\frac{q}{4\pi\epsilon_0}\frac{1}{x^2}$ is chunked as the electric field of the middle charge and $\frac{2x}{(x^2+d^2)^{3/2}}$ is chunked as the electric field due to the top and bottom charges. Similarly, some rewrote $\frac{2x}{(x^2+d^2)^{3/2}}$ as $2\left(\frac{x}{(x^2+d^2)^{3/2}}\right)$ to highlight the association between the term $\frac{2x}{(x^2+d^2)^{3/2}}$ and the two equidistant point charges, e.g., sample responses in section 4.3.1.2.

Likewise, in the inclined plane context, many responses coded as showing grouping included reparsing $v = \sqrt{2gd(\sin\theta - \mu\cos\theta)}$ as $v = \sqrt{2(gd\sin\theta - gd\mu\cos\theta)}$ or even $v^2 = 2(gd\sin\theta - gd\mu\cos\theta)$. These new representations might have aided the recognition of chunks $gd\sin\theta$ as the potential energy of the block at any point on the incline, $d\sin\theta$ as the height of the incline, and $gd\mu\cos\theta$ as the energy lost to friction as the block comes down the incline. Note that the energy examples of these terms are not actually energy terms; they are missing a mass, possibly suggesting a further reparsing.

Grouping is also consistent with chunk decomposition since pointing out the chunks within the given expression is somewhat like breaking the expression into its constituent chunks. In this regard, following the steps of Knoblich, Ohlsson, Haider, and Rhenius [49], we believe that the ease of decomposing the given into chunks is dictated by chunk tightness: how easy it is to perceptually divide the given expression into useful chunks. In the context of our study, a useful chunk is a term that is familiar and recognizable in the context of physical scenario outlined in a research task. As with chunking, we believe that reparsing the given expression facilitates chunk decomposition.

Implementing the definition of a useful chunk, we believe that the tightest expression is that of the bubble skating task as perceptually, it does not resemble anything that students have seen before or have experience with. The tightness of the expression might explain why grouping was not a productive evaluation strategy for the bubble skating context as grouping does not provide enough information for a decision to be made about the validity of the expression.

On the other hand, the expressions in the inclined plane and point charge tasks are loose as they both contain groups of symbols that are familiar to students. As we have discussed above, and shown in our results, the expression in the inclined plane task is associated with the following quantities: projection, kinematic quantities, forces, energies, and distances. Similarly, the expression in the point charge task is associated with distances, projection, and electric field. However, unlike the bubble skating task, these associations are productive to evaluating the respective expressions.

We believe that the projection term in the expression for the point charge task is tighter than projection term in the expression for the inclined plane task. Both the inclined plane and point charge tasks were coded as including grouping for projection. However, the terms that are associated with projection in the inclined plane ($\sin\theta$ and $\cos\theta$) may be more salient than the term associated with projection in the point charge task $\left(\frac{x}{(x^2+d^2)^{1/2}}\right)$. The projection term in the point charge task is less familiar to students. This difference in salience might explain why some students had issues with the projection term in the point charge task, and why fewer students were categorized as grouping for projection than distance and electric field.

Lastly, consistent with chunking theory, grouping seems to be correlated with expertise [45]–[47], [53], [54]. On both the point charge and incline plane task, the percentage of students categorized as using this strategy increased as students moved up the physics curriculum. In written responses, chunking does not show up after the first year. This is probably because students became more sophisticated and realized that the tool was not productive in that problem context. Due to the small size of our data sample, it is hard to tell if the size of groups increased as students moved up the physics curriculum. However, it is worth looking into as the percentage of juniors coded as grouping for electric field (the biggest group) in the point charge task was higher than the sophomores and first years.

In conclusion, we argue that grouping is a form [instantiation] of chunking in physics problem solving. First, grouping is consistent with the definition of chunking and the nature of chunks as described in prior literature. Grouping shares characteristics with both automatic and deliberate chunking. Specifically, grouping is very similar to characterization since it involves grouping te rms for meaning, utility, or significance. Grouping is also consistent with chunking template theory. A few templates are identified cross all three tasks. The salience of some grouped terms is aided by reparsing the given expression to familiar or recognizable form. Grouping is also consistent with chunk decomposition, since pointing out the chunks within the given expression is somewhat like breaking the expression into its constituent chunks. We believe that the expression in the bubble skating task is tighter than the expressions in the other task because it does not resemble expressions that students have seen or had experience working with. Lastly, consistent with chunking theory, grouping seems to be correlated with expertise. Truly identifying groups created by chunking is more problematic and could be verified more precisely by examining eye movement (eye tracking) and exploring pauses in speech [58][69].

4.4.2 **Connecting our results to frameworks in physics and mathematics education research**

To zoom out of our in-depth comparison of grouping with chunking, we now examine grouping from the perspective of frameworks in physics and mathematics education. Specifically, we compare grouping to Sherin's symbolic forms and interpretative devices [39], [41].

Some of the kinds of grouping that we coded for share semblance with symbolic forms. Specifically, some of the responses that were coded as grouping are consistent with the *parts of a whole, opposition*, and *coefficient* symbolic forms.

Responses coded as grouping for electric field in the point charge task include associating each term in the expression to part of the total field due to individual charges ($E_{Total} = E_1 + E_2 + E_3$), thus invoking the *parts of a whole* form. Expressions for the electric field due to a set of point charges exhibit properties of superposition upon inspection, possibly cueing the connection of individual terms with individual charges (and locations).

In the inclined plane task, responses coded as showing grouping for forces are consistent with the *opposition* symbolic form. The expression for velocity is made up of two groups: a force of gravity group in opposition of the force of friction group. This explains responses coded as grouping in which students say that friction opposes motion. Responses coded as showing grouping for energy are consistent with the *whole – part* symbolic form such that the expression is grouped into the total potential energy of the system minus the energy lost to friction. Similarly, grouping the expression for *x* and *y* velocity components is consistent with *parts of a whole*.

Finally, in the bubble skating task, responses coded as grouping for velocity are consistent with a combination of the *parts of a whole* and *coefficient* symbolic forms such that the given expression is a sum of terms that are coefficients of the initial velocities of the skaters involved in the collision. We believe that the mathematical operators ($-$ and $+$) separating terms make them more salient.

From the perspective of interpretative devices [41], grouping is used as a narrative or static class device. When students tell the story of the given equation, they allow some terms or variables to vary while others are held fixed. However, the parameters that the students allow to vary are those that actually vary during the scenario that the equation describes, just like in the use of the physical change interpretative device. Sometimes, when using grouping, an equation is interpreted by comparing the situation the equation describes with a different situation, e.g., when students compare the inclined plane task to one where there is not friction between the box and the incline. In the point charge task, students use the generic moment interpretative device because the expression for an equation is viewed as describing any moment in a motion or statements that are true at any time during a motion. This is because in the steady state moment interpretative device, the equation for electric field describes a system where no parameters vary with time.

In addition to Sherin's symbolic forms and interpretative devices, other frameworks give some insight into the use of grouping as an evaluation strategy. In Chapter 3, we described how grouping and other strategies in the comparing to the physical world categories fit with the frameworks of epistemic frames in PER, proofs and justifications in mathematics education research, and metacognition in cognitive science. From the perspective of epistemic frames in physics, grouping is consistent with Bing's *physical mapping* frame [20]. While using grouping, 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. Students also attach physical information to symbols, signs and operations. Grouping is also consistent with the *mapping mathematics to meaning* epistemological game. When grouping, students develop a conceptual story about the physics equation that they have been given to evaluate. However, unlike the context of the epistemic game, our students are not trying to solve a problem; instead, they are evaluating the solution to one.

From the perspective of proofs and justification, grouping fits in the transformational proof scheme. Grouping involves perceiving the underlying structure behind patterns, including unpacking the contextual meaning of symbols in a mathematical expression and involving reasoning aimed at settling the conjecture put forth: in this case, evaluating the given expression. Finally, from the perspective of metacognition, grouping is consistent with control and the ability to self-evaluate. Grouping is also consistent with a belief that physics is consistent with the real world.

4.5 Conclusion

The main intent of our study was to probe the kind of strategies students use when they evaluate the solutions to physics problems. In analyzing tasks that probe student use of evaluation strategies in problem solving, we documented several instances of students making sense of an expression by grouping terms into sub-expressions that have physical significance (e.g., Coulomb's Law). While we expected strategies like solving for the given expression, unit analysis, and special case analysis, we did not expect the strategy of grouping because it was not observed in prior studies on evaluation strategies. We outlined instances and examples coded as showing grouping inclined plane, point charge and bubble skating tasks. We also gave some general notes on our results and argued for the sophistication of grouping from the perspective of the use of mathematics in physics. To delve into the phenomenon of grouping, we presented an in-depth comparison between grouping and the phenomenon of chunking in cognitive science. We also briefly examined grouping from the perspectives of epistemic games and frames, proofs and justifications, and metacognition.

The consistency of grouping with chunking, symbolic forms, and epistemic games suggests that grouping is a general sensemaking strategy that we could see in non-evaluation examples. Because of the coupling of mathematical and physical meaning, grouping should be considered an expert-like skill, consistent with frameworks that model mathematical reasoning in physics [11], [20], [29]. Chunking theory suggests that grouping is correlated with expertise. Consequently, we encourage physics instruction to include the strategy of grouping whenever it is a productive to an equation or physics context. Grouping is not a canonical evaluation strategy; however, it is a sophisticated evaluation and sensemaking strategy that instructors should look out for and encourage when they see.

CHAPTER 5

5. EVALUATION STRATEGIES: EXPERT, NOVICE, AND THE JUICY IN-BETWEEN

5.1 Introduction

One preeminent goal of a science education us the ability to think critically [22]. One way that physics fosters critical thinking is through problem solving. In physics, evaluation entails checking to make sure 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 dimensional analysis, considering limiting cases, using approximations, predicting the effects of changes in problems and identifying errors in solutions [2]. As described in earlier chapters, the ability to evaluate a solution is one of the unspoken examples of what it means to "think like a physicist" [3]. Furthermore, in both physics and mathematics, there is consensus that expert problem solving entails evaluation [5], [6], [26], [61].

Expertise is a well-studied area in many fields. For instance, psychologists have studied expertise from the perspective of memory and perception in the fields of medicine, sports, chess, engineering, and physics [54]. Physics education researchers have also studied expertise in problem solving. For instance, novice vs. expert studies in PER have shown that experts categorize problems based on the physics concepts behind them while novices categorize them based on surface features such as a falling object or an object on a ramp [66]. In mathematics education research, one finding from a study that compares novice and expert problem solving is that experts spent more time on planning a problem-solving route/technique than novices do [6][14]. Novices tended to go straight to implementing whatever problem-solving techniques they had in mind while experts spent more time planning and reflecting on their progress. Other studies show that while experts base their solution techniques on the fit between mathematics and the laws of physics at play in a physics problem, novices tend to base their solution on information from an authority, or the correctness of computation or an algorithm [20], [21]. We add to this existing body of knowledge in two ways: a) we focus on the step of evaluation in problem solving, b) we not only talk about the ends of the novice-expert spectrum, but we also examine phases in between.

5.2 Research design/methods

5.2.1 Research questions

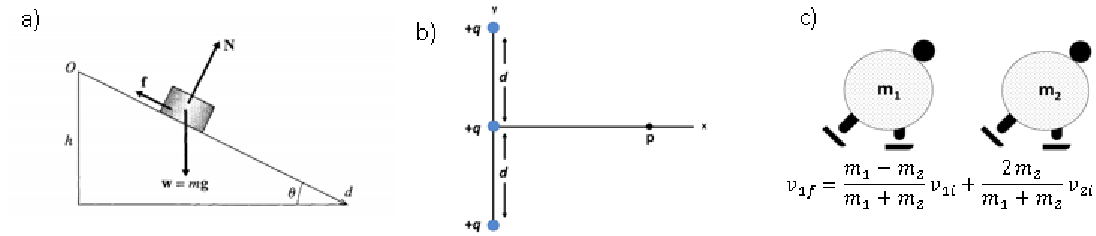
We aim to contribute to the prior research on evaluation by studying how students' use of evaluation strategies vary over the course of the major. To this end, we intend to answer the following questions:

1. What are the evaluation strategies that students employ to evaluate the solutions of physics problems?
2. What are the similarities and differences between strategies used by students across the physics curriculum?
3. What are some differences and similarities between novice and expert evaluation?

5.2.2 Research design

To answer these questions, we designed tasks that prompted students to evaluate solutions to physics problems. The provided solutions were in form of mathematical expressions that described the physical quantity that was being sought or calculated in the problem statement. These tasks were given in both interview and written form and administered at different levels of the curriculum as well as with different problem contexts. However, to limit the scope of this paper, we focus on three introductory-level tasks (see Fig. 5.1). In each of these tasks, students were given a correct expression for a quantity: the velocity of a block at the bottom of an incline with friction; the electric field at a point some distance from three point charges of equal magnitude; or the final velocities of two masses in an elastic collision. The students were first prompted to describe how they would go about checking whether the expression was reasonable and then asked to use their suggested approaches to determine whether the expression was likely to be correct.

The written tasks were administered in the calculus-based introductory physics sequence for engineers at a public research university in New England. The textbook used for the courses was *Physics for Scientists and Engineers: A Strategic Approach* by Knight [63]. By the time the tasks were administered in both interview and written formats, all participants had covered the relevant physics content in class. All the students received instruction through 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 the inclined plane and point charge data were collected were taught by the same instructor. Both courses had both lecture and recitation component but weekly homework was almost completely quantitative. On the other hand, the course in which the conservation of momentum task data were collected had two sections co-taught by different instructors so that students received similar instruction and assessment. The courses had both lecture and recitation component and weekly homework had both quantitative and conceptual components. The written data collection depended on the 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 ($5). Some of the interviews were individual, while others were paired. While it is not possible to eliminate all potential variables, the phenomena described appeared in our data across variation in our approach, format, and level.



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

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

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

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

Figure 5.1: 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 involved in an elastic collision

The first question was largely answered in Chapter 3. However, in this chapter, we briefly revisit categories that we found. We then show how the prevalence of these strategies evolve in different populations. We explore these questions using both quantitative and qualitative data from students' responses to research tasks administered at different levels of the physics curriculum. Furthermore, we examine our data from different perspectives including proofs and justifications in mathematics, the use of mathematics in physics, and students' understanding of equations.

To answer our research questions, we will be using both qualitative and quantitative data. The quantitative data is from written student data at the introductory, sophomore, and junior/senior levels of the physics curriculum. The number of written responses collected are summarized in table 5.1. The qualitative data involves case studies of two pair interviews, one from the introductory level and the other from the junior/senior level of the physics curriculum. The introductory case study was one of five pair interviews, while the junior case study was one of three pair interviews. These two pairs of individuals were chosen because they were a good representation of other interviews of their colleagues for the same level of the physics curriculum. Every strategy used in both interviews was also used in at least one other interview for each respective curriculum level. Both groups of students were also the most concise in their answering of the questions on the tasks. Consequently, these pair of interviews allow us to hit the most points of comparison between both groups of students. They highlight the differences between upper-division and lower-division students and speak to the trends that we saw in our overall qualitative data set. The responses of the chosen pair of students are consistent with many hours of interviews with both upper and lower division students. The students' responses are also consistent with evaluation strategies that we observed in students' written responses.

5.2.3 **Data Analysis**

Written data were analyzed using modified grounded theory/phenomenography, as the analysis was in part based on previous literature and there were some expectations of certain categories. For instance, interviews were conducted after the tasks had been conducted in written form, thus data acquired from interviews were analyzed with some expectation of certain categories. Also, data analysis was done with previous work like Loverude's study and Bing's epistemological frames in mind [2], [20]. We hoped to be able to identify recurring themes in student responses/reasoning. Our research design and data analysis have focused on emergent patterns in the data. Written data were open-coded, with phrases in a response categorized based on an overall theme. For instance, on the inclined plane task, responses in which students suggested plugging in numbers to check a velocity value were coded as "plug in numbers." To analyze interview data, we transcribed the videos and coded for approaches that were also present in the written data, then for new ones that emerged in the interview. Like the written responses, the interview codes were not based on the presence or absence of certain words or phrases but in the overall approach with which the student seemed to tackle the prompt.

On both the written and interview formats of the task, there were many different kinds of responses given, and most students suggested and/or used more than one approach. Furthermore, several (written) responses were not clear in describing what the student would do [give example]. In order to account for this, we rated written responses from 0 to 3 based on clarity of explanation (3 being the clearest). After performing interviews in an attempt to clarify and shed light on the written responses, we re-analyzed the written responses for clarity; some of the response ratings were changed when deemed appropriate.

Table 5.1: Summary of the number of written responses collected

	Number of Students		
Year	Inclined Plane	Point Charge	Bubble Skating
First Year	215	174	191
Sophomore	11	22	22
Junior	20	18	N/A

5.3 Findings

In earlier work, we described in detail the kinds of evaluation strategies observed at the introductory level of the physics curriculum. Here we present instances of the use of evaluation strategies at the intermediate and junior/senior level. We use quantitative data to see whole group trends, with particular emphasis on how the students' responses compare at different levels of the physics curriculum. Thus, the quantitative data addresses the first and second research question. Then, to illustrate some of the phenomenon that we observe in the quantitative results, we take an extended look at a pair of interviews as case studies that encapsulate similarities and differences between lower- and upper-division students. Consequently, we will use qualitative data to answer our third research question.

In chapter 3, 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* (Figure 5.2)[68]. Strategies in the *comparing to the physical world* category involve evaluating the given expression by checking whether it is consistent with prior 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 source* category involve evaluating given expressions by checking with a trusted external source.

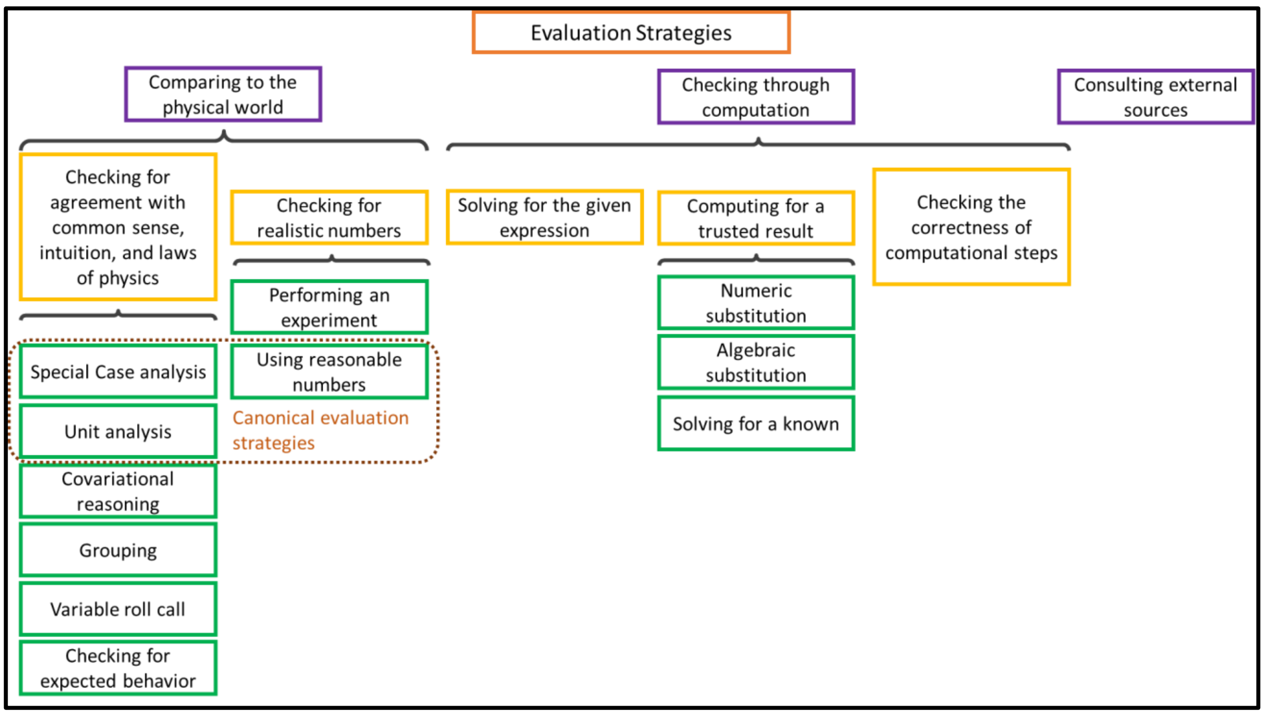
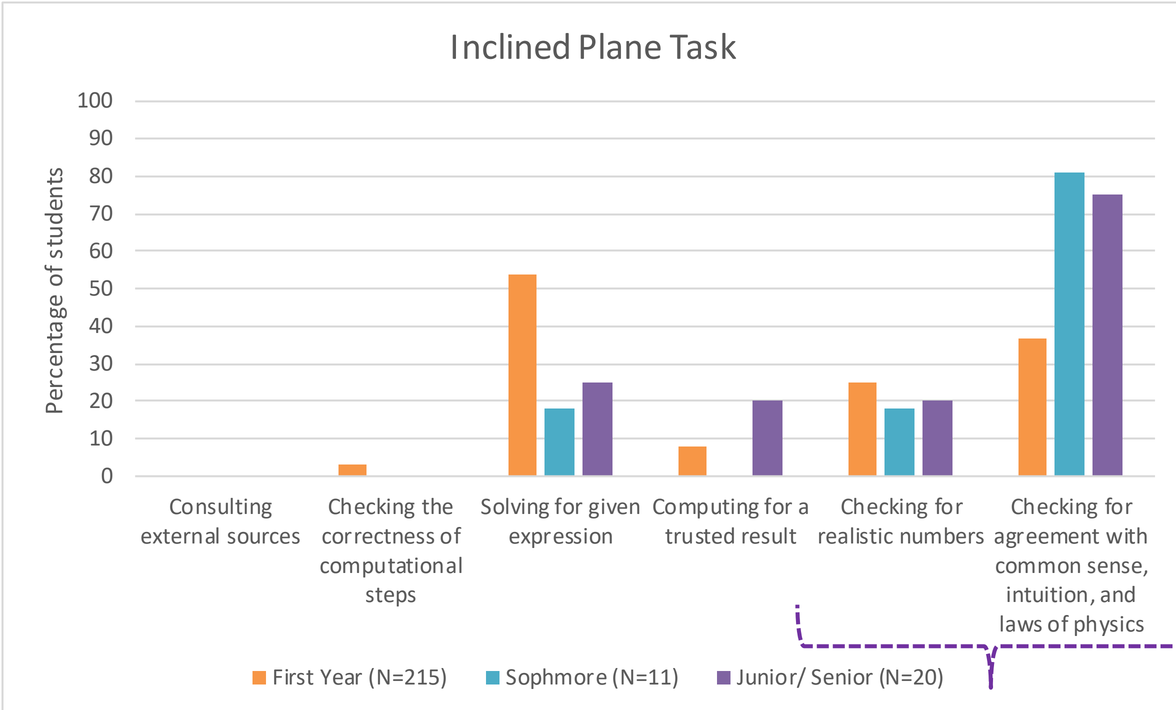


Figure 5.2: Breakdown of categories of evaluation strategies

5.3.1 **Quantitative Data**

We present the results of coding students' written responses in three different contexts at the first year, sophomore, and junior levels of the physics curriculum. For efficiency of representing these data sets, some to the code categories have been combined. Unless explicitly stated, any evaluation strategy that is missing from/unmentioned in the graph was not used by students in the given task context. Figure 5.3, 5.4, and 5.5 show the changes in the prevalence of the use of evaluation strategies in the inclined plane, point charge and bubble skating contexts respectively.



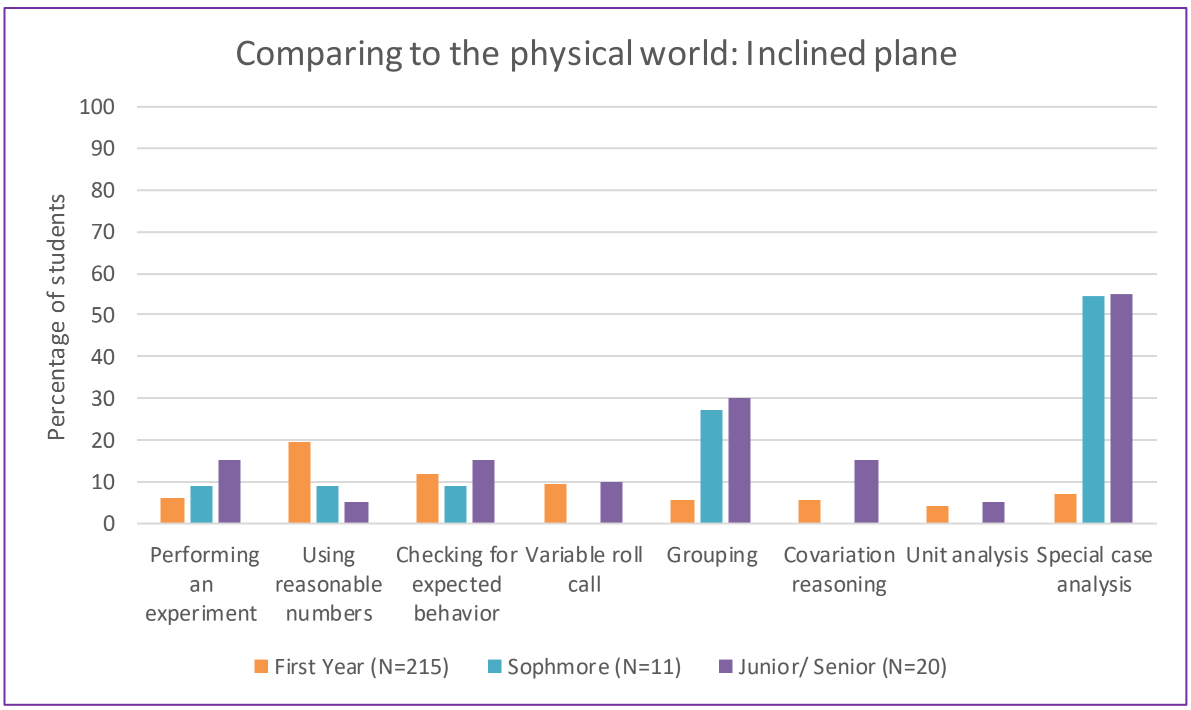
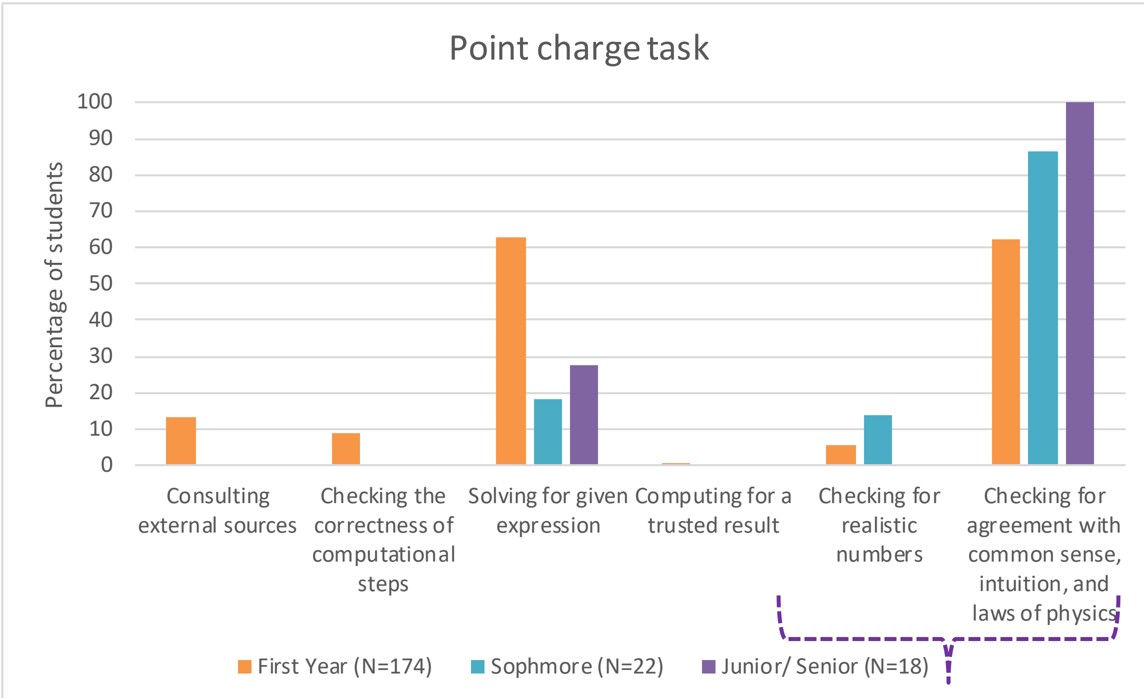


Figure 5.3: Prevalence of evaluation strategies across the curriculum in the Inclined plane task



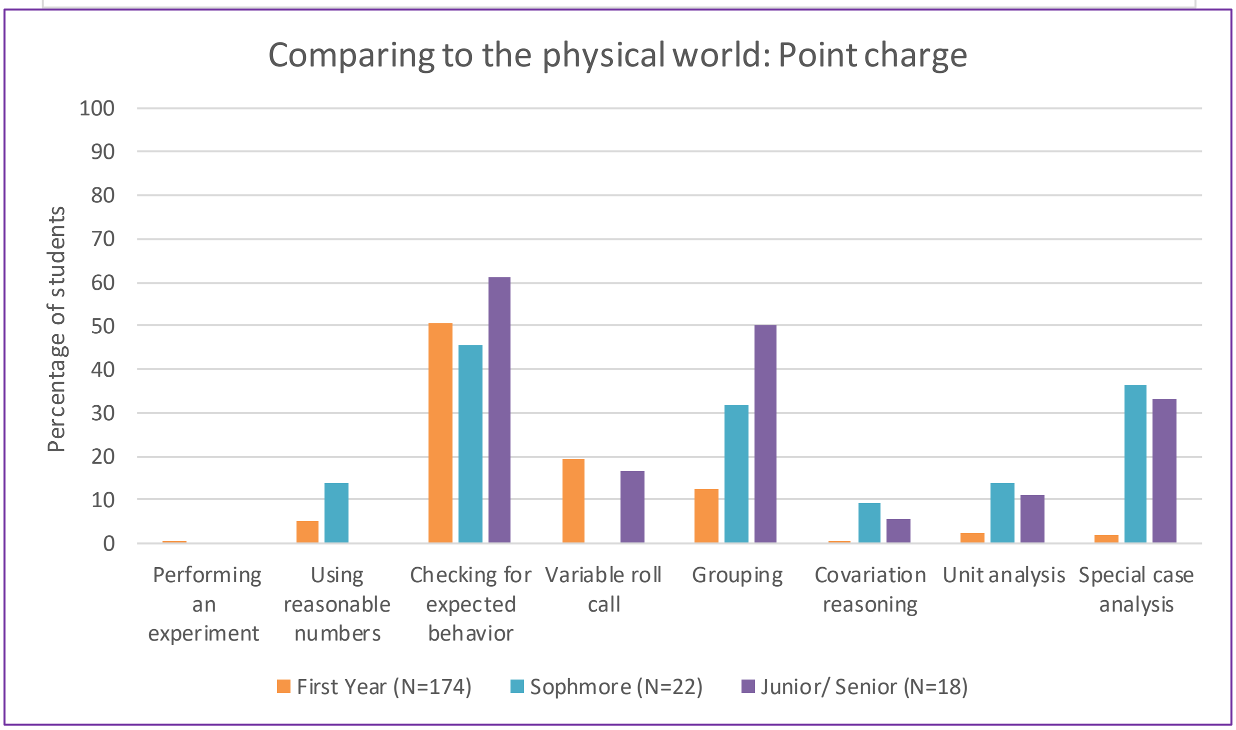
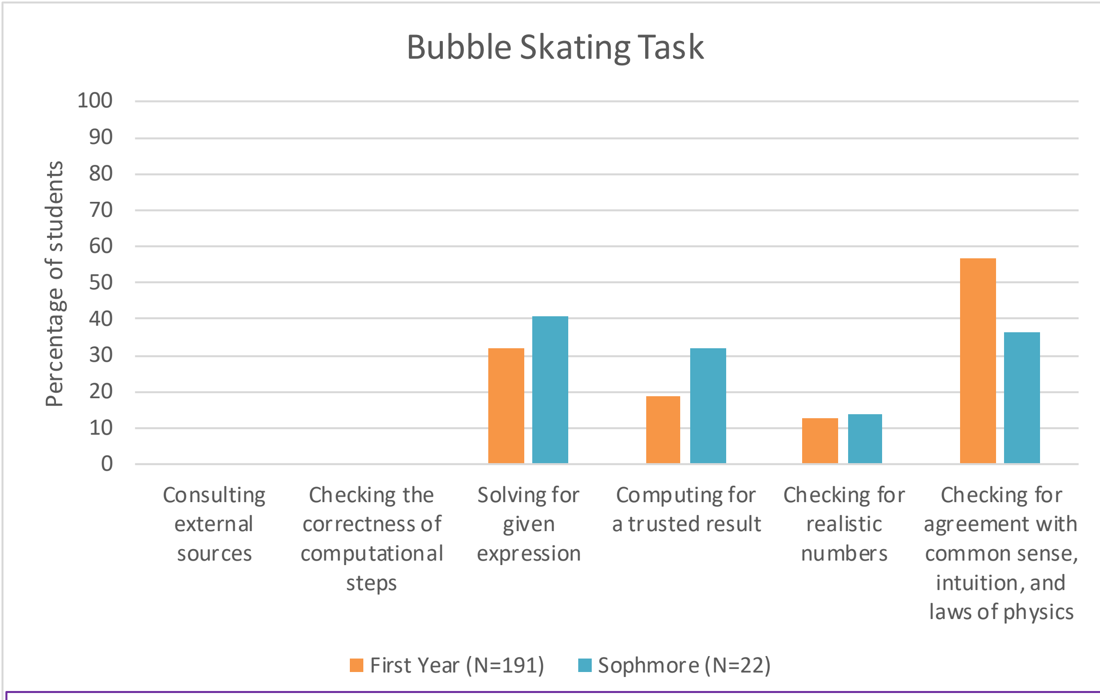


Figure 5.4: Prevalence of evaluation strategies across the curriculum in the Point charge task



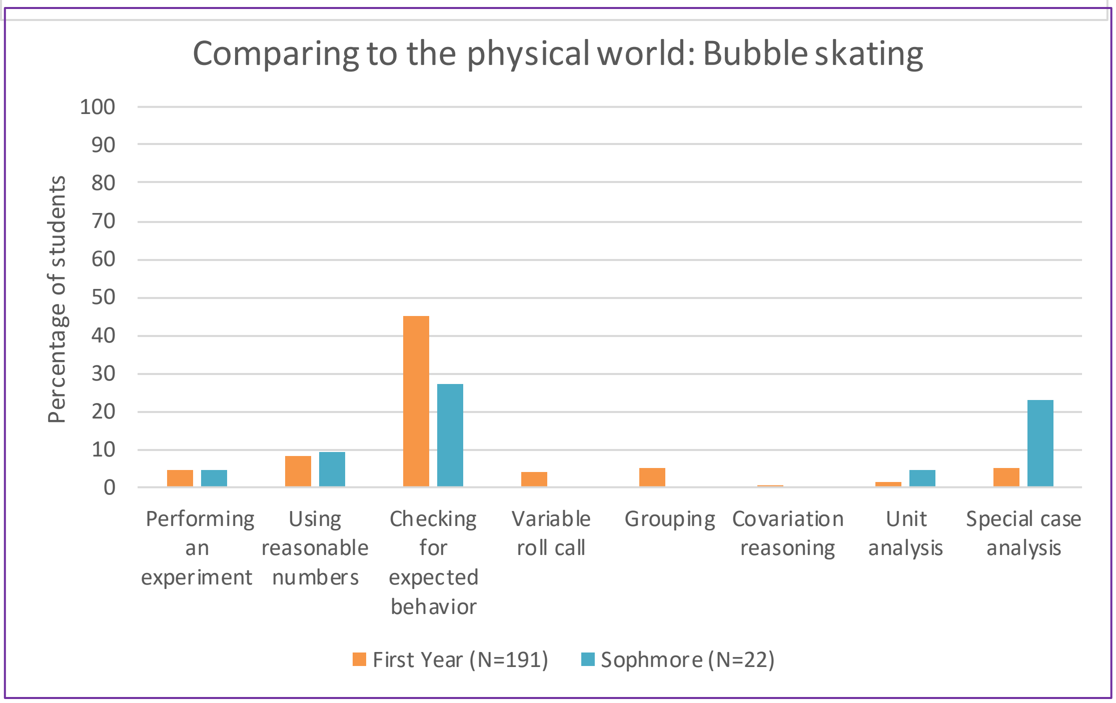


Figure 5.5: Prevalence of evaluation strategies across the curriculum in the Bubble skating task

There are a few observations that we can make about our quantitative results. First, we observe a shift from the use of evaluation strategies that rely on computation in our first-year student population, to the use of evaluation strategies that verify the ability of the equation to describe the given physical scenario in our junior student population. This is an expected trend, but we have documented it, and shown some specific expressions of the underlying changes that cause this trend. Most notably, on all three tasks, the use of sophisticated strategies such as special case analysis and unit analysis increases considerably among the sophomore and junior student populations. In the inclined plane and point charge task, the percentage of student population that use strategies in the *checking through computation* category is smaller at the sophomore and junior levels than at the introductory level, while the percentage of students that use strategies in the *comparing to the physical world* category is greater at the sophomore and junior/senior levels than at the introductory level.

Secondly, our data suggest that students at the sophomore and junior/senior level are more selective in their choice of task-specific strategies. Strategies that are not productive in a given task context are largely not present in the responses these student populations. For instance, in the bubble skating task, *grouping* and *quantity roll call* are not present beyond the first-year level. This may reflect that more experienced students recognize that *grouping* and *quantity roll call* are not productive in the bubble skating context. On the flip side, *grouping can be a productive strategy* in the point charge task context and indeed it is more prevalent in the sophomore and junior student population.

Lastly, we observed that some strategies used by students in our introductory population are not used by our sophomore and junior/senior student population. Specifically, the strategies of checking for correctness of computation steps and consulting external sources are not used by any students in our sophomore and junior/senior student population.

The quantitative results appear to show the trend of a shift from strategies that focus on computational correctness at the first-year level, to strategies that focus on the fit between the given expression and the physical scenario it describes at the upper level of the curriculum. Howeve r, the low number of surveyed students at the upper level of the curriculum warrants the question of validity. How real is the difference between students' performance real considering the low N at the upper level? To address this concern, we calculated the standard error of the mean of and performed the chi-squared test of our measurements.

To perform this statistical test, we split every population that we surveyed into two groups: students that used at least one sophisticated evaluation strategy, and students that used only unsophisticated strategies. For this purpose, we considered the following strategies as sophisticated: *special case analysis*, *unit analysis, using reasonable numbers*, *covariational reasoning*, and *grouping*. All other strategies were considered unsophisticated strategies.

While this split obscures some of the subtle features of the data set, it created a binary response pattern for use of at least one sophisticated strategy, which allowed the use of simple statistical tools. This allows us to consider the question: were the rates of use of sophisticated strategies in first-year and higher-level courses likely to arise simply by chance? Treating the responses as categorical and binary enables the use of chi-squared tests for statistical significance. Figure 5.6 shows the percentage of students at each level using at least one sophisticated strategy, for each task; error bars show the standard error of the mean. Table 5.2 shows the chi-squared analysis for the data.

The chi-squared tests show statistically significant differences at the $p<.05$ level between the first years and all higher-level populations on all tasks, suggesting that more post-introductory students use at least one sophisticated evaluation strategy than first-year students in all tasks. While a promising result, we note that this is a rudimentary statistical analysis; more detailed analysis in the future will yield more robust results.

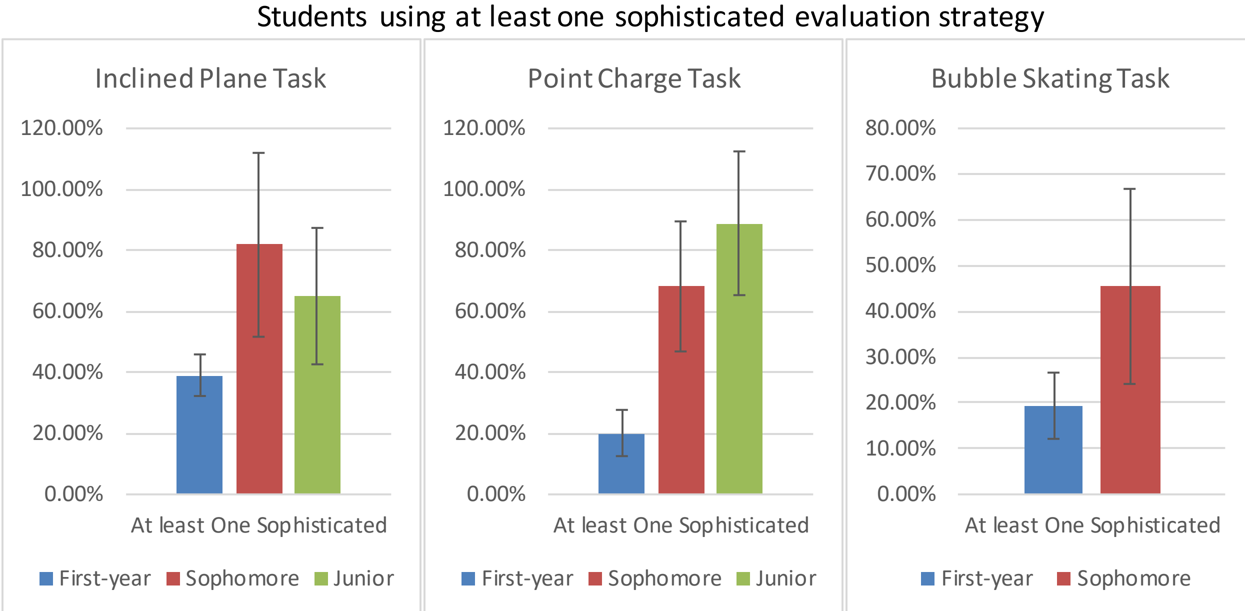


Figure 5.6: Percentage of students at each level using at least one sophisticated strategy in each task context. Error bars show the standard error of the mean

Table 5.2: Summary of p- values comparing first-year and post introductory use of at least one sophisticated strategy in each task context

Task	Inclined Plane	Point Charge	Bubble Skating
Level	First-year		
Sophomore	.00495	<0.00001	.00521
Junior	.0242	<0.00001	N/A

5.3.2 **Case studies**

For our case studies, we chose to examine two representative evaluation strategies used by our students during pair interviews. In both interviews, the students were asked to evaluate an expression for the final velocities of two skaters involved in a one-dimensional elastic collision (bubble skating task, Figure 5.1). The introductory students, Frodo and Sam, performed *arithmetic substitution:* they assigned numbers to variables in the given expression and used the result to check that numerically, a physics rule or concept holds for the given expression. In this case, numeric substitution was used to verify that the given expression was consistent with conservation of momentum. During the interview, the students said they would do this (arithmetic substitution) if they had numbers, and the interviewer suggested that the students could make up numbers. The interviewer did not provide numbers but facilitated the students' choice to choose their own numbers. On the other hand, our juniors, Jack and Jill, performed *special case analysis:* they checked whether the expression describes the real world or laws of physics as expected under certain physical or corresponding mathematical conditions. We outline and narrate these two evaluation strategies in the excerpt below. First, we present excerpts from the first-year interview, then we present excerpts from the junior interview. To make referencing easier, we number segments for the excerpts.

5.3.2.1 **First year interview excerpts**

To evaluate the given expression, the first-year students used *arithmetic substitution.* They assigned numbers to the masses and initial velocities of the skaters. The students used these numbers and the given equation to calculate the final velocities of the skaters. Finally, they used the masses and final and initial velocities of the skaters to solve for their initial and final momentum. The values of the calculated momentum were then used to verify that the given expression is consistent with conservation of momentum. This process is outlined and narrated in the excerpt below.

1. Sam: So your, momentum has to remain law of conservation of energy says that [trails off] I don't know.
2. Frodo: Oh, I didn't think it was conservation of momentum and conservation of energy. We already used those so that would not make any sense[…].
3. Sam: We could say that if this [$mv_{1i} + mv_{2i} = mv_{1f} + mv_{2f}$] is the case, then if these two [skaters] have different masses and then the next part we had different velocities, then

this […] this $[mv_{1i} + mv_{2i} = mv_{1f} + mv_{2f}]$ part would have to hold true wouldn't it if we put in the actual masses and velocities of both of them.

4.Frodo: I yeah, I agree with that. But I don't know. I just think that we have to think of something else because we derived that from this. So it obviously going to be true, I think. I don't' know, I could be wrong.

5.Sam: […] I could probably just like make up like this is so many kilograms, this is so many kilograms […] and then I could just go in, I could just put in two different speeds for the two velocities, and then see[…] the momentum of those two [skaters] are equal to each other afterwards, […]. But at the same time, I mean if the numbers come up correctly then […] it should equal to the same equation, which kind of says that it's true…

6.Sam: Let's try this, let's say m_1 is equal to 2kg and then m_2 is equal to 5kg and then v_1 is equal to 1m/s and then v_2 is equal to 2m/s. […] so v final has to be equal to $m_1 - m_2$,so two minus five kilograms over and $m_1 + m_2$ so two kilograms plus five kilograms times v_{1i} which is 1m/s[….].

Sam [incorrectly] solves for $v_{1f} = \frac{2kg-5kg}{2kg+5kg} * 1\,{}^{m}/_{s} + \frac{2*5kg}{2kg+5kg} * 2\,{}^{m}/_{s} = -\frac{13}{7}$.

Frodo first solves for the initial momentum: $2kg * 1\,{}^{m}/_{s} + 5kg * 2\,{}^{m}/_{s} = 12kg\,{}^{m}/_{s}$. Then Frodo solves for $v_{2f} = \frac{2*2kg}{2kg+5kg} * 1\,{}^{m}/_{s} + \frac{5kg-2kg}{2kg+5kg} 2\,{}^{m}/_{s} = -\frac{10}{7} kg\,{}^{m}/_{s}$. Lastly, he uses his value for v_{2f} and Sam's[wrong] value for v_{1f} to solve for $p_f = \frac{76}{7} kg\,{}^{m}/_{s}$ which is not equal to the calculated initial momentum. Once they both realize this inconsistency, they both go through their calculations.

7.Frodo: That's not right. Maybe I did something wrong. Maybe that's [inaudible] go back to this [written work to calculate p_i]

8.Sam: [Going through his calculations for v_{1f}]. So two minus five, that's fine. That's fine. Two plus seven. That's fine. And then that's seven. That's seven. Two, two plus 10 times times two times two because v_2 . twenty. That’s three, Twenty minus three. ... Oh wait, it's not 13 it's 17 that's maybe, okay.

9.Frodo: So where oh seventeen so that changes to 34 which is four and six sevenths. That's still wrong is it?

10.Sam: No cause it's, that's one plus seven is eight plus four...

11.Frodo: Oh I was adding that to that my bad. That that's what I supposed to add it to. Whoops. Alright.

12.Sam: So that's, numerically it makes sense.

5.3.2.2 **Junior students interview excerpts**

To evaluate the give expression, the juniors used *special case analysis.* They checked if the given expression was consistent with the laws of physics, and their intuitions and real life experiences under certain physical or corresponding mathematical conditions. Specifically, the students checked the case where the masses of the skaters were equal, and where the mass of one skater was much larger than the other and the initial velocity of one skater was zero. For each of these cases, the students calculated the final velocities of the skaters using the given equations. The students then compared their result with laws of physics, and real-life experiences. This process is outlined and narrated in the excerpt below.

1.Jack: Yeah but that would, yeah that would be just checking another boundary. Like let – Let $m_1 = 10m_2$ or something. Yeah like something much larger or a $100m_2$.

2.Jill: Yeah I think that if we picked another case it would be just more of the same.

3.Jack: [...] If that's a 100 m_2 and that $\left[\frac{m_1-m_2}{m_1+m_2}\right]$ would just be $99m_2$ over $101m_2$ that's roughly 1. This $\left[\frac{2m_2}{m_1+m_2}\right]$ would be basically zero. [...]

4.Jill: And I guess the idea behind your idea of doing limiting cases would be where one of the velocities is zero. And one of them can be really large. [Already drew an arrow and dot to represent p_{1i} and p_{2i} respectively]. So, like this arrow [draws arrow to represent p_{1f}] could be super big, so that is consistent.

5.Jack: [Working out the algebra on the board] Roughly get v_{1f}=v_{1i} so if I just had a really heavy first object relative to my second object situation again and I like had a certain speed and then hit it I wouldn't like you wouldn't stop me that much. Right. Like I would just keep going. Yeah

[The students continue examining the special case of masses ($m_1 = 100m_2$) they (incorrectly) arrive at the result $v_{2f} \approx 2v_{1f} + v_{2i}$ and are uncomfortable with it.]

6.Jack: And if I do that on a second expression [Jill pointing to v_{2f} expression] on this one equals we get $200m_2$. Jesus is that that's right. It's right. m_1 m_2. Is that's right? Yeah. No I'm not crazy. Two times 100 is 200 right? [...]. $\frac{99m_2}{101m_2}v_{2i}$. Does that make sense?

7.Jill: I don't know. Now I am worried. [Jack: I know!] because I was trying to think about it in terms of my vector addition thing.

8.Jack: and equals $2v_{1i} + v_{2i}$? Let me think about that for a second [Looks over his work on the board].

8.Jill: [Looking over what Jack wrote on the board] wait. I think you're missing...

9.Jack: Did I miss something?

10.Jill: This [The $99m_2$ in $\frac{99m_2}{101m_2}v_{2i}$] should be minus. $m_2 - 100m_2$ so that's

11.Jack: Oh oh oh. Oh. You caught me. You caught me. I'm not sure. You are. You are right. So, it's you. Oh yeah, that's a vector addition works.

[Now, they have the result: $v_{2f} \approx 2v_{1f} - v_{2i}$]

12.Jack: Yeah. So let's pick a case like setting v_{2i} to be zero. I feel like I am confusing myself now.

13.Jill: [...] something's odd. I'm confused because when you apply the same, when you plug in this [$v_{2f} \approx 2v_{1f}$] condition to both, then you're suggesting that the initial velocity. Or the initial velocity of the first one is equal to the final velocity of the second one of the first one which means that the final velocity of two should be zero for that to work out.

14. Jack: But wouldn't it have to be moving because if we draw but wouldn't it just wouldn't be moving like if I if I like ran and hit you with a sphere and I didn't like stop then you would just be moving along in the same velocity as me?

15. Jack: So. the final velocity of the second object is somehow twice the initial velocity. Like how. Where did that come from?

16. Jill: Wait I'm sorry. OK. So, you said you're trying to make

17. Jack: if I'm coming if I'm coming let's put this into numbers [make v_{2i} zero.] So, let's have let's have this coming at me at some like wait.

18. Jack: Let's just do velocity vectors I guess so $v_{1i} = 100\,{}^{miles}/_{hour}$ let's be crazy. I love being crazy umm meters per second what am I, a monster? meters per second.

And so and then we have $v_{2i} = 0$. This [$v_{1f} \approx v_{1i}$] says here that

v_{1f} still equals $100\,{}^{m}/_{s}$. But then this one says that somehow this [$v_{2f} \approx 2v_{1f}$] would be $v_{2f} = 200\,{}^{m}/_{s} - v_{2i}$. Yeah, that's minus zero. So how is that a thing?

In our qualitative analysis of our case studies, we made some observations. To discuss our observations, we will make a claim about a difference or similarity about how the first years and juniors evaluated the given expression and then support it with excerpts from both interviews. Our observations from the qualitative data are summarized in table 5.1.

First, we observed that to find consistency between the given expression and the laws of physics, the first-year students sought to generate an arithmetic equality, while the juniors looked for consistency between the given expression, experience, and physics concepts under certain conditions. Both the first-years and juniors evaluated the given expression by checking to see whether it obeyed the law of conservation of momentum. The first-year students went about this by evaluating the expression numerically (fig. 5.7). Specifically, in segments 3 and 5, Sam suggested that if actual masses and velocities were substituted into the given equations, then the calculated final velocities should be consistent with the statement of conservation of momentum $[mv_{1i} + mv_{2i} = mv_{1f} + mv_{2f}]$. In segment 6, Sam and Frodo plugged in numerical values for the masses and initial velocities of the skaters, then used these numbers to solve for the initial momentum and final velocities of the skaters. Finally, they used their chosen values for masses and initial velocities and calculated values of the final velocities of the skaters. In segment 12, the students arrive at the result that the initial and final momenta of the skaters are equal $(p_i = p_f = 12kgm/s)$, indicating that the expression is consistent with conservation of momentum and therefore correct. They do not reason with ratios or variables, and indeed spend a significant amount of time and effort on the arithmetic.

Table 5.3: Summary of the similarities and differences between first years and juniors in the case studies

How do the students…	First years	Juniors
Find consistency	Generate arithmetic equality	Find consistency between equation, experience and physics concepts under certain conditions
View result	As an end in itself	Something from which to extract physical meaning
Engage variables	Use arithmetic equality (e.g., m_1=2kg, m_2=5kg	Primarily use symbolic equality (e.g., m_1=100m_2) but also use arithmetic equality.
Perform operations	Perform mathematical operations meticulously	Perform mathematical operations less strictly *and* employ qualitative comparisons
Troubleshoot	Check calculation procedures	Check calculation procedures and also try to reconcile results with experience and physics knowledge.

On the other hand, the juniors evaluated the expression by checking to see if it satisfied the law of conservation of momentum under certain conditions. Specifically, here, the students performed two special case analyses: one for when the masses are equal, and one for when one mass is much larger than the other. We focus on the second case where one mass is much larger than the other. In segment 1, the students decide to evaluate the case where one mass is one hundred times larger than the other ($m_1 = 100m_2$). In segments 3 through 5, they arrive at, make sense of, and agree with, the result $v_{1f} = v_{1i}$. They conclude that the given expression is consistent with the laws of conservation of energy.

Secondly, we noticed that while evaluating the given expression, both the first-years and juniors used equations as a computational means to verify the conservation of momentum. However, juniors also extracted physical meaning out of the result of their computation, while the first-years seemed to see the result as an end in itself. In segment 6 through 12, while verifying that the given expression was consistent with momentum conservation, the first years performed mathematical operations to confirm that momentum is conserved. However, the only sense they make of their result is that $p_i = p_f = 12kgm/s$, so momentum is conserved. They do not triangulate this result with anything else or extract more physical meaning from it. They also do not make physical sense of any intermediate results or mathematical operations.

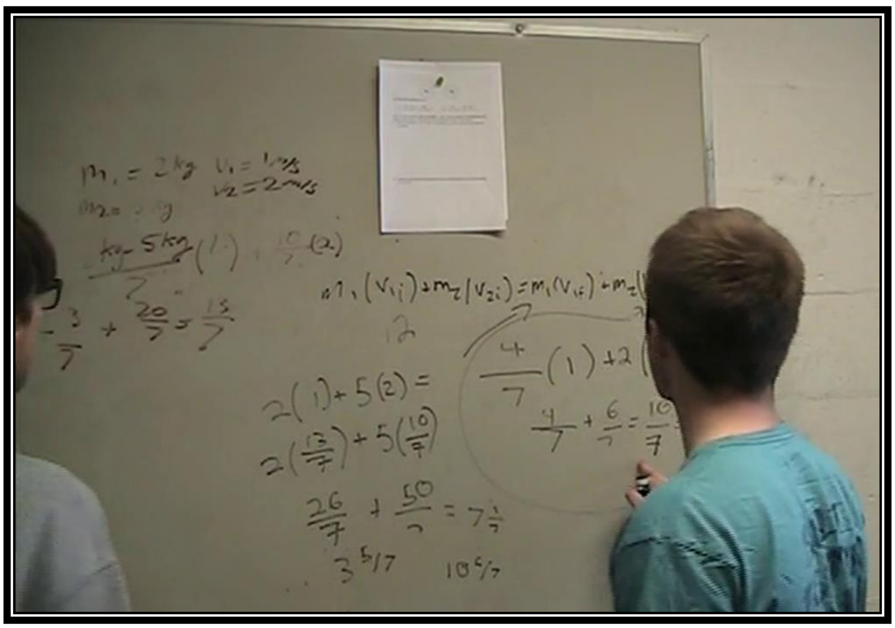


Figure 5.7: Sam and Frodo using Arithmetic Substitution

On the other hand, while verifying that the given expression was consistent with momentum conservation, the juniors made sense of their results by corroborating it with other representations of conservation of energy and lived experience. For instance, in segments 4 and 13, Jill crosschecks the results $v_{1f} = v_{1i}$ and $v_{2f} \approx 2v_{1f}$ by representing the velocities as vectors and confirming the vector diagrams are consistent with conservation of momentum. In segments 5 and 15, Jack makes sense of the results by comparing the physical meaning of the mathematical result to a real-life scenario of running into Jill. In the rest of the interview, Jack compares the $m_1 \gg m_2$ case to the lecture demonstration known as Newton's cradle, and the $m_1 \gg m_2$ to a giant boulder rolling towards him like the scene in the movie *Raiders of the Lost Ark*. Not only did the juniors perform calculations, but they also corroborated their results with other representations of their system and extracted physical meaning from their mathematical results.

Thirdly, we also observed that while evaluating the given expression/ finding consistency between the given expression and the laws of physics, both first year students and juniors used arithmetic

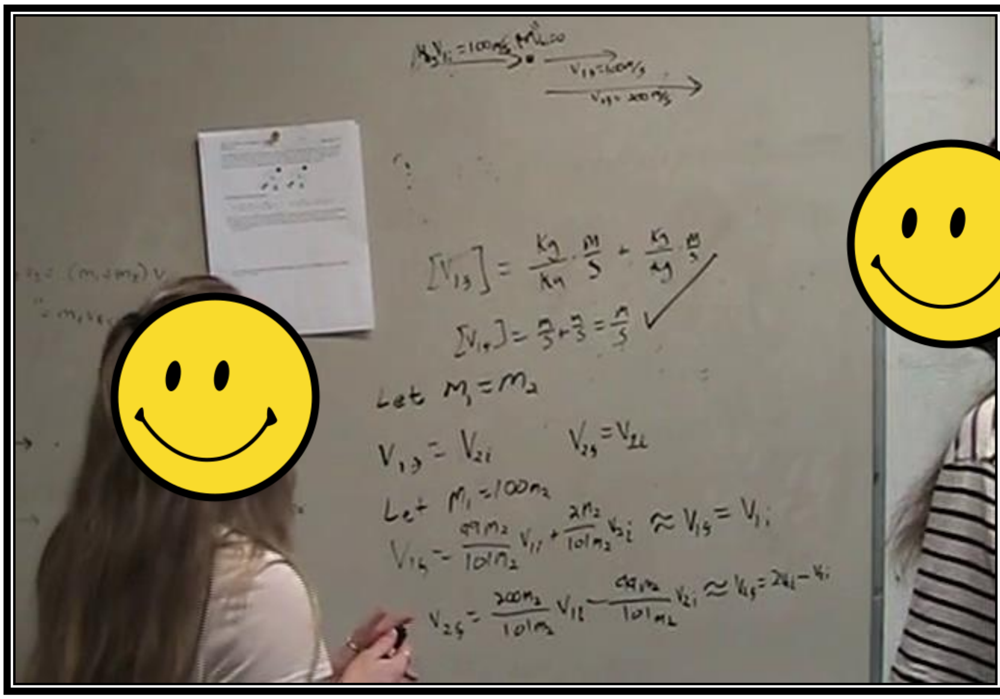


Figure 5.8: Jack and Jill using Special Case Analysis

equalities. However, the juniors primarily used symbolic equalities and ratios. While verifying that given expression was consistent with the law of conservation of momentum, the first-year students checked for numerical equalities, using specific values for the masses and initial velocities of the skaters. In segment 3 through 12 of the excerpt, Sam and Frodo use the values $m_1 = 2kg,\ m_2 = 5kg, v_{1i} = 1m/s,$ and $v_{2i} = 2m/s$ to solve for final velocities and ultimately verify that the given expressions are consistent with the law of conservation of momentum.

On the other hand, Jack and Jill do not use specific values for the masses of the skaters in their special case analysis (fig. 5.8), using symbolic ratios instead. In segment 1 and 3, Jack uses the ratio $m_1 = 100m_2$ to describe the case where the mass is one skater is arbitrary larger than the other. Earlier (before the excerpt segment), the pair use the ratio $m_1 = m_2$ to describe the special case of when both masses are equal.

While examining the limit of one mass being much greater than the other, an expert would most likely use the representation and symbolic equality $m_1 \gg m_2$. Jack's and Jill's choice of $m_1 = 100m_2$ is an intermediate version between two numbers and a specific mass proportion; this provides insight about the behavior between expert and novice evaluation and may suggest an avenue for instructors to help promote symbol-based evaluation.

However, similar to the first-year students, the juniors also used numerical values for variables while verifying that given expression was consistent with the law of conservation of momentum. In segment 16, Jack plugged numbers into the result $v_{2f} \approx 2v_{1f}$. However, unlike the first-years, Jack did not use this number to check for an arithmetic equality or as an end in itself. Instead, he extracted physical meaning from the result, using it to elaborate how unusual the result $v_{2f} \approx 2v_{1f}$ was.

Fourthly, we noted that while evaluating the given expression and finding consistency between the given expression and the laws of physics, the first-year students performed mathematical operations punctiliously while the juniors employed more qualitative comparisons. In lines 6 - 12, Sam and Frodo compute specific numerical values for the initial momentum, final momentum, and final velocities of the skaters. While solving for these quantities, no qualitative comparisons were made, and no computational result or operations are approximated.

On the other hand, while verifying that given expression was consistent with the law of conservation of momentum, the juniors were not strict in their computation for expressions. In segment 3, Jack approximates $\frac{99m_2}{101m_2}$ as "roughly 1" and $\frac{2m_2}{101m_2}$ as "basically zero". At the beginning of segment 6 and 13, Jack describes their result for v_{2f} as $v_{2f} \approx 2v_{1f} + v_{2i}$ and $v_{2f} \approx 2v_{1f}$ respectively. In this way, the juniors are less strict/rigid about their calculations as they evaluate.

Finally, we also observed that while evaluating the given expression, both first-years and juniors debugged their work by checking their computational procedures. However, in addition, the juniors also tried to reconcile their results with intuition, experience, and knowledge of physics.

While evaluating the given expression, (line 7), Sam and Frodo realized that their calculated values for p_i and p_f did not match. To address this problem, they looked over their calculations again meticulously, making sure that correct numbers have been entered and mathematical operations had been carried out correctly (segment 7-11). In segment 8, Sam realized that he had incorrectly calculated v_{1f}, and Frodo had used this result to calculate p_f. In segment 9 through 11, the pair re-did their calculations for p_f, and arrive at the correct results.

Similarly, while evaluating the given expression, the juniors arrived at the result $v_{2f} \approx 2v_{1f} + v_{2i}$ and were uncomfortable with it (segments 6-14). Like the first-year students, Jack and Jill cautiously looked through their calculations (segment 6, and 8-11). For instance, in segment 6, Jack wondered whether two times 100 is 200, and in segment 9, he asked Jill if he missed something in his calculations. Upon inspection of their work, Jill found that Jack dropped a negative sign (segment 10-11). However, in segment 7, prior to this check of procedures, Jill had attempted to reconcile their result with her vector diagram.

Furthermore, (in segment 12) after arriving at the result $v_{2f} \approx 2v_{1f} - v_{2i}$, the pair first simplified their results by making $v_{2i} = 0$, leaving them with the result $v_{2f} \approx 2v_{1f}$. Jill thought this result was "odd" and explained how it was inconsistent with conservation of momentum the vector re presentation of the final velocities of the skaters. To counter this, (in segment 14), Jack described a real-life scenario that matches the result: one where he is running into Jill while carrying a heavy sphere causing Jill to move at his speed.

In the rest of the interview, Jack and Jill went back and forth to make sense of the result $v_{2f} \approx 2v_{1f}$: Jill talked about the vector diagram, and Jack used real life examples including bowling while throwing a heavy ball at a much lighter ball, and a scene from the movie *Raiders of the Lost Ark* where someone is being chased by a rolling boulder. These comparisons bring up a discussion about the differences between elastic and inelastic collisions, and both students discuss which type of collisions apply to Jack's scenarios. In the end, Jill realized that her vector diagram had represented velocity and did not take into account the mass of the skaters. After this error is resolved, the two agreed that the result $v_{2f} \approx 2v_{1f}$ was reasonable.

5.4 Discussion and conclusion

To further delve into the observations from our quantitative data and case studies, we examine our results from existing perspectives in PER. We analyze our results from the lens of the epistemic complexity of equations [20]. We also compare our results to the results of expert-novice studies in problem solving. Observations from our case studies are used to interpret our quantitative data and consequent claims. We also provide some implications for instruction and outline directions for future work.

First, from the perspective of Bing and Redish's epistemic complexity of equations [20], both first-years and juniors use equations as a calculation scheme. Both groups of students use the given equation to calculate results. The first years use the given equation and conservation of momentum to calculate the numerical values for the masses, initial and final velocities of the skaters. The Juniors use the given equations to calculate the final velocities of the skaters when their masses are equal and when one mass is much bigger than the other.

Secondly, the juniors seem to consider the given equations as a physical relation among measurements – one that describes the final velocities of two bodies involved in a collision as a function of their initial velocities and the ratio of, and differences in, their masses. Unlike the first-year students, juniors extracted physical information from the results of their calculations. The juniors also corroborated the equation with real-life experiences and with other representations of conservation of momentum.

The results of our study are consistent with prior work in PER that describes the similarities between expert and novice behaviors in problem solving using epistemic games and frames. For instance, the high prevalence of the use of evaluation strategies that focus on calculation observed in our first-year population is consistent with a physics preference for playing the recursive plug and chug game, solving problems in a rote equation-chasing or calculation frame [20], [21], [29]. This result is also consistent with novices solving physics problems using a plug and chug approach [28]. Our observations are consistent with the finding that inexperienced physics students tend to focus on meticulously calculating and finding the "right result". However, here, we get to see these behaviors play out in a different context (evaluation) as opposed to those of prior studies in which students solved problems.

Similarly, our results are consistent with the findings that during problem solving, experienced physics students connect their calculations to physical meaning. The use of evaluation strategies in the comparing to the physical world strategy in our sophomore and junior-senior student population is consistent with experts' preference for solving physics problems using a scientific approach [28], or in the quantitative sensemaking and physical mapping frames [20], [29]. Our observations are consistent with the finding that experienced physics students and physicists corroborate and find coherence between their calculations, knowledge of the laws of physics, and intuition/real life experience.

Our results also suggest that expert and novice uses of evaluation do not constitute a dichotomy but rather form a spectrum. Our quantitative data showed that, generally, all our surveyed population used the same evaluation strategies, but some strategies were more prevalent in one group (level) than the other. There are no new strategies that the upper-level students used that the first-year students did not. The percentage of students who used sophisticated strategies was greater in the sophomore and junior/senior student population than in the first-year student population. Similarly, in our case studies, our qualitative results also suggest that while the juniors used a more sophisticated evaluation strategy than the first-year students did, the more advanced physics students and first-year students performed some similar actions while evaluating the solution to the physics problem, e.g., substituting numbers into the equation, and meticulously checking calculation steps while troubleshooting.

Our interview/qualitative data suggest that there is more to the shift in the prevalence of evaluation strategies observed in our quantitative data than meets the eye. For instance, for the first-year students, evaluating an equation involved dealing with numbers, and a final decision was made by checking whether one number equals another ($p_i = p_f$), whereas ironically (from the perspective of the first-year students), the juniors are less committed to such strict equalities. The juniors seem to be more fluid in their thinking, but to a novice it may seem sloppy and imprecise, e.g., in using approximations.

Furthermore, the first-year students' use of numbers at the beginning of the evaluation sequence is consequential because it constrains the rest of the evaluation process. Specifically, the substitution of numbers for variables makes it hard to recover the symbolic relationship between the physical quantities in the given equation. On the hand, the approach of the juniors is a much more flexible way of thinking about mathematical relationships between the quantities in the equation, and how the equation describes the physical world. The juniors' use of symbolic relationships and ratios is consistent with a dynamic view of mathematics, such that one equation describes a range of physical phenomenon. This is very different from what the first-year students do.

Another subtle difference between first-years and more advanced students, as illustrated by our case studies, is the interpretation and triangulation of results. While evaluating the given expression, the only physical meaning the first-year students extracted from their result was its consistency with conservation of momentum. Furthermore, they did not seek to triangulate the result of their calculations with anything outside their computation, e.g., other physics concepts, other representation of their results, and real-life scenarios. The only thing that made Sam and Frodo pause was getting a value for the initial momentum that was not equal to the calculated final momentum.

On the other hand, while evaluating the given expression, the juniors thought about the physical meaning of their results and used sensemaking tools that were outside their computation. Specifically, they compared the physical meaning of their results with vector representation of conservation of momentum, and a few real-life scenarios that reflected their chosen special cases. Like the first-year students, the juniors looked through their arithmetic while deliberating a result with which they were disgruntled. However, in addition to looking over their work, they also thought about the physical meaning of their result. Unlike the first-year students, Jack and Jill have developed approximation skills and intuition about what they expect in real life situations. This observation is consistent with the finding that triangulation of, and fluid movement between, different sensemaking approaches characterize expert-like behavior in physics [21], [70].

Finally, our study provides some insight for potential approaches for teaching evaluation strategies. Jack and Jill's suggestion of using $10m_1$ or $100m_1$, and later $1000m_1$, may suggest a possible bridge between a novice approach to evaluation and an expert approach toward evaluation. As students at the introductory level seem more comfortable with numbers, perhaps guiding them through a special case analysis with $m_1 = 100kg$ and $m_1 = 1kg$ would be a productive activity. If we look closely at what the first-year students are trying to do, they are going about their goal in the right way. However, as instructors, we want to nudge their approach into a direction that is more productive.

While the shift in the student's use of physics and mathematics is interesting to see, it is not surprising as we hope that this is the development students go through during their training as physics majors. Some first-year students do use strategies from the *compare to the physical world* group of strategies. This brings up the question of whether we train students to use comparing to the physical world strategies, or instead select for students that are already inclined to do so? This question is beyond the scope of this paper but, may be an avenue for future work. It is not clear what the causes in skill level are, but future work might probe the impact of specific classes between first year and junior/senior level by collecting research data at beginning of the semester and end of the semester.

In conclusion, the goal of our study was to document and describe the differences between expert and novice use of evaluation strategies. We did this using analysis of written responses and interview data at different levels of the physics curriculum. From written responses, we demonstrated differences in the use of different evaluation strategies at the introductory, sophomore, and junior levels. We then examined two case studies, one each from the first year and junior level, to gain insight into the results from our quantitative data.

From our quantitative data, we found that while all the surveyed student populations drew from the same set of evaluation strategies, the percentage of students who used sophisticated evaluation strategies was higher in the sophomore and junior/senior student populations than in the first-year population. Our simple statistical analysis suggests that more post-introductory students use at least one sophisticated evaluation strategy than first-year students in all tasks. From our case study, we found that that while evaluating an expression, both juniors and first years performed similar actions.

However, while the first-year students focused on computation and checked for arithmetic consistency with the laws of physics, juniors checked for computational correctness *and* probed whether the equation accurately described the physical world and obeyed the laws of physics.

Our case study suggests that a key difference between expert and novice evaluation is that experts extract physical meaning from their result and make sense of them by comparing them to other representations of laws of physics, and real-life experience. These results are consistent with previous descriptions of expert and novice behaviors in physics problem solving. Future work on this project can show whether and when students are taught to evaluate as they get into advanced physics courses. Future work could also specify what courses equip student with the ability to evaluate solutions.

CHAPTER 6

6. CONCLUSIONS

The aim of this project was to study students' use and understanding of evaluation strategies. This project explored evaluation strategies as an avenue for students to find connections between mathematical operations, physics concepts, intuition, and real-life experience. This project is also one of a group of studies at the confluence of mathematics and physics. Evaluation strategies entail integrating mathematics and physics in a way that makes sense physically and checking that physics is consistent with itself. Consequently, one goal of our project was to examine how students ground their use of mathematics and mathematical reasoning in physics in the context of using evaluation strategies. The project also studied how students' use of evaluation strategies evolves as students gain expertise on physics.

We sought to meet these goals by answering the following research questions.

1. To what extent, and in what ways, do students evaluate the validity of derived expressions or solutions when prompted?
2. How do students' use of evaluation strategies fit current models from PER and adjoining fields?
3. How do students' use of evaluation strategies compare at different levels the physics curriculum?

In Chapter 3 we addressed the first and second research questions. We found that students used a myriad of evaluation strategies when prompted, including special case analysis, grouping, performing an experiment, solving for a known result, and consulting external sources. We classified students' responses into 3 categories: consulting external sources, checking with computation, and comparing to the physical world. However, at the introductory level, only a few students used canonical evaluation strategies (*special case analysis*, *unit analysis*, and *use of reasonable numbers*). We also found that instead of evaluating, most introductory students attempted or suggested solving for the given solution using first principles. We also compared our classifications of evaluation strategies to proof/justifications in mathematics education research, epistemic frames in PER, and control beliefs about knowledge in metacognition. The analysis of strategies in the comparing to the physical world category from the perspective of the use of mathematics in physics showed that evaluation strategies are indeed a great avenue for students to ground their use of mathematics in physics and find coherence between calculation, physics concepts and intuition/real life experience.

In Chapter 4, we addressed the first and second research questions. We focused on *grouping*, one of the novel evaluation strategies identified in student responses. We established that grouping is a sophisticated strategy using perspectives of mathematical modeling and mathematical reasoning in physics. We then showed that grouping is consistent with the phenomenon of chunking in cognitive psychology. We also demonstrated how grouping also had some attributes of symbolic forms and the use of symbolic forms. We also showed that on the inclined plane and point charge task, the percentage of students who used grouping was greater. However, on the bubble skating task, the trend is reversed.

In Chapter 5, we addressed the first and third research questions. We focused on the evolution of the use of evaluation strategies as students move through the physics curriculum. We showed that on the written task, the percentage of students coded as using canonical evaluation strategies and other comparing to the physical world strategies increased in sophomore and junior/senior student populations surveyed. On the other hand, the percentage of students coded as using strategies in the *checking through calculation* and *consulting external source*s category decreased in sophomore and junior/senior student populations surveyed. We also used two case studies to show the key similarities and differences between the way first years and juniors/seniors evaluate solutions to physics problems.

We found that to ascertain the validity of an equation, some of our first-year students checked for computational consistency, while juniors/seniors checked for computational correctness *and* also probed the ability of the equation to accurately describe the physical world / obey the laws of physics.

In conclusion, we found that students use both canonical and non-canonical evaluation strategies. At the introductory level, many students solve for the given expression from first principles instead of evaluating. The evaluation strategies that we observed our students using are consistent with models in PER, mathematics education research, and cognitive science. Specifically, we found our classifications of evaluation strategies consistent with Bing and Redish's epistemic frames, Sherin's symbolic forms, and models of mathematical reasoning in physics [10]–[12], [20], [39]. Our findings were also consistent with prior research on proofs and justifications in mathematics education research. Lastly, from the perspective of metacognition, our results were also consistent with control and beliefs about knowledge. Furthermore, the strategy of grouping is consistent with the phenomenon of chunking in cognitive science [69]. These perspectives show that evaluation strategies in the *comparing to the physical world* category are consistent with the definition of evaluation, and help students to find connections between mathematical operations, physics concepts, intuition, and real-life experience. As students move up the physics curriculum, they become better evaluators. For instance, between the first and second year of the physics curriculum, there is considerable increase in the percentage of students that use sophisticated evaluation strategies. Our simple statistical analysis suggests that more post-introductory students use at least one sophisticated evaluation strategy than first-year students in all tasks.

There are a few limitations of our results. First, from the perspective of design, providing the students with symbolic solutions might have pushed them toward more qualitative and symbolic reasoning. However, despite this possible cuing, our results at the first-year level show that very few students used strategies canonically taught for evaluation. Furthermore, it could also be argued that despite the attempt to let students choose their own strategy, that the tasks might well have pushed students to the modes of evaluation favored by researchers. As a result, although there were several strategies that had not been previously documented, perhaps other strategies would have been seen in tasks that were more open ended. An expansion of our work in the future might include tasks with a mix of both symbolic and arithmetic solutions. Another limitation of our design is that it is hard to directly compare the results of our task because they were not strictly isomorphic. While the multiple contexts of our task allow us to see how the use of evaluation strategies vary with context, it does not allow for direct across- task comparison of the prevalence of the use of observed evaluation strategies.

Aside from self-regulation, none of the frameworks we adopted were developed or adopted to describe evaluating or evaluation strategies. Nonetheless, these frameworks were useful for describing, giving language for, and explaining the patterns we observed in students' responses. Evaluation is a step in the larger processes of problem solving and mathematical modeling. Consequently, the frameworks of mathematical modeling, epistemic games, and epistemic frames are useful perspectives to describe and analyze our results. Furthermore, as physics education researchers, we are not only concerned about what students do when they solve problems, but also how they are thinking about the problem-solving activity. As a result, the framework of metacognition and chunking are helpful perspectives to explain patterns in students' responses.

Specifically, the framework of mathematical modeling describes how to model systems using mathematical representations. However, this framework was useful for describing our task, including the steps taken to get to the provided solution, and the steps we expected our students to perform on the equation we provided. Similarly, Sowder and Harel's proofs and justifications describe how students justify that their mathematical proofs are right. However, the proof schemes Sowder and Harel observed their students use were useful for describing our categories of the evaluation strategies our students used. Finally, Bing and Redish's epistemic frames describe how students justify their problem-solving approach. However, the epistemic frames they observed in their study are consistent with our categories of evaluation strategies.

While different, these frameworks sometimes complement each other. For instance, as we discussed, both *metacognition* and *epistemic frames* can describe the students' attitude/perception of problems/problem solving. However, Schoenfeld's work on metacognition is discussed in the contexts of problem solving in mathematics. Consequently, Bing and Redish's discussion epistemic frames complement Schoenfeld's discussion of students' attitude towards problem solving because it is done in the context of physics. Similarly, while models of mathematical modeling are useful in describing how to model any system using mathematical representations, models of using mathematics in physics complement models of mathematical modeling as they give insight into how to use mathematics to model physical systems. Lastly, chunking describes how stimuli (in this case, symbols in an equation) are perceived in groups, while Sherin's symbolic forms provides specific descriptions of how chunking can be expressed in a physics problem solving/evaluation context.

Nonetheless, the similarity between the frameworks brings up the issue of the lack of communication between different fields doing similar work. For instance, Bing and Redish's epistemic frames and Schoenfeld's modes are quite similar. However, Bing and Redish do not cite or refer to Schoenfeld's work. Similarly, Sowder and Harel's types of student's justifications, and Bing and Redish's epistemic frames are similar. However, the authors do not reference each other. The lack of cross reference between Bing and Redish's epistemic games and Sowder and Harel's students' justification is a good example of the lack of communication between the fields of physics and mathematics education research.

Physics as a subject area is mathematics-intensive, and as a result, it is not unusual for studies in PER and mathematics education research to be parallel or share similar results. Furthermore, the field of mathematics education research is older than that of physics education research. Consequently, physics education research can be informed by already existing frameworks in mathematics education research. However, it is important to note that this type of adaption of frameworks from mathematics education research can prove difficult for many reasons including one as trivial as differences in nomenclature. For instance, during this project, our first attempt to find relevant frameworks in mathematics education literature proved unsuccessful as a search using the keyword "evaluation" did not yield any results. The suggestion to use proofs and justifications was from feedback after giving a talk at the RUME conference. Nonetheless, the consistency of our results with frameworks in mathematics education research suggest the need for collaboration and streamlining of research in the fields of mathematics and physics education research.

The consistency between our results and all the frameworks we chose help solidify our results. Even though the frameworks we are adopted are not specific to evaluating and evaluation strategies, the consistency between the frameworks used, and the consistency between the frameworks and our results help strengthen our conclusions. The consistency of our results with prior research allows us to be confident in our results.

The wide range of frameworks that describe and explain students' use of evaluation strategies echo the importance of evaluation strategies including but not limited to their relationship with self-regulation, beliefs about knowledge, critical thinking and problem-solving performance. However, despite the far-reaching importance of evaluation, and its' importance to problem solving in physics and mathematics, there is not a lot of research on evaluation in both fields. The closest thing to evaluation in mathematics education is students use of proofs and justifications in mathematics. This gap tells us that evaluation is a field that needs to be further studied.

The use of different frameworks in our study is aided by the extensiveness of the importance of problem solving. Evaluation is an important step in mathematics and physics problems solving. However, problem solving is common to fields outside physics and mathematics. For instance, many fields of Discipline Based Education Research (DBER) study problem solving in the context of their respective discipline. However, because problem solving is an activity that cuts across many fields, there are similarities between problem solving frameworks in DBER. Another reason why we can use frameworks from other DBER fields is that every DBER field studies the same thing: what students think learning is, how students think, the beliefs, mindsets and perspectives students bring into the classroom, and how students' attitudes interact with their learning.

In conclusion, to analyze and discuss the results of our project, we got insights from many perspectives including mathematical modeling, metacognition, using mathematics in physics, chunking, epistemic frames and symbolic forms. Some of these insights would be impossible, were it not for a careful review of literature outside of what many PER scholars generally attend to.

One take away from this project is that physics education researchers can find and productively utilize frameworks in other fields. Particularly, as a field, we can glean some wisdom from older fields like mathematics education research and cognitive science. In our experience, we found that these fields had good input for patterns that we observed in students' work. For instance, we found that Carlson's discussion of mental actions to be very useful in capturing students' descriptions of the presence of, and covariational relationships between symbols in the provided equations. Mental actions framework should be widely used by researchers who study the use of mathematics in physics, but it is not. Our project is enhanced by its interdisciplinary focus: from cognitive science, and the RUME literature. This is something there should be more of, and the cross-pollination of our communities is important to the growth of PER and other DBER fields.

There are few other take home points from this project. First, while most students at the introductory level do not use canonical evaluation strategies, many use other potentially useful strategies like covariational reasoning, grouping, and quantity roll call. Physics instructors should look out for the use of these non-canonical strategies and encourage students when they employ them. Furthermore, our results suggest that students can evaluate solutions when they are explicitly taught to do so. Our results also suggest that evaluation strategies are not generalizable to every task, i.e., the use of some evolution strategies is task dependent. Finally, the results of study of the evolution of strategy use suggests that teaching introductory students how to evaluate using arithmetic versions of canonical evaluation strategies might provide some scaffolding for teaching/learning how to evaluate.

While we have provided at least partial answers to each of our research questions, our results prompt additional questions and suggest further studies and projects for future scholars. In the future, we hope to expand and build on our results. First, we hope to delve further into the evolution of students use of evaluation strategies. Specifically, we want to examine the responses of introductory physics majors explicitly to determine whether our students learned how to evaluate as part of the physics curriculum, or whether the physics major selects for students who already possess more sophisticated evaluation skills. We also plan to use the results of this project to develop instructional materials to guide students to evaluate. In the near future, we plan to expand upon our study by asking students to evaluate solutions to physics problems in unexplored contexts such as quantum physics, relativity, and thermodynamics. Furthermore, we hope to examine the interaction between physics and mathematics during evaluation by giving our evaluation tasks to students who have taken mathematics courses but have not taken any physics classes.

7. BIBLIOGRAPHY

[1] L. Greenfield, "Engineering Student Probem Solving," in *Cognitive process instruction, research on teaching thinking skills*, 1979, pp. 233–254.

[2] M. E. Loverude, "Quantitative reasoning skills in math methods," *PERC Procceedings*, pp. 203–206, 2015, doi: 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.*, vol. 59, no. 10, pp. 891–897, Oct. 1991, doi: 10.1119/1.16667.

[4] D. S. Wright and C. D. Williams, "A wise strategy for introductory physics," *Phys. Teach.*, vol. 24, no. 4, pp. 211–216, 1986, doi: 10.1119/1.2341986.

[5] G. Polya, *How To Solve It*. 1945.

[6] A. H. Schoenfeld, *Mathematical Problem Solving*. New York: Academic Press, 1985.

[7] P. Heller, R. Keith, and S. Anderson, "Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving," *Am. J. Phys.*, vol. 60, no. 7, pp. 627–636, 1992, doi: 10.1119/1.17117.

[8] W. Blum and D. Leiß, "How do students' and teachers deal with modelling problems?," in *Mathematical Modelling: Engineering and Econonomics.*, C. et. al . (Eds. Haines, Ed. Chichester: Horwood, 2007, pp. 222–231.

[9] J. Perrenet and B. Zwaneveld, "The many faces of the mathematical modeling cycle," *J. Math. Model. Appl.*, vol. 1, no. 6, pp. 3–21, 2012.

[10] E. F. Redish and E. Kuo, "Language of Physics, Language of Math: Disciplinary Culture and Dynamic Epistemology," *Sci. Educ.*, vol. 24, no. 5–6, pp. 561–590, Jul. 2015, doi: 10.1007/s11191-015-9749-7.

[11] 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. Spec. Top. - Phys. Educ. Res.*, vol. 9, no. 2, p. 020119, Nov. 2013, doi: 10.1103/PhysRevSTPER.9.020119.

[12] O. Uhden, R. Karam, M. Pietrocola, and G. Pospiech, "Modelling Mathematical Reasoning in Physics Education," *Sci. Educ.*, vol. 21, no. 4, pp. 485–506, 2012, doi: 10.1007/s11191-011-9396-6.

[13] L. S. Vygotsky, *Mind in society: The development of higher psychological processes (A. R. Luria, M. Lopez-Morillas & M. Cole [with J. V. Wertsch], Trans.)*. 1978.

[14] A. H. Schoenfeld, Ed., *Cognitive Science and Mathematics Education*. Routledge, 1987.

[15] A. H. Schoenfeld, "Learning To Think Mathematically," pp. 334–370, 1992.

[16] A. R. Warren, "Impact of teaching students to use evaluation strategies," *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 6, no. 2, pp. 1–12, 2010, doi: 10.1103/PhysRevSTPER.6.020103.

[17] M. Lenz, P. J. Emigh, and E. Gire, "Surprise! students don't do special-case analysis when unaware of it," *PERC Procceedings*, 2019, doi: 10.1119/perc.2018.pr.lenz.

[18] T.-R. Sikorski, G. D. White, and J. Landay, "Uptake of solution checks by undergraduate physics students," in *2017 Physics Education Research Conference Proceedings*, Jan. 2018, pp. 368–371, doi: 10.1119/perc.2017.pr.087.

[19] E. Burkholder, L. Blackmon, and C. Wieman, “Characterizing the mathematical problem-solving strategies of transitioning novice physics students,” *Phys. Rev. Phys. Educ. Res.*, vol. 16, no. 2, p. 20134, 2020, doi: 10.1103/PhysRevPhysEducRes.16.020134.

[20] T. J. Bing and E. F. Redish, “Analyzing problem solving using math in physics: Epistemological framing via warrants,” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 5, no. 2, pp. 1–15, 2009, doi: 10.1103/physrevstper.5.020108.

[21] T. J. Bing and E. F. Redish, “Epistemic complexity and the journeyman-expert transition,” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 8, no. 1, pp. 1–11, 2012, doi: 10.1103/PhysRevSTPER.8.010105.

[22] National Research Council, “Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics Education,” National Academies Press, Washington, D.C., Apr. 2011. doi: 10.17226/13099.

[23] M. Scriven and R. Paul, “A statement by Michael Scriven & Richard Paul, presented at the 8th Annual International Conference on Critical Thinking and Education Reform, Summer 1987.” 8th Annual International Conference on Critical Thinking and Education Reform, Summer 1987.

[24] J. I. Heller and F. Reif, “Prescribing Effective Human Problem-Solving Processes : Problem Description in Physics Linked references are available on JSTOR for this article : Prescribing Effective Human Problem- Solving Processes : Problem Description in Physics,” *Cogn. Instr.*, vol. 1, no. 2, pp. 177–216, 1984.

[25] D. Huffman, “Effect of Explicit Problem Solving Instruction on High School Students’ Problem-Solving Performance and Conceptual Understanding of Physics,” *J. Res. Sci. Teach.*, vol. 34, no. 6, pp. 551–570, 1997, doi: 10.1002/(SICI)1098-2736(199708)34:6<551::AID-TEA2>3.0.CO;2-M.

[26] J. L. Docktor *et al.*, “Assessing student written problem solutions: A problem-solving rubric with application to introductory physics,” *Phys. Rev. Phys. Educ. Res.*, vol. 12, no. 1, pp. 1–18, 2016, doi: 10.1103/PhysRevPhysEducRes.12.010130.

[27] J. L. Docktor and J. P. Mestre, “Synthesis of discipline-based education research in physics,” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 10, no. 2, pp. 1–58, 2014, doi: 10.1103/PhysRevSTPER.10.020119.

[28] L. N. Walsh, R. G. Howard, and B. Bowe, “Phenomenographic study of students’ problem solving approaches in physics,” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 3, no. 2, pp. 1–12, 2007, doi: 10.1103/PhysRevSTPER.3.020108.

[29] J. Tuminaro and E. F. Redish, “Elements of a cognitive model of physics problem solving: Epistemic games,” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 3, no. 2, pp. 1–22, 2007, doi: 10.1103/PhysRevSTPER.3.020101.

[30] G. Kaiser, “Realitätsbezüge im Mathematikunterricht – Ein Überblick über die aktuelle und historische Diskussion,” in *Materialen für einen realitätsbezogenen Mathematikunterricht*, et al. (Eds. G. Graumann, Ed. Bad Salzdetfurth: Franzbecker, 1995, pp. 64–84.

[31] W. Blum, “Anwendungsbezüge im Mathematik-unterricht – Trends und Perspektiven,” in *Trends und Perspektiven. Schriftenreihe Didaktik der Mathematik, Vol. 23.*, G. et al. (Eds. . Kadunz, Ed. Wein: Hölder-Pichler- Tempsky, 1996, pp. 15–38.

[32] 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.*, vol. 33, no. 5, pp. 352–378, 2002, doi: 10.2307/4149958.

[33] M. Carlson, "A cross-sectional investigation of the development of the function concept," no. October, pp. 114–162, 1998, doi: 10.1090/cbmath/007/04.

[34] J. A. Czocher, "How can emphasizing mathematical modeling principles benefit students in a traditionally taught differential equations course?," *J. Math. Behav.*, vol. 45, pp. 78–94, 2017, doi: 10.1016/j.jmathb.2016.10.006.

[35] E. Goffman, *Frame Analysis:An Essay on the Organization of Experience*. Chicago: Harper & Row, 1974.

[36] J. Tuminaro, "A Cognitive Framework for Analyzing and," University of Maryland, 2004.

[37] R. R. Bajracharya and J. R. Thompson, "Analytical derivation: An epistemic game for solving mathematically based physics problems," *Phys. Rev. Phys. Educ. Res.*, vol. 12, no. 1, pp. 1–21, 2016, doi: 10.1103/PhysRevPhysEducRes.12.010124.

[38] A. Collins and W. Ferguson, "CollinsFerguson1993EpistemicGames.pdf," *Educational Psychologist28*, vol. 28, no. 1. pp. 25–42, 1993.

[39] B. L. Sherin, "How Students Understand Physics Equations," *Cogn. Instr.*, vol. 19, no. 4, pp. 479–541, 2001.

[40] B. P. Schermerhorn and J. R. Thompson, "Physics students' construction and checking of differential volume elements in an unconventional spherical coordinate system," *Phys. Rev. Phys. Educ. Res.*, vol. 15, no. 1, p. 10112, 2019, doi: 10.1103/PhysRevPhysEducRes.15.010112.

[41] B. L. Sherin, "The Symbolic Basis of Physical Intuition: A Study of Two Symbol Systems in Physics Instruction," University of California Berkely, 1996.

[42] G. Harel and L. Sowder, "Types of Students' Justifications," *Math. Teach.*, vol. 91, no. 8, pp. 670–675, 1998.

[43] 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.*, vol. 23, no. 4, pp. 614–628, 1986, doi: 10.3102/00028312023004614.

[44] N. E. Perry, "Young Children's Self-Regulated Learning and Contexts That Support It," *J. Educ. Psychol.*, vol. 90, no. 4, pp. 715–729, 1998, doi: 10.1037/0022-0663.90.4.715.

[45] A. de Groot, *Thought and Choice in Chess (first Dutch edition in 1946).* The Hauge: Mouton Publishers, 1965.

[46] W. G. Chase and H. A. Simon, "THE MIND'S EYE IN CHESS," in *Visual Information Processing*, W. G. Chase, Ed. Elsevier, 1973, pp. 215–281.

[47] W. G. Chase and H. A. Simon, "Perception in chess," *Cogn. Psychol.*, vol. 4, no. 1, pp. 55–81, 1973, doi: 10.1016/0010-0285(73)90004-2.

[48] F. Gobet *et al.*, "Chunking mechanisms in human learning," *Trends Cogn. Sci.*, vol. 5, no. 6, pp. 236–243, 2001, doi: 10.1016/S1364-6613(00)01662-4.

[49] G. Knoblich, S. Ohlsson, H. Haider, and D. Rhenius, "Constraint Relaxation and Chunk Decomposition in Insight Problem Solving," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 25, no. 6, pp. 1534–1555, 1999, doi: 10.1037/0278-7393.25.6.1534.

[50] A. L. Gilchrist and N. Cowan, “What Is Chunking?,” pp. 476–483, 2021.

[51] A. Newell and H. A. Simon, *Human problem solving*. Englewood Cliffs: Prentice-Hall, 1972.

[52] F. Gobet and H. A. Simon, “Templates in chess memory: A mechanism for recalling several boards,” *Cogn. Psychol.*, vol. 31, no. 1, pp. 1–40, 1996, doi: 10.1006/cogp.1996.0011.

[53] F. Gobet and H. A. Simon, “Five seconds or sixty? Presentation time in expert memory,” *Cogn. Sci.*, vol. 24, no. 4, pp. 651–682, 2000, doi: 10.1207/s15516709cog2404_4.

[54] F. Gobet, *Understanding Expertise: A Multi-Disciplinary Approach*. New York, NY: Palgrave Macmillan Education, 2016.

[55] K. A. Ericsson and A. C. Lehmann, “Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints,” *Annu. Rev. Psychol.*, vol. 47, no. August 2015, pp. 273–305, 1996, doi: 10.1146/annurev.psych.47.1.273.

[56] D. H. Holding, *The Psychology of chess skill*. Hilldsale, NJ: L.Erlbaum, 1985.

[57] F. Gobet, “Expert memory: A comparison of four theories,” *Cognition*, vol. 66, no. 2, pp. 115–152, 1998, doi: 10.1016/S0010-0277(98)00020-1.

[58] W. G. Chase and H. A. Simon, “Perception in chess,” *Cogn. Psychol.*, vol. 4, no. 1, pp. 55–81, Jan. 1973, doi: 10.1016/0010-0285(73)90004-2.

[59] D. E. Egan and B. J. Schwartz, “Chunking in recall of symbolic drawings,” *Mem. Cognit.*, vol. 7, no. 2, pp. 149–158, 1979, doi: 10.3758/BF03197595.

[60] S. Ohlsson, “The mechanism of restructuring in geometry,” in *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 1990, pp. 237–244.

[61] 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.*, vol. 64, no. 12, pp. 1495–1503, 1996, doi: 10.1119/1.18409.

[62] A. F. Heckler, *The Role of Automatic, Bottom-Up Processes: In the Ubiquitous Patterns of Incorrect Answers to Science Questions*, vol. 55. 2011.

[63] R. Knight, *Physics for Scientists and Engineers: A Strategic Approach with Modern Physics*, 4th ed. Pearson, 2016.

[64] J. A. Maxwell, *Qualitative Research Design. An Interactive Approach*. Thousand Oaks, California: Sage Publications, Inc, 2005.

[65] M. C. Wittmann and K. E. Black, “Mathematical actions as procedural resources: An example from the separation of variables,” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 11, no. 2, pp. 1–13, 2015, doi: 10.1103/PhysRevSTPER.11.020114.

[66] M. T. H. Chi, P. J. Feltovich, and R. Glaser, “Categorization and representation of physics problems by experts and novices,” *Cogn. Sci.*, vol. 5, no. 2, pp. 121–152, 1981, doi: 10.1207/s15516709cog0502_2.

[67] M. D. George *et al.*, “Shaping the Future: New Expectations for Undergraduate Education in Science, Mathematics, Engineering, and Technology,” *J. Coll. Sci. Teach.*, pp. 1–95, 1997.

[68] A. R. Akinyemi, J. R. Thompson, and M. E. Loverude, “Summary of evaluation strategys (in preparation),” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, 2021.

[69] F. Gobet, M. Lloyd-Kelly, and P. C. R. Lane, “What’s in a Name? The Multiple Meanings of ‘Chunk’ and ‘Chunking,’” *Front. Psychol.*, vol. 7, no. February, pp. 1–5, 2016, doi: 10.3389/fpsyg.2016.00102.

[70] E. Kuo, M. M. Hull, A. Elby, and A. Gupta, “Assessing mathematical sensemaking in physics through calculation-concept crossover,” *Phys. Rev. Phys. Educ. Res.*, vol. 16, no. 2, p. 20109, 2020, doi: 10.1103/PhysRevPhysEducRes.16.020109.

8. APPENDIX

8.1 Code tables

Table 8.1: Summary of codes in the inclined plane task

Category	Inclined plane examples
Consulting external sources	N/A
Checking correctness of computational steps	*"I would check to see if the values of each variable were put into the equation correctly"*
Quoting equation	*"I would check if my solution is reasonable by first determining the velocity of the block using a different verified equation"*
Derivative/integral	*"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* $\int \left(2gd\,(sin\,\theta\, - \mu cos\,\theta\,)\right)^{\frac{1}{2}}\,dv$.
Solving for given expression	"I would sum all of the forces in the x and y directions and derive an equation that way. $x{:}\, f - \sin w = ma,\ \ y{:}\, N - \cos w = ma$."
Solve from a different perspective	"solve *using another method like finding* $\hat{\imath}$ *and* $\hat{\jmath}$*. If both methods yield the same answer & the numbers are reasonable, the answer is most likely correct."*
Legitimate derivative/integral	*"Use forces in the x-direction, find acceleration.* $f - mgsin\,\theta\, = ma$*. Use acceleration and take the antiderivative to find velocity."*

Table 8.1 Continued

Solving for a known	*"plugging in the answer I got for v, and then using that to solve for d or another variable."*
Algebraic substitution	"I would plug the equation into another equation that needs the velocity or I could rearrange another equation to equal velocity and see if I can get the two to equal... $\hat{v}_f = \hat{v}_i + \hat{a}t$. $\sqrt{2gd\,(sin\,\theta\, - \mu cos\,\theta\,)} = 0 + \hat{a}t$. $\sqrt{2gd\,(sin\,\theta\, - \mu cos\,\theta\,)} = \frac{f}{m}t$. $F_{total\,y} = W\cos\theta - f$, . $F_{total\,x} = \mu - W\sin\theta$. $\sqrt{2gd\,(sin\,\theta\, - \mu cos\,\theta\,)} = \frac{(W\cos\theta - f)\hat{\imath} + (\mu - W\sin\theta)\hat{\jmath}}{m}t$."
Arithmetic substitution	"Assuming I came up with a numerical answer, I could check it by looking at the energy at the top (potential energy) and energy at the bottom (Kinetic energy). If my calculations show that PE=Kei (Or $E_{initial}$=E_{final}) then I'd assume (I could be wrong) I had the correct answer"
Performing an experiment	*"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"*
Using reasonable numbers	"you should be able to tell if v is too slow or too fast for example, if d =10m, a v of 100,000m/s wouldn't make sense"
Special case analysis	If I were to put the block at the bottom of the ramp (makes d=0) then I would have $v = \sqrt{2(9.8)0\,(sin\,\theta\, - \mu cos\,\theta\,)}$ meaning $v = \sqrt{0}$,which seems okay"
Unit analysis	"I would check to see if the units were reasonable, as velocity is *m/s* and in this case, it is $\sqrt{\frac{m^2}{s^2}} = m/s$."

Table 8.1 continued

Covariational reasoning	“the velocity increases with the θ increasing which makes sense”
Grouping	“we need a *d(*m) so the *d* makes sense. Need an initial force so (g). $\sin\theta$ reasonable because hypotenuse is known, *h* isn’t. $\mu\cos\theta$ got to factor in friction…no idea why there is a *2* or why there is a sqrt function.”
Quantity roll call	“This is not likely the correct choice. Mass is not included in the equation”
Checking for expected behavior	“Make sure the solution is negative since the block is sliding in the –y direction.”
Plug in	Plugging in the numbers for each variable and see what the result is
0's and 1's	"Solve for a variable when v=0, plug that into the equation and solve for remanding [remaining] variable"

Table 8.2: Summary of codes in the point charge task

Category	**Point charge examples**
Consulting external source	"email the professor or a recitation TA to see of the answer is correct, go check at the physics learning center if the result is correct."
Checking correctness of computational steps	"By making sure the correct formula is used, the values are plugged into the correct places, and that the calculations are correct."
Quoting equation	"Use the equation for an electric field for each of the three +q charges to find their vector components, then sum them to determine the net electric field at p."
Derivative/integral	"I would take the antiderivative and see if it matches the expression."
Solving for given expression	*"you know that $\underline{E}$ at any point is $= \sum \frac{q_i}{4\pi\varepsilon_0} . \frac{1}{r^2}\hat{r}$ so reference that equation[...] Yes, since the sum would be* $\frac{1}{4\pi\varepsilon_0} . \frac{q}{x^2} \frac{1}{4\pi\varepsilon_0} . \frac{q}{(x^2+d^2)} + \frac{1}{4\pi\varepsilon_0} . \frac{q}{(x^2+d^2)}$ ".
Legitimate derivative/integral	"You could treat the three +q points as one big bar with charge +q spread across it, then integrate from –d to +d with respect to y. $dE_x = \frac{kdq}{r^2}\cos\theta, dQ = \frac{\theta}{d}dy$"
Arithmetic substitution	N/A
Algebraic substitution	$\bar{E}_{net} = q\bar{F}\hat{\imath} \quad F = \frac{k\|q_1 q_2\|}{r^2}$ If we were given the force due to the points on p and the charge of P. Then I would use $F = q.E$ to see if my friend's answer is valid.
Solving for a known	N/A

Table 8.2 Continued

Performing an experiment	"Test it…"
Using reasonable numbers	"I might pick values and "test run" the equation to make sure it has a reasonable answer"
Special case analysis	Check that the result is somewhat less than $E_{net} = \frac{3q}{4\pi\varepsilon_0 x^2}$ but more than $E_{net} = \frac{2q}{4\pi\varepsilon_0 x^2}$"
Unit analysis	"Check units to see if answer in N/C"
Covariational reasoning	"I would plug in random equal values for d and a possible value of p… as the x values increase, the E_{net} should decrease."
Grouping	" I will look @ the equation $E = \frac{q_1 q_1}{r^2}$ and compare what is written to the equation. q is the charge, k is $\frac{1}{4\pi\epsilon_0}$, x is the radius @ zero ,$\sqrt{x^2 + d^2}$ is the radius when +q is @d or -d. I think the result is correct because there is a k &q. The it is separated by the 2 different radi"
Quantity roll call	"First, I would ask how and why π is involved… "why is there a π in this?" -no response"
Checking for expected behavior	"Observe whether they added a y component because we know that the vectors would cancel each other out. No $\hat{\jmath}$ or y component."
Plug in	"I would use the equation and plug the numbers in to check the answer"
0's and 1's	"Derive the y axis from the x axis"

Table 8.3: Summary of codes in the Bubble Skating Task

Category	**Bubble Skating example**
Consulting external source	N/A
Checking correctness of computational steps	N/A
Quoting equation	"Using physics, I'd see if there's an equation that better resembles what I'm thinking ..."
Derivative/integral	N/A
Solving for given expression	Finally, an example in the bubble skating task, is *"...Furthermore, elastic collisions are conserved in terms of energy, ...*$\frac{1}{2}m_1{v_i}^2 + \frac{1}{2}m_2{v_i}^2 = \frac{1}{2}m_2{v_f}^2 + \frac{1}{2}m_1{v_1}^2$*[...]* $v_f = \sqrt{\frac{m_1v_i^2 + m_2v_i^2 - m_1v_i^2}{m_2}}$ *.No as it doesn't contain a square root, so by my method it's incorrect"*
Solve from a different perspective	N.A
Legitimate derivative/integral	N.A

Table 8.3 Continued

Solving for a known	"…I would plug it back in and solve for a known variable"
Algebraic substitution	*"check to see if kinetic energy is conserved.* $\frac{1}{2}m_1v_{1i}^2 + \frac{1}{2}m_2v_{2i}^2 = \frac{1}{2}m_1v_{1f}^2 + \frac{1}{2}m_2v_{2f}^2$. $m_1v_{1i}^2 + m_2v_{2i}^2 = m_1v_{1f}^2 + m_2v_{2f}^2$. $m_1v_{1i}^2 + m_2v_{2i}^2 = m_1\left[\frac{m_1-m_2}{m_1+m_2}v_{1i} + \frac{2m_2}{m_1+m_2}v_{2i}\right]^2 + m_2\left[\frac{2m_1}{m_1+m_2}v_{1i} + \frac{m_2-m_1}{m_1+m_2}v_{2i}\right]^2$"
Numeric substitution	*To check if your answer is reasonable, you can check to see if momentum is conserved with the velocities you calculated. You calculated the velocities and know the masses.* $m_1 = 20kg, m_2 = 25kg, v_{1i} = 5\frac{m}{s}, v_{2i} = -5\frac{m}{s}$. $v_{1i} = \left(\frac{20-25}{20+25}kg\right)\left(-5\frac{m}{s}\right) + \left(\frac{50}{20+25}\right)\left(-5\frac{m}{s}\right)$ *[…]* $P_i = 100 + 125$. $v_{2f} = \left(\frac{40}{45}\right)(5) + \left(\frac{5}{45}\right)(-5)$. $v_{2f} = \left(\frac{175}{45}\right)\frac{m}{s}$. $p_i = -25kgm$ $p_f = -25kgm$. *Yes this will yield the correct result."*
Performing an experiment	"Actually do the experiment…"
Using reasonable numbers	"Compare to how fast humans can run…"
Special case analysis	"You can make vi be equal and make m1 be extremely large to see if its momentum is barely changed."

Table 8.3 Continued

Unit analysis	"The idea is to verify by checking if both sides have the same units [m/s] and [m/s]"
Covariational reasoning	"$p_i = p_f$ we know if $m_1 > m_2$ that $p_1 = p_2$ and vice versa..."
Grouping	"Since it is elastic the masses will stay together so half of the results seems odd"
Quantity roll call	"this is likely the correct result because both masses are taken into account before and after collision..."
Checking for expected behavior	"Since its a collision, the answers should be equal but opposite"
Check Momentum/Energy conservation	"You could check your answer by checking to see if momentum and kinetic energy is conserved because elastic collisions conserve both energy and momentum"
Plug in	"I would check it by plugging in values into because momentum is conserved"
0's and 1's	"Setting the two equations equal to each other and solving to see if it is inelastic"

BIOGRAPHY OF THE AUTHOR

Abolaji Akinyemi was born in Kaduna, Nigeria on October 25, 1992. She is $23 + 5i$ years old. She was raised in Abuja, Nigeria and graduated from Federal Government Girls College Bwari, Abuja in 2008. She attended the Minnesota State University and graduated in 2015 with Bachelor's degrees in Physics and Astronomy. She joined the Department of Physics (Physics Education Research Group) at The University of Maine in the fall of 2015. After receiving her degree, Abolaji will be joining the University of Minnesota, Duluth, to begin her career in teaching physics and making delicious curries. Abolaji is a candidate for the Doctorate degree in physics from the University of Maine in August 2021.