



Measuring and Assessing Introductory Students' Physics Problem-Solving Ability

Citation

Miller, Olivia Colleen. 2023. Measuring and Assessing Introductory Students' Physics Problem-Solving Ability. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37375759>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.

Please share how this access benefits you. [Submit a story](#)

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Department of Physics
have examined a dissertation entitled

Measuring and Assessing Introductory Students' Physics Problem-Solving Ability

presented by Olivia Colleen Miller

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature _____

Typed name: Professor Eric Mazur, Chair

Signature _____

Typed name: Professor Melissa Franklin

Signature _____

Typed name: Professor Vinothan Manoharan

Date: April 25, 2023

Measuring and Assessing Introductory Students' Physics Problem- Solving Ability

A dissertation presented
by

Olivia Colleen Miller

to

The Department of Physics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Physics

Harvard University
Cambridge, Massachusetts
April 25, 2023

© 2023

Olivia Colleen Miller

All Rights Reserved

Eric Mazur**Olivia Colleen Miller**

Measuring and Assessing Introductory Students' Physics Problem-Solving Ability

Abstract

This dissertation investigates the complexities associated with developing expertise in physics problem-solving and the methods for measuring and assessing the skills involved. Problem-solving skills are critical in today's rapidly changing world, and thus, it is important to emphasize them in education. However, lacking precise methods for measuring problem-solving abilities makes it challenging to improve techniques used for teaching these skills. The goal of this dissertation is to broaden our understanding of problem-solving skills in introductory physics classrooms at the undergraduate level. We present four studies that investigate distinct aspects of physics problem-solving skills, each providing better insight into the obstacles students encounter in honing these skills.

The first study focuses on the design and development of three computer-based problem-solving diagnostic instruments to measure respondents' problem-solving processes. From the first two surveys we investigate the novice-expert differences in the problem-solving process, which provide insight into the design and future analysis of the third problem-solving tool. This study also highlights the nonlinearity of problem-solving development, and the third survey will aim to investigate the difference in problem-solving ability in different educational settings.

The second study explores the practices involved in solving a complex, authentic problem in physics and the effectiveness of worked examples in facilitating the learning of these problem-solving practices. While the results suggest that worked examples can help students adopt more effective problem-solving practices, this improvement does not result in successfully solving the problem.

The next study we present investigates two analyses of student epistemologies while reflecting on solutions to physics problems. In the first analysis, we examine the strategies students use to check their answers, which provides insight into the factors that may influence student epistemology of student reflection. In the second analysis, we examine the epistemological framing associated with answer checking, revealing the differences in how students view solution reflection tasks by physics background.

Finally, we present a study exploring how students engage in the metacognitive subcomponent of evaluation on their weekly homework assignments over a semester in an introductory physics course. We investigate the relationship between students' evaluation ability and their course performance as well as trends in evaluation strategies students use by the specific task. We further highlight the difference in evaluation ability by students' college concentration. From our findings on evaluation ability, we provide insights into how this problem-solving skill can be improved.

Table of Contents

Title Page	i
Abstract	iii
Table of Contents	v
List of figures	viii
List of tables	xi
Citations to previously published work	xiii
Acknowledgements	xiv
Dedication	xvi
1 Introduction	1
1.1 Overview	1
1.2 Overview	2
2 Physics problem-solving diagnostic tools	6
2.1 Introduction	6
2.1.1 Importance of problem-solving skills and authentic problems	7
2.1.2 Difference in problem-solving skills between experts and novices	9
2.1.3 Measuring and assessing problem-solving skills	11
2.2 Instrument development and design	14
2.3 Instrument validation	20
2.4 Data collection	22
2.5 Coding framework	23
2.5.1 Mechanics diagnostic tool coding	23
2.5.2 E&M diagnostic tool coding	28
2.5.3 Blended diagnostic tool coding	30
2.6 Results	32
2.6.1 Mechanics diagnostic tool results	32
2.6.2 E&M diagnostic tool results	43
2.7 Discussion	49
2.8 Conclusions, limitations, and future work	52
3 Measuring and teaching problem-solving practices using an interactive task in PhET simulation	54
3.1 Introduction	54

3.2 Theoretical framework	55
3.2.1 Worked examples as a learning mechanism	56
3.2.2 Interactive simulations to engage in problem-solving tasks	57
3.3 Materials	58
3.4 Method	63
3.5 Results	66
3.6 Discussion	71
4 Epistemological framing and strategy use during answer checking	73
4.1 Introduction	73
4.2 Analysis 1	78
4.3 Methods	81
4.4 Results	88
4.4.1 Differences by question type	89
4.4.2 Differences by student level	91
4.4.3 Shifts in strategy	92
4.5 Discussion	93
4.6 Analysis 2	95
4.7 Methods	97
4.8 Results	98
4.8.1 Differences by question type	98
4.8.2 Differences by student level	99
4.8.3 Epistemic frame and strategy shifts	100
4.9 Discussion	101
4.10 Conclusions, limitations, and future work	102
5 Student metacognitive ability on introductory physics assignments	106
5.1 Introduction	106
5.2 Methods	109
5.2.1 Weekly homework assignments	111
5.2.2 Evaluation framework	115
5.3 Results and discussion	118
5.3.1 Importance of evaluation on course score	118
5.3.2 Trends in evaluation over a semester	127
5.3.3 Differences in evaluation ability by concentration	147
5.4 Conclusion	153
6 Conclusions	156
6.1 Review	156
6.2 Importance of teaching problem-solving skills	162
References	163

Appendix A	172
A.1 Mechanics problem-solving diagnostic tool	172
A.2 E&M problem-solving diagnostic tool	180
A.3 Blended problem-solving diagnostic tool	195
 Appendix B	 202
B.1 Complete transcripts for selection of students	202
 Appendix C	 207
C.1 Solutions to the estimation problems	207

List of Figures

Figure 2.1 – Problem stems for mechanics, E&M, and blended problem-solving diagnostic instruments	15
Figure 2.2 – Assessment template based on decisions involved in problem-solving practices from Price et al. (2022).	17
Figure 2.3 – Information respondents are provided in the mechanics problem-solving diagnostic tool	26
Figure 2.4 – List of assumptions for respondents to select from in order to solve the problem in the mechanics problem-solving diagnostic tool.....	27
Figure 2.5 – List of criteria for respondents to bin into three groups based on importance for evaluating the problem-solving process of a student example of attempting the problem in the mechanics problem-solving diagnostic tool.....	28
Figure 2.6 – Four possible circuits solutions for the hidden mystery circuit in the E&M problem-solving diagnostic tool	30
Figure 2.7 – Respondents’ process for devising a plan to solve the problem by physics expertise for the mechanics problem-solving diagnostic tool	36
Figure 2.8 – Respondents’ physics content for devising a plan to solve the problem by physics expertise for the mechanics problem-solving diagnostic tool.....	37
Figure 2.9 – Respondents’ ranking criterion for evaluating problem-solving process by physics expertise for the mechanics problem-solving diagnostic tool. (a) Ranking for criterion “Free body diagram includes all forces in the correct direction” (b) Ranking for criterion “The conditions for static equilibrium have been determined” (c) Ranking for the criterion “Important variables are clearly defined”	43
Figure 2.10 – Respondents’ process for devising a plan to solve the problem by physics expertise for the E&M problem-solving diagnostic tool.....	46
Figure 2.11 – Novices referencing math or equations in devising a plan to solve the problem by whether they selected the correct possible solution for the E&M problem-solving diagnostic tool	49

Figure 3.1 – The black box problem user interface	59
Figure 3.2 – The hidden circuits behind the black-box problems used in the written-response (left) and simulation (right) tasks	60
Figure 3.3 – A data collection example presented to students in the written-response task.....	62
Figure 3.4 – Study design for black box problem: n = 43 participants assigned to the guided written-response task first and the simulation task second, n = 40 assigned to just the unguided, simulation task	64
Figure 3.5 – The problem statement used in both the simulation and written-response tasks	65
Figure 3.6 – Two examples of solutions for the hidden circuit and their score	67
Figure 3.7 – Example of simple vs complex circuits constructed around the black box	68
Figure 3.8 – Percentage of complex circuits built by group	69
Figure 3.9 – Number of times voltmeter was used by group	70
Figure 4.1 – The Chandelier problem	84
Figure 4.2 – The Ladder problem	85
Figure 4.3 – Student strategy use by problem type. The different colors indicate different strategies defined in Table 4.2	90
Figure 4.4 – Student strategy use by students' prior physics experience. The colors correspond to the same strategies as Figure 4.3.....	92
Figure 5.1 – Trellis plot depicting the course scores for students who did not successfully evaluate their solutions and students who did by assignment	119
Figure 5.2 – Histogram depicting the distribution of the frequency students successfully evaluated their solutions for the six assignments.....	124
Figure 5.3 – Percentage of students who successfully evaluated their solution by assignment .	128
Figure 5.4 – Percentage of students who evaluated their solution through reflection using problem definitions and assumptions by assignment.....	130
Figure 5.5 – Percentage of students who evaluated their solution through reflection using knowledge by assignment	132

Figure 5.6 – Percentage of students who evaluated their solution through reflection using strategy by assignment.....	134
Figure 5.7 – Percentage of students who evaluated their solution through reflection using solution by assignment.....	135
Figure 5.8 – Percentage of students who evaluated their solution through mathematical answer-checking strategies by assignment	137
Figure 5.9 – Percentage of students who evaluated their solution using more than one evaluation technique by assignment	139
Figure 5.10 – Percentage of students whose answers were 0 to ± 1 OoM, ± 2 to ± 4 OoM, or $> \pm 4$ OoM from the correct answer for each assignment	141
Figure 5.11 – Percentage of students who successfully evaluated their solution by problem difficulty.....	142
Figure 5.12 – Percentage of students who evaluated their solution through reflection using problem definitions and assumptions by problem difficulty	143
Figure 5.13 – Percentage of students who evaluated their solution through mathematical answer-checking strategies by problem difficulty	144
Figure 5.14 – Percentage of students who evaluated their solution through more than one evaluation strategy by problem difficulty	145
Figure 5.15 – Percentage of students who evaluated their solution through reflection using solution by problem difficulty	146

List of Tables

Table 2.1 – Cognitive processes associated with the levels of Bloom’s Taxonomy	9
Table 2.2 – Qualitative codes for the ‘devise a plan’ problem from the mechanics problem-solving diagnostic tool	25
Table 2.3 – Codes for ‘devise a plan’ problem in E&M problem-solving diagnostic tool.....	29
Table 2.4 – Summary of problem-solving decisions developed by Price et al (2021) applicable for coding ‘devise a plan’ problem in blended problem-solving diagnostic tool	31
Table 2.5 – Process code counts for respondents’ approach to the ‘devise a plan’ question in the mechanics problem-solving diagnostic tool. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem	33
Table 2.6 – Content code counts for respondents’ approach to the ‘devise a plan’ question in the mechanics problem-solving diagnostic tool. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem	34
Table 2.7 – Assumptions needed to solve the problem in the mechanics problem-solving diagnostic tool. Percentages represent the percentage of respondents that selected that assumption as being needed to correctly solve the problem	38
Table 2.8 – Criteria for evaluating the problem-solving process in the mechanics problem-solving diagnostic tool. Percentages reported are the percentage of respondents who ranked a particular criterion in each category (most important, medium importance, or least/no importance)	41
Table 2.9 – Process code counts for respondents’ approach to the ‘devise a plan’ question in the E&M problem-solving diagnostic tool. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem	44
Table 2.10 – Content code counts for respondents’ approach to the ‘devise a plan’ question in the E&M problem-solving diagnostic tool. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem	45
Table 2.11 – Process code counts for novices’ approach to the ‘devise a plan’ question in the E&M problem-solving diagnostic tool by correct solution. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem	47

Table 2.12 – Content code counts for novices’ approach to the ‘devise a plan’ question in the E&M problem-solving diagnostic tool by correct solution. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem	48
Table 3.1 – Rubric for scoring participants’ solutions for the hidden circuit	67
Table 3.2 – Summary of whether students referenced applying Ohm’s Law to solve the hidden circuit problem by group.....	71
Table 4.1 – Number of students by level and type of problem. The number of students at each level who completed each problem was not statistically different (Fisher's exact test, $p = 0.77$)	82
Table 4.2 – Coding scheme for strategy use including code name, code definition, and an example of each code from the data	87
Table 4.3 – Code counts for students' strategy use during solution reflection. Percentages represent the total percentage of lines within each column that students were using a particular strategy. The total number of lines is given in the header	89
Table 4.4 – Mean and standard deviation (in parentheses) of strategy shifts per student by student level.....	93
Table 4.5 – Epistemic frames and resource definitions adapted from Shar et al. (2020), including examples of each code from the current data set	98
Table 4.6 – Counts of epistemic frames by student level and problem type. N is the total number of codable lines within each column and category, and the percentage represents the fraction of responses that were coded as each frame/resource within that category by column count	99
Table 4.7 – Mean (and standard deviation) of epistemic frame shifts per student by student level	100
Table 5.1 – Number of students by undergraduate year and concentration.....	111
Table 5.2 – Six estimation problems.....	115
Table 5.3 – Framework for how students evaluated their solutions on the estimation problems	116
Table 5.4 – Summary of fixed effects estimates from multi-level analysis investigating the relationship between whether students successfully evaluated their solutions and concentration controlling for college year and previous physics experience	150

Citations to previously published work

Parts of this dissertation cover research reported in the following articles:

Burkholder, E. W., Miller, O. C., & Blackmon, L. (2023). Investigating the epistemology of physics students while reflecting on solutions. *European Journal of Physics*.
<http://iopscience.iop.org/article/10.1088/1361-6404/acce0c>

Wang, K., Miller, O., Miller, K., Burkholder, E., & Wieman, C. (2022, April 21-26). *Measuring and teaching problem-solving practices using an interactive task in PhET simulation* [Paper presentation]. AERA, San Diego, CA, United States.

Acknowledgments

Before starting my journey in education research, I was engaged in a completely different research area but with no clear goals for my future in that field. I decided to take a leave of absence to explore a teaching opportunity at an international school teaching physics and math to middle and high school students. While I loved teaching and helping students develop a better understanding of the world around them, I knew I still wanted to pursue research and my Ph.D. Upon returning to graduate school I was uncertain as to what research I wanted to focus on but I knew I wanted to somehow incorporate my love for working with students. I am immensely grateful to Jacob Barandes for not only supporting me throughout this journey but also for recommending that I explore physics education research.

I have been very fortunate to have many incredible mentors along this journey. I want to first thank the physics department at East Tennessee State University for fostering a love for physics and for supporting me in developing research skills. I want to especially thank Dr. Mark Giroux and Dr. Beverley Smith for first introducing me to physics research opportunities and to Dr. David Close and Dr. Yuriy Razskazovski for mentoring me in my undergraduate thesis. My experience in the ETSU physics department is one that I will always cherish and has helped lead me to where I am today.

I would especially like to thank Eric Mazur for welcoming me into his group and for his tireless mentorship and support. I feel very lucky to have him as an advisor. I also want to thank Melissa Franklin for helping me find my path, especially when I felt very unsure of myself. I am also grateful for the support of Vinny Manoharan in my journey as a graduate student. And of

course, I cannot forget the countless hours I have spent in Carol Davis's office. I am so thankful for all of her support through the ups and downs of graduate school and life.

I would also like to thank the Mazur Group and the education subgroup. I especially want to thank Kelly Miller who helped me get started in physics education research projects and who introduced me to a network of other education researchers. I would also like to thank Catherine Crouch, Greg Kestin and the Science Education Research Group, and Louis Deslauriers who have all helped provide me with a great learning community. I am immensely thankful for my colleague and dear friend Isaura Gallegos who has not only taught me so much about education research but has also been an incredible support throughout. And a special thanks to Virginia Casas for taking care of the necessary details to help the Mazur Group run smoothly.

I am grateful for the opportunity to learn from and collaborate with the Wieman Group at Stanford. They have provided me with a rich learning environment and have motivated me to challenge myself further. I specifically want to thank Carl Wieman, Karen Wang, Eric Burkholder, Argenta Price, and Shima Salehi. Thank you for welcoming me into your group and providing me with thoughtful feedback on my research endeavors.

Lastly, I would like to thank my family for their never-ending support and encouragement. I especially want to thank my parents for providing me with amazing opportunities to grow as a person and for always supporting my dreams. Without you, none of this would have been possible.

To my parents, Vickie and Terry Miller, for their endless support and encouragement, and for teaching me the importance of a sense of humor.

And, to my cats, Bella and Marcel, who have been distracting during this process but have kept me in good spirits.

Chapter 1

Introduction

1.1 Overview

Problem solving is an essential skill not only in academia but in everyday life. In today's rapidly changing world, where new challenges and problems emerge frequently, the ability to solve problems has become more critical than ever. With technology advancing and new problems arising, individuals need to be adaptable, flexible, and able to think critically to overcome these challenges. Thus problem solving skills should be a central focus of a student's education.

It is crucial to comprehend the complexities associated with developing expertise in physics problem solving to facilitate students in becoming proficient problem solvers. Facilitating the development of effective problem-solving skills among students is particularly an important objective of many physics courses, and it should be a fundamental aspect of any instructional approach aimed at guiding novices toward expertise. However, finding effective methods for both assessing and teaching problem-solving skills remains a challenge. Assessment and instruction are interconnected, and thus lacking precise methods for measuring problem-solving abilities makes it impractical to improve techniques used for teaching these problem-solving skills (Price et al., 2022). Therefore, in order to better understand the struggles students have in developing expert-level problem-solving skills and then implement teaching methods to address these challenges, we need to be able to measure these skills. The goal of this

dissertation is to broaden our understanding of problem-solving skills in the context of introductory physics classrooms at the undergraduate level. We present studies on measuring problem-solving ability in introductory physics students, providing better insight into the obstacles students encounter in honing these skills.

1.2 Organization of dissertation

The dissertation is composed of four main projects, each of which investigates a distinct aspect of physics problem-solving skills. The focus of all four endeavors is to enhance our understanding of the barriers introductory physics students face in developing these skills. Through measuring and assessing specific problem-solving skills we can better understand how to improve methods for teaching students to develop expertise.

We first present a study focusing on the development and implementation of three separate computer-based problem-solving diagnostic instruments aimed to elicit the respondent's problem-solving process. Each of the three problems presented in the diagnostic surveys presents the respondent with a complex and authentic problem because, as emphasized by Newell and Simon (1972), engagement in problem solving necessitates a novel activity/experience where one does not know the steps involved beforehand, and the solver must make decisions based on available information. Therefore, in order to investigate the differences in problem-solving processes between experts and novices, the problems must also be novel to the experts.

As problem solving is complex there need to be measurement techniques to capture the multifaceted nature of the process and skills involved. Early studies examining the expert-novice differences in problem solving did so through interviews and think-aloud protocol (Chi et al.,

1981; Reif & Heller, 1982; Heller & Reif, 1984). While these methods are important for developing a problem-solving diagnostic tool, they are very tedious as well as time-consuming. We have thus focused on a computer-delivered assessment to allow us to capture problem-solving ability of many respondents with varying expertise. Furthermore, this method can assist in examining the difference in problem-solving development among novices in various educational environments.

The next study we present explores the practices involved in solving a complex, authentic problem in physics as well as whether and how a worked example can help students learn these practices. To solve novel problems, individuals must utilize their working memory to process information (Newell and Simon, 1972; Simon 1974; Anderson, 1993). However, working memory has its limitations, and an individual's reasoning ability is directly related to their capacity for processing information in working memory when making decisions (Kyllonen & Christal, 1990). To enable students to become proficient problem solvers, it is essential to recognize the constraints of their information-processing ability while learning. This means fostering effective problem-solving methods by taking into account their prior knowledge and providing appropriate scaffolding to minimize their cognitive load while attempting to solve problems (Sweller, 1988; Anderson, 1995). Worked examples enable learners to concentrate on a particular problem state and its corresponding solution by providing them with the necessary solution steps to arrive at the answer. This, in turn, allows the learners to direct their cognitive resources toward constructing solution schemas for specific types of problems (Sweller, 2006). However, it is unclear how worked examples may facilitate learning in complex and novel problem-solving tasks. In this study, we thus examine how a brief task embedded in an

interactive circuit simulation can be used to elicit students' problem-solving performance and if a worked example helps students learn problem-solving practices.

In the third study presented in this dissertation, we are interested in the epistemology of reflection on a solution, not just the actions involved in such reflection, but also why students may approach reflection tasks the way they do. We are particularly interested in understanding the variation in reflection approaches within individuals, as well as how it is evident among individuals who share a similar background in physics. Becoming a proficient problem solver in physics not only requires establishing links between mathematics and physics principles, so as to effectively unpack, interpret, and apply the laws of physics, but it also requires the process of sense-making to develop a foundation of knowledge in physics and solve novel problems in diverse situations (Maries & Singh, 2023). It is known that novices tend to struggle with engaging in sense-making while solving problems and tend not to reflect on their solutions automatically (Ambrose et al., 2010; Burkholder et al., 2020). It is unclear whether this is due to the difficulty of the task and whether novices lack the appropriate mental models and content knowledge or whether they do not use these strategies because they do not perceive them as relevant. Investigating this question, however, is complex because of the dynamic and ever-changing nature of epistemology, both on an individual level and across individuals within a system. We, therefore, aim to investigate students' approach to reflecting on a solution to a physics problem from an epistemological perspective by analyzing the prevalence of different answer-checking strategies across student levels and different problem types as well as examining the epistemological framing associated with answer checking to provide a more detailed picture of this epistemology both within and across individuals.

The final study we present in this dissertation investigates the metacognitive ability of introductory physics students on their weekly homework assignments. Metacognition refers to a person's ability to reflect on and regulate their own thinking processes and learning (Shraw, 1998), and metacognitive skills are particularly important for problem solving and learning in knowledge-rich domains like physics. Research has indicated that students who are more metacognitive during physics problem solving are more likely to correctly solve physics problems (Neto & Valente, 1997; Rozenchwajg, 2003). This is likely because metacognitive skills enable individuals to monitor their knowledge and skill levels, plan and allocate learning resources efficiently, and evaluate their learning state.

Evaluation is one of the six subcomponents of metacognition, and it involves appraising the products and regulatory process of one's learning (Taasoobshirazi & Farley, 2013). Experts are distinguished by their ability to effectively apply reflective practices and evaluation techniques during problem solving (Price et al., 2021) while novice problem solvers tend to adopt a more linear approach to problem solving (Heller et al., 2001). In this study, we aim to better understand the trends in how students evaluate their homework solutions over the course of a semester of an introductory physics course. By doing so, we seek to gain a deeper understanding of the metacognitive ability of students in an introductory physics course and pave the way for future work on enhancing their problem-solving skills in physics.

In the final chapter of the dissertation, we summarize the findings and suggest implications for instruction to promote problem-solving ability.

Chapter 2

Physics problem-solving diagnostic tools

2.1 Introduction

Problem solving is arguably one of the most important skills an undergraduate education in science, technology, engineering, and mathematics (STEM) can help develop. Problem-solving skills are not only important within the classroom and research projects but are also valuable skills in all areas of life. In a 2011 study of physics PhDs ten years into their careers, AIP reported that out of forty-seven themes described to contribute to career success problem-solving skills were the second most important factor, second only to hard work (Porter, 2019). Moreover, problem-solving skills were indicated as a frequent theme that contributed to career success in private sector careers, academia, and even government positions. Given its importance in post-college career success, pedagogies focused on problem-solving skills is currently one of the most important topics of physics education.

While a key goal of science and engineering education is to produce good problem solvers, it remains unclear how to best teach and measure these skills. In this study we aim to first highlight the importance of problem-solving skills in an authentic context and the key features of expert versus novice problem solvers. We then address how these skills can be measured and assessed. Our study focuses on the development and implementation of three separate problem-

solving diagnostic instruments (surveys) to elicit the problem-solving process of novices and experts.

2.1.1 Importance of problem-solving skills and authentic problems

The Program for International Student Assessment (PISA) defines problem-solving competency as “an individual’s capacity to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious” (OECD, 2014). Problem solving has also been defined as “the cognitive process of moving towards a goal when the path is uncertain” (Heller, 2010). That is, the desired outcome is known, but how to get there is unknown. Similarly, Carl Wieman and Wendy Adams define problem solving as “cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver” (2015). These definitions emphasize that real problem solving does not have a direct path. Furthermore, the designation of a task as a problem is dependent upon the solver. What is a problem to students might not be, and is probably not, a problem to the instructor in that if they know how to solve it, then it is by definition not a true problem. This brings about an important distinction between ‘authentic’ problems and exercises. Authentic physics problems that students are expected to solve after college are unrestricted because “they possess conflicting goals, multiple solution methods, (nonscientific success standards, outside constraints), unanticipated problems, distributed knowledge, collaborative activity systems, and multiple forms of representation” (Jonassen, 2014). In contrast, exercises, like the typical textbook-like ‘problems’, usually include all the information needed for the solution’s single objective, and the assumptions are already made for the students. Students can thus apply a known procedure to find/ an unknown answer in exercises.

Researchers agree that problem solving requires both relevant knowledge and appropriate cognitive strategies. The problem solver must construct knowledge and organize cognitive and effective behavioral processes toward a specific goal, which is a process closely related to creativity (Ince, 2018). These cognitive behaviors involved in problem solving align with the educational goals at the top of Bloom's Taxonomy (Armstrong 2010). Bloom's Taxonomy is a framework for categorizing and describing the cognitive processes by which thinkers encounter and work with knowledge. There are six levels of cognitive learning according to Bloom's Taxonomy: remember, understand, apply, analyze, evaluate, and create. Table 2.1 describes each of the levels in more detail. Most exercises, or typical textbook problems, only reach the bottom category of 'remember' on Bloom's Taxonomy because they only involve information retrieval and following a familiar procedure from previously solved problems. Authentic problems, however, require cognitive processes that reach the upper levels of the taxonomy because they require the solver to analyze, plan, and monitor their own behavior to formulate unfamiliar solutions.

Table 2.1: Cognitive processes associated with the levels of Bloom's Taxonomy

Bloom's Level	Descriptor	Cognitive Processes
Create	Produce new or original work	Design, assemble, construct, develop, formulate, investigate
Evaluate	Justify a stand or decision	Appraise, argue, defend, select, support, critique, weigh
Analyze	Draw connections among ideas	Differentiate, organize, relate, compare, contrast, examine, experiment, question, test
Apply	Use information in new situations	Execute, implement, solve, use, demonstrate, interpret, operate
Understand	Explain ideas or concepts	Classify, describe, discuss, explain, identify, report, select
Remember	Recall facts and basic concepts	Define, duplicate, list, memorize, repeat, state

2.1.2 Difference in problem-solving skills between experts and novices

In order to help students develop more expert-level problem-solving skills, it is necessary to first better understand the difference between novice and expert problem solvers. Many of these differences have been well-studied in physics education. Ince defines expert problem solvers as those who can use problem-solving strategies efficiently and continuously, whereas those who cannot use these strategies adequately are defined as novice problem solvers (Ince, 2018). When first presented with a problem, research has shown that experts and novices differ in how they represent the problem. Representations during problem-solving can describe the

depth of consideration of the problem solver, that is, whether they are representing surface features of the problem or representing the deeper physics (Chi et al., 1982). Experts have been found to represent problems by their deeper conceptual structures while novices tend to categorize problems by surface features (Larkin et al., 1980; Chi et al., 1981; Heller & Reif, 1984). Kohl and Finkelstein (2008) found that a further distinguishing feature between expert and novice problem solvers is how they used their problem representations. Experts engage in more analysis while novices are “more likely to behave mechanically or algorithmically and to produce multiple representations without being able to make much use of them” (Kohl & Finkelstein, 2008).

Studies have also found differences in expert and novice knowledge structure and strategy in problem solving. Reif and Heller (1982) describe experts to organize their knowledge in a structured, cohesive fashion, and they are able to activate the required knowledge structure as needed in problem solving. Furthermore, experts are reported to often use qualitative arguments to first plan solutions before approaching them mathematically as well as make problem-solving decisions by first exploring their consequences (Larkin & Reif, 1979; Reif & Heller, 1982; Eylon & Reif, 1984). Novices, however, lack the same knowledge structure as experts and tend to encounter difficulties in adequately describing problems qualitatively. Because they struggle to establish connections between the problem information and physics concepts, novices tend to engage in limited prior planning and thus try to assemble problem solutions by stringing together loosely connected mathematical formulas from their repertoire, especially when presented with complex problems (Reif & Heller, 1982). This expert-novice difference in approach emphasizes the importance of framing and planning in problem solving.

While solving authentic problems, Wieman and Adams (2015) found experts spend more time analyzing, planning, and managing their own behavior, demonstrating a holistic view and systematic problem-solving approach. Such systematic steps to solving a problem include understanding the problem, determining the concepts, making the plan, calculating a solution, and evaluating the outcome (Heller & Heller, 2010). Experts engage in self-monitoring and evaluation throughout this process whereas novices tend to view problem solving as a more linear process (Heller et al., 2001). Novices thus generally have a more superficial approach to problem solving while expert problem solvers have an in-depth decision-making mechanism (Hardiman et al., 1989). However, if the problem is devoid of an authentic context, experts may behave like novices because the problem is more formulaic for the experts to solve (Singh, 2002). This further supports the need for authentic problems in investigating problem-solving skills.

2.1.3 Measuring and assessing problem-solving skills

The aforementioned studies have shown that an individual's problem-solving skills are related to the performances of the problem-solving sub-steps (Ince, 2018). Thus in order to better understand problem-solving ability, it is necessary to measure and assess the problem-solving decisions. A range of methodologies have been employed to capture the many aspects of problem solving, including, interviews (Price et al., 2020), think-aloud protocols (Alii et al., 2016), expert self-reflections (Polya, 1945), educational problem-solving video games (Chopade, 2018), and computer-delivered assessments (PISA, 2012). Below we highlight some of the broad categories of methodologies and provide an overview of the challenges faced in measuring problem-solving skills.

Most studies focused on measuring and assessing problem solving use interviews and/or think-aloud protocols. In these studies, novices and often experts are interviewed by researchers who pose specific questions designed to understand the approach being taken to solving a given problem. Similarly, think-aloud protocols involve a semi-structured cognitive interviewing method where a person is asked to verbalize their thought process as they engage in a specific task, and during this they are recorded for further analysis. Cognitive interviewing techniques are specifically used to study the manner in which a targeted audience understands, processes, and responds to materials presented to them, with a particular emphasis on the potential breakdowns in this process (Willis, 2005). In several studies on cognitive interviewing techniques, including that of Ali et al. (2016), students are first asked to narrate their thinking while solving physics problems, followed by qualitative interviews to elicit student recollection of their thinking while solving the problems. These interviews are typically transcribed and then analyzed and coded by subject-domain experts. Interrater reliability is monitored by measuring the level of agreement between the expert coders to ensure validity. The generation of codes is often an iterative process. For example, Price et al. (2020) interviewed over 50 scientists and engineers about problem-solving expertise in their respective fields. From the interviews, they generated a list of decisions commonly made by experts while solving an authentic problem in their field. This list was then refined and validated via semi-structured interviews. These interviews were then coded for the decisions represented and provided a framework of 29 different decisions that characterize the expert problem-solving process.

In addition to interviews and think-aloud protocols, rubrics are a common tool used to evaluate written work and are often used as part of the methodology in problem-solving studies. In a 2016 study, Haim et al. investigated students' ability to apply problem-solving strategies in

physics. Students were asked to solve routine problems on paper and rubrics were utilized as the measurement tool. Similarly, Burkholder et al. (2020) developed a problem-solving template for students to use while working through problems. This template was implemented to not only scaffold the problem-solving process for students but to also serve as a rubric to measure how expert-like students are in their problem solving.

A less common methodology for measuring problem-solving ability is computer-delivered assessments. A key example of such an assessment is the PISA computer-delivered assessment which first assessed problem solving in 2012. The problems used in the assessments do not require specific disciplinary knowledge, but instead, focus on measuring the cognitive processes that are fundamental to problem solving. During these assessments, students are presented with an authentic scenario and a series of test items designed to measure their performance in four cognitive processes: exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting. These test items are presented to the student one at a time, and the student must proceed from item to item without being able to skip forward or return to previous items. The assessment consists of a variety of response formats which can be automatically coded including multiple-choice, drag-and-drop, constructed-response as well as free-text response which requires coding by experts. Test item scores are combined to provide an overall problem-solving score for each student.

When constructed appropriately, computer-delivered assessments can be an efficient method for measuring problem-solving ability in many novices and experts within a short time. Though interviews and think-aloud protocols are important for the development of such tasks, they are very time-consuming to conduct, transcribe, and then analyze. In the following sections,

we present the development and implementation of three separate computer-delivered problem-solving diagnostic surveys.

2.2 Instrument development and design

We have developed three physics problem-solving diagnostic instruments. The first instrument focuses on a problem in mechanics while the second one focuses on a problem in electricity and magnetism (E&M). The third instrument is less clearly focused on one particular physics topic and incorporates elements of both mechanics and E&M topics. This third survey was developed based on the findings and feedback from the other two surveys. Each of these diagnostic tools is designed to elicit the problem-solving process and decisions a respondent would take to solve a complex, authentic physics problem without actually having to solve the problem. The stems for each of the three problems are presented in Figure 2.1.

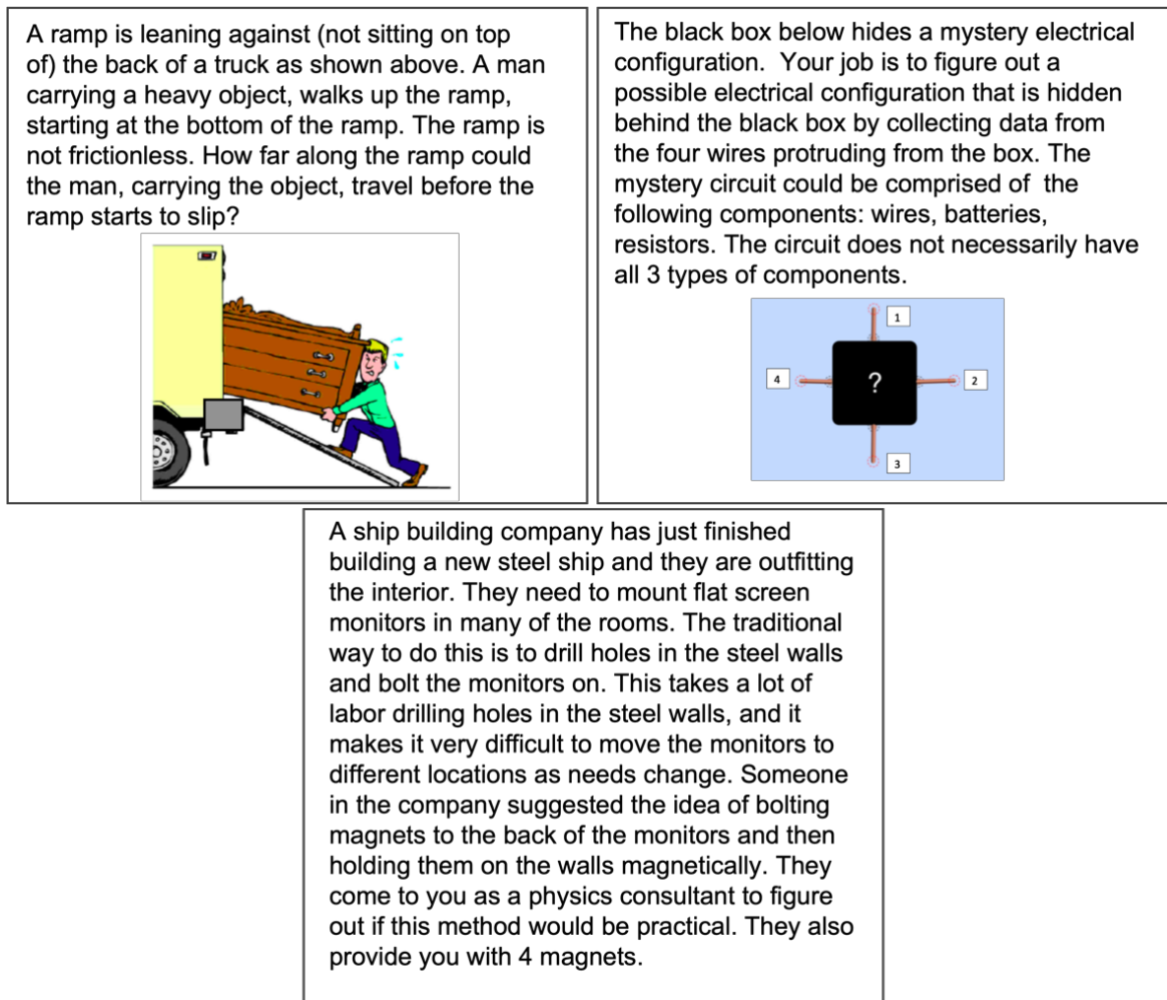


Figure 2.1: Problem stems for mechanics (left), E&M (right), and blended (bottom) problem-solving diagnostic instruments

The surveys were developed similar to the PISA 2012 problem-solving assessment in that the questions consist of a variety of response formats: free-response, multiple choice, many-choice, and drag-and-drop. Our surveys do, however, focus on more so free-response prompts as they allow for the respondent to describe a wide range of ideas pertaining to the given situation as opposed to having to choose a single or subset of responses that are most salient to them, as is the case for most multiple choice or many-choice surveys (Hull et al., 2022). The disadvantage of free-response prompts, though, is the time demanded for the respondents to fill out the survey

as well as the time demanded for assessing the responses. This aspect of the survey development will be discussed later in how to further improve the design of such diagnostic tools for both the respondents and the education researchers analyzing the responses.

Unlike the PISA assessments which aim to measure general problem-solving skills, our goal is to measure problem-solving skills of students (novices) training to become experts at solving problems in the specific domain of physics just as the problem-solving assessments developed by Price et al. (2022) aim to do in the domain of science and engineering. Our problem-solving surveys thus follow a very similar structure as that developed by Price et al. (2022) which involves the respondent recognizing and applying appropriate specialized knowledge given limited information when presented with an authentic discipline-specific problem scenario. Also, unlike the PISA assessment, our task, like Price et al.'s (2022), requires the respondent to decide “what are the criteria for a good solution, what kind of information is needed to solve the problem, [and] how to find that information” instead of simply assessing whether students can collect from the task the information necessary to solve the problem.

The general assessment structure from Price et al. (2022) is shown in Figure 2.2. The components of this assessment template are based on the decisions involved in problem-solving practices (Price et al., 2021; Salehi, 2018). We implemented a similar structure in our diagnostic tools following the steps: 1) provide an authentic problem scenario; 2) ask a series of questions that require the respondent to make decisions about problem definition and planning how to solve; 3) provide more information; 4) ask a series of questions that require decisions about interpreting information and drawing conclusions; and 5) have respondents choose and reflect on their solution. In order to further emulate the problem-solving process used in real situations, multiple rounds of providing new information followed by steps B-F can be included, as we did

for our E&M survey. Moreover, all three of our surveys are designed to start with broad, un-scaffolded questions about the process the respondent would use to solve the problem. Subsequent questions are progressively more specific to constrain the planning and interpretation.

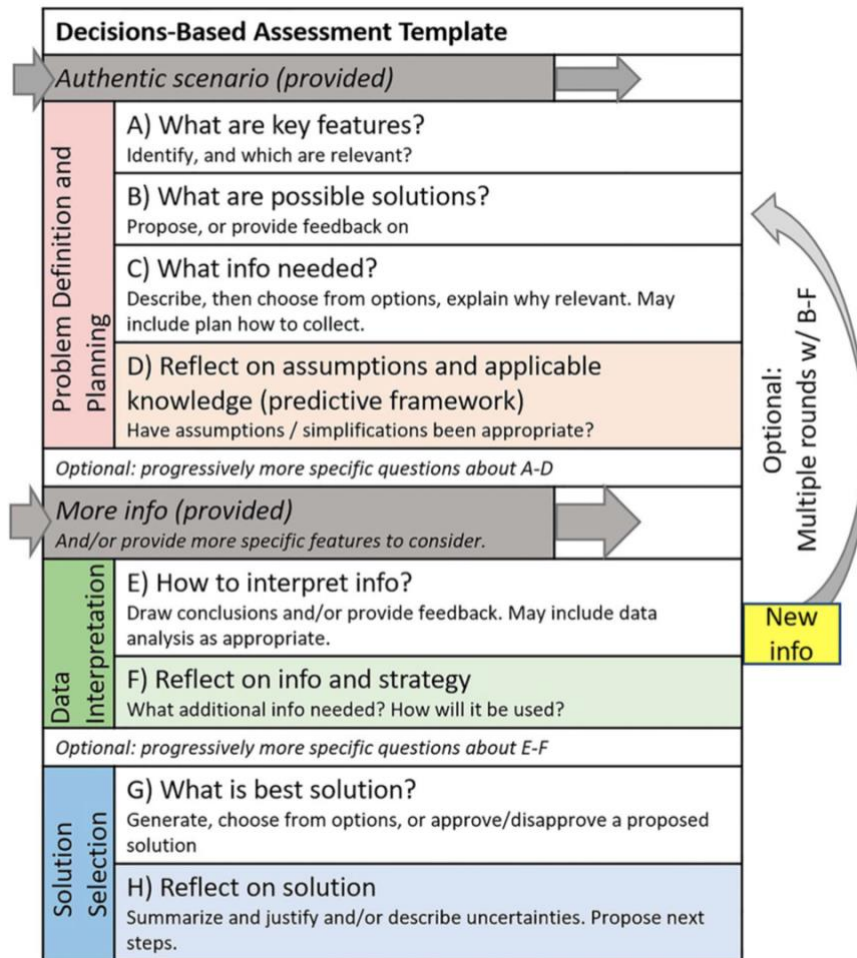


Figure 2.2: Assessment template based on decisions involved in problem-solving practices from Price et al. (2022).

The mechanics survey and the blended mechanics and E&M survey, which we will henceforth refer to as the blended survey, both have 14 questions, while the E&M survey has 15 questions. (The full surveys can be found in Appendix A). The first question on the mechanics survey is “Write out a detailed plan for how you would solve the problem. Be sure to include a

list of quantities that you would need to know and assumptions you would have to make.”

Similarly, the first question on the E&M survey is “Write out an approach for how you would figure out the mystery circuit configuration and the values (resistances, voltages) for any of the components.” Based on feedback from these two surveys, we aimed for the prompt of the blended survey to elicit the respondent’s initial problem-solving approach without any scaffolding to include assumptions or quantities needed to solve the problem. The prompt for the blended survey is “What would you do to determine if this method of hanging the monitor would work? Write a detailed plan.” This prompt is then followed by a question asking what information would be needed to determine if the method of hanging the monitor would work, which is followed by a question asking what assumptions or approximations would need to be made to determine if the method of hanging the monitor would work. These two questions are designed to prompt the respondent to provide further insight into what is needed to solve the problem if they did not already address these in their first response describing a detailed plan.

After committing to a detailed plan in the first questions for the mechanics and E&M surveys, respondents are then provided with gradually more information about the problem in subsequent questions. For example, the second question on the mechanics survey provides a list of quantities and then asks the respondent, “Given the quantities provided, which of the following assumptions would you need to make to be able to correctly solve the problem?” A list of 14 possible assumptions is provided and respondents must select the ones they think are necessary. The next mechanics question provides a sample student solution which contains errors. Respondents are first asked to provide a list of criteria they would use to evaluate the sample solution. Then, in the next section, they are given a list of 12 criteria and asked to group them into bins according to their importance for evaluating the solution. For the E&M survey,

the respondent is prompted to address how they would collect data and then how they would specifically use electrical components to collect this data. They are next presented with a specific configuration of connecting the terminals of the black box without and later with a battery along with tables of voltage and current data. The respondent is asked to interpret this data and then describe an updated plan on how they would proceed to determine the hidden circuit configuration.

Unlike the mechanics and E&M surveys, the blended survey focuses less on giving the respondent data to interpret but instead focuses on eliciting decisions the respondent would need to make to solve the problem. For example, respondents are asked to represent the problem with a sketch, decide what conditions need to be satisfied and decide what physics concepts would be needed in solving the problem. The blended survey also prompts the respondent to write an equation that would represent conditions that need to be satisfied followed by a chance to provide an updated diagram to represent the problem. Similarly, the survey allows the respondent to update their initial assumptions after they are prompted to address what measurements would need to be made and how a specific location can affect the experiment. In providing the opportunity to update their sketch and assumptions, the blended survey prompts the respondent to reflect on the planning and interpretation decisions they made and allows us to observe whether the respondents engaged in such reflective practices, more so than the mechanics and E&M surveys.

The final questions of the mechanics survey involve interpreting the answer given by the flawed student solution and providing conditions needed for something not to slide. This gives the respondent a chance to demonstrate their understanding with additional scaffolding as well as allows us to investigate how they interpret information provided and whether they reflect on the

implications of the answer. For both the E&M and blended surveys, the respondent is presented with possible answers to the problem that they must select from as the best solution possible provided. They are then asked to justify their choice. These final questions allow us to probe whether the respondent is able to reach an appropriate conclusion based on the data and information provided and if and how well they are able to reflect on this solution.

For each of the three surveys, we start with the broad, un-scaffolded “devise a plan” question so that we do not influence the respondent’s process-oriented thinking with the more scaffolded questions. This progression from the open-ended to the scaffolded questions is an intentional structure designed to gather as much information as possible from the respondents with a wide range of problem-solving skills. Respondents are not able to return to previous questions, so increasing the scaffolding as the respondent progresses from one section to the next allows us to gather information about a respondent’s problem-solving process even if they are not able to devise a cohesive plan in the first section.

2.3 Instrument validation

To validate our problem-solving diagnostic surveys, we followed methodology described in “Development and Validation of Instruments to Measure Learning of Expert-Like Thinking” (Adams & Wieman, 1992). This methodology is based on the “Standards for educational and psychological testing” American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME, 1999). The methodology stresses the use of student interviews as the most important part of developing and validating assessments designed to measure the expertness of students’ thinking. Below is an outline of the procedure we

undertook to develop and validate each of the three problem-solving diagnostic surveys. The process was iterative and involved student and expert interviews at several stages.

1. We consulted several experts to solicit advice and feedback on the initial problem stems. These experts were all college instructors of introductory physics who had been teaching for at least ten years. Several candidate problems were considered and iterated. The following selection criteria was considered:
 - i. Authenticity – solution method is not immediately obvious to even expert problem solvers
 - ii. Complexity – problems complex enough that the solver must state their assumptions and have a detailed method to be able to solve them
 - iii. Open-ended – problems with multiple solution paths which depend on initial assumptions.
2. We conducted interviews of both students and experts. We interviewed six introductory physics students and four introductory physics instructors. We asked them to talk through the ways in which they would solve the problems. These interviews allowed us to identify student thinking about the problems and the various ways this was different from expert thinking.
3. We created open-ended survey questions to probe student thinking more broadly. The first version of the surveys was administered to a group of introductory physics students at Harvard University. We subsequently interviewed a subset of these students chosen based on the diversity of their responses to the first “devise a plan” question.

4. We adjusted the surveys based on students' responses and interview findings. We converted some of the open response questions into multiple-choice questions based on common student responses.
5. We conducted validation interviews on the updated versions of the survey with novices and experts different from those interviewed in step 2.
6. We administered the surveys to students and experts on a wider scale.

2.4 Data collection

The diagnostic problem-solving surveys are administered through the Harvard Qualtrics interface. Qualtrics is a web-based software for creating surveys and generating reports on the collected data. Upon logging on to each survey, respondents are told that they will be presented with an introductory physics problem with questions designed to probe their thinking about how to solve the problem. Respondents are told that the questions will be presented in sections, and they will not be able to go back to previous sections once they advance to the next one. They are also instructed to only use the information that is provided in the problem and to not use any outside resources, including collaboration with others. Finally, the respondents are informed that the survey should be completed in a single sitting, and while there is no time-limit, the survey should take around 30-45 minutes to complete.

We have collected data with both the mechanics and the E&M surveys ($N = 138$ for the mechanics survey and $N = 96$ for the E&M survey). The majority of these respondents were introductory physics students at Harvard College, the Harvard Extension School (post-baccalaureate students) and Stanford University. For the mechanics survey, 116 respondents of the 138 were introductory physics students, 16 were graduate students (from Harvard and

Stanford), and 6 were instructors (all of whom had been teaching college-level introductory physics for at least 10 years). For the E&M survey, the breakdown of respondents was 88 introductory students, 3 graduate students, and 5 instructors. We have collected preliminary data on the blended survey and have plans to collect data in two different introductory physics classes at Harvard College during the Spring 2023 semester to investigate the development of problem-solving skills in a certain educational experience.

2.5 Coding framework

To investigate the trends in the responses to the problem-solving diagnostic tool data, we have developed sets of codes by question to capture the qualitative data. This type of qualitative coding involves systematically labeling and grouping similar types of data to generate themes and make analyzing the data possible. Qualitative coding of the responses to the problem-solving surveys allows us to interpret and analyze the differences in problem-solving ability more readily and accurately. Below we highlight the qualitative codes developed for select questions from each of the three problem-solving diagnostic tools.

2.5.1 Mechanics diagnostic tool coding

For the mechanics problem-solving diagnostic tool, we have developed a set of codes addressing the first three questions of the survey. These questions are focused on how the respondent would devise a plan to solve the problem, what assumptions they would need to make to solve the problem given a list of quantities, and how they would group criteria for evaluating the problem-solving process for the provided worked example. Codes for subsequent problems have been developed but were later dismissed due to their focus only on content-specific

understanding rather than problem-solving process. The first question we coded for is an open-ended question and asks the respondent to “Write out a detailed plan for how you would solve the problem. Be sure to include a list of quantities that you would need to know and assumptions you would have to make.” This question generated the greatest diversity of responses and was captured by a set of 22 codes. We developed codes for this question through iterative rounds. Two experts studied the responses independently and generated labels (codes) independently based on emergent features in the data. These codes were compared and revised, and then through several rounds, a common set of codes was established. Each expert then independently coded a subset of about twenty responses, and these codes were then compared. In any cases in which the codes did not match up they were discussed and resolved. The codes for this ‘devise a plan’ question consisted of a subset of process-specific codes (10) and content-specific codes (12) as represented in Table 2.2.

Table 2.2: Qualitative codes for the ‘device plan’ problem from the mechanics problem-solving diagnostic tool

Process Codes	Content Codes
Procedural versus holistic approach	Pivot
Detailed versus brief	Normal force
States assumptions	Torque
States quantities needed	Friction
Math or solving equations	Static friction
Free body diagram (or mention of drawing a diagram)	Work energy theorem
Labels problem as a ‘torque problem’ or a ‘statics problem’	Sum forces
Evidence of simplifying the problem	Separate problem components into x and y axes
Defines specific set of axes	Mentions constant velocity
Defines what it means for ladder to slide	Kinematics
	Newton’s 2 nd law
	Static equilibrium

After providing the respondent with a set of quantities relevant to solving the problem as shown in Figure 2.3, the second question we coded for asked the respondent which of the following assumptions out of a list of fourteen would need to be made to be able to correctly solve the problem. The list of assumptions provided is shown in Figure 2.4. For this problem, we initially coded for the differences between which assumptions were selected among different respondent groups by expertise (novices, intermediates, experts). After feedback on possible ambiguities in the assumptions that would need to be made, we then coded by expert consensus meaning that we investigated/coded for how closely aligned novice and intermediate responses were to what experts agreed were the correct assumptions that would need to be made.

The following information is provided by the problem:

Length of ramp $l = 5 \text{ m}$

angle between floor and ramp $= 53^\circ$

Coefficient of static friction between the ramp and the floor is the same as the coefficient of static friction between ramp and the truck $\mu_s=0.4$

Mass of ramp $m_1 = 50 \text{ kg}$

Mass of man and heavy object $m_2 = 100 \text{ kg}$

Figure 2.3: Information respondents are provided in the mechanics problem-solving diagnostic tool

Given the quantities provided, which of the following assumptions would you **NEED** to make to be able to correctly solve this problem?

- ☐ the pivot point is at the center of mass of the ramp
- ☐ the pivot point is at the top of the ramp
- ☐ the pivot point is at the bottom of the ramp
- ☐ the pivot point is at some point along the ramp
- ☐ the heavy object and man are in the same position on the ramp
- ☐ the static friction forces prevent the ramp from sliding until the man gets far enough up the ramp
- ☐ the ramp will start to slide before the man reaches the truck
- ☐ all the forces are initially balanced
- ☐ friction between the ramp and the truck is negligible
- ☐ friction between the ramp and the floor is negligible
- ☐ the truck won't roll during the process
- ☐ the man's feet won't slip on the ramp
- ☐ the ramp will not bend
- ☐ the man and the heavy object are not accelerating as they travel up the ramp

Figure 2.4: List of assumptions for respondents to select from in order to solve the problem in the mechanics problem-solving diagnostic tool

The third and final question we focused on for the mechanics problem-solving diagnostic tool presented the respondent with three bins labeled “most important for evaluating problem-solving process”, “medium importance for evaluating problem-solving process”, “least/no importance for evaluating problem-solving process”. Respondents were asked to group a list of 12 criteria according to the importance for evaluating the problem-solving process of a student example of attempting the problem. The list of criteria and three bins is shown in Figure 2.5.

Similar to the previous problem, we focused on how the different expertise groups (novice, intermediates, and experts) differed in grouping the criteria.

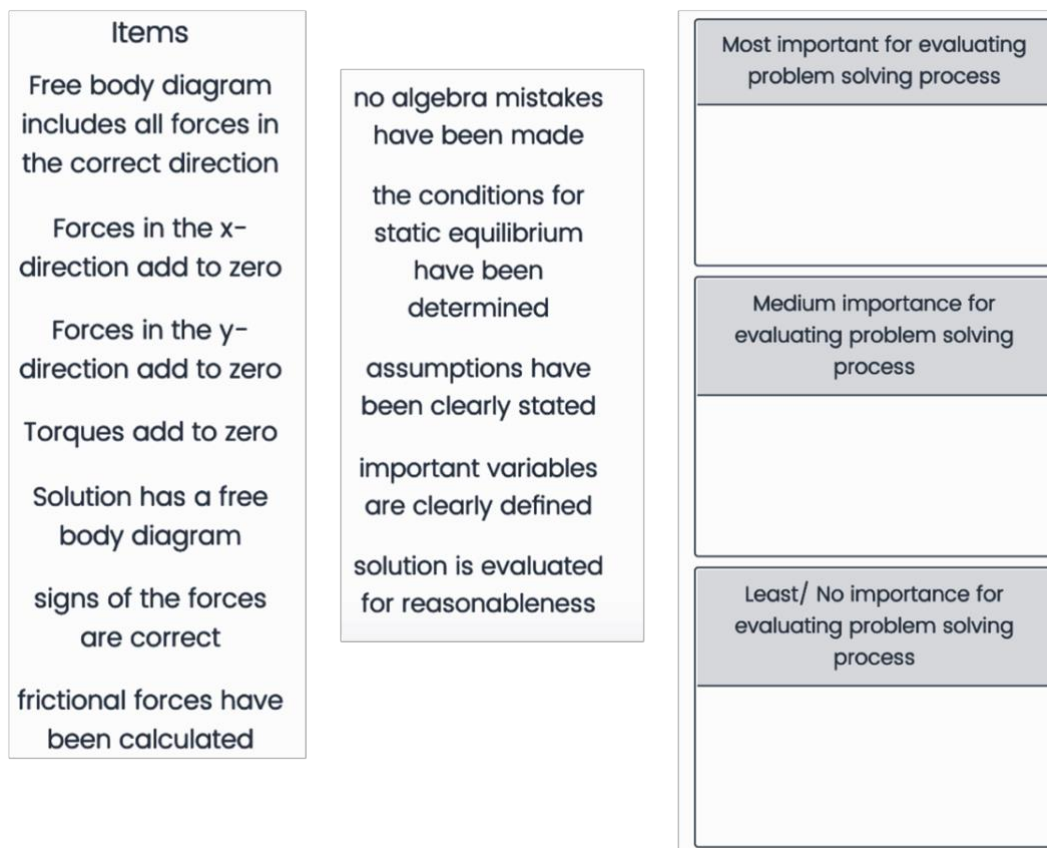


Figure 2.5: List of criteria for respondents to bin into three groups based on importance for evaluating the problem-solving process of a student example of attempting the problem in the mechanics problem-solving diagnostic tool

2.5.2 E&M diagnostic tool coding

For the E&M diagnostic problem-solving tool we have developed a set of codes for the first question where respondents are asked to devise a plan for how they would figure out the mystery circuit configuration and the values for its components. We followed a similar procedure as with the mechanics survey where we developed the codes through iterative rounds independently and then collaboratively. The codes were generated based on emergent features in

the data, and each expert independently coded for a subset of about twenty responses that were compared and discussed. Any differences in coding were discussed and resolved. The codes for this ‘devise a plan’ question consisted of 15 codes addressing both process-specific and content-specific features. These codes are represented in Table 2.3.

Table 2.3: Codes for ‘devise a plan’ problem in E&M problem-solving diagnostic tool

Process Codes	Content Codes
States quantities needed	Checks for battery
States assumptions	Uses voltmeter
References math or equation	Uses ammeter
Procedural versus holistic approach	Measures voltage
Detailed versus brief	Measures current
Evidence of simplifying problem	Determines or considers resistance
Acknowledges (infinitely) many possibilities	Connects battery
	Connects light bulb

In addition to the first question, we recorded whether respondents were able to successfully identify a plausible solution for the hidden mystery circuit given an option of four circuits as shown in Figure 2.6. We aimed to investigate whether there was a correlation between how respondents devised a plan for determining the hidden circuit and whether they select the correct answer. We decided to solely focus on the ‘devise a plan’ question and the

final question of selecting a possible solution because as students progressed through the survey and were progressively given more data, we found that many of them did not provide further insight as to how they would address solving the problem.

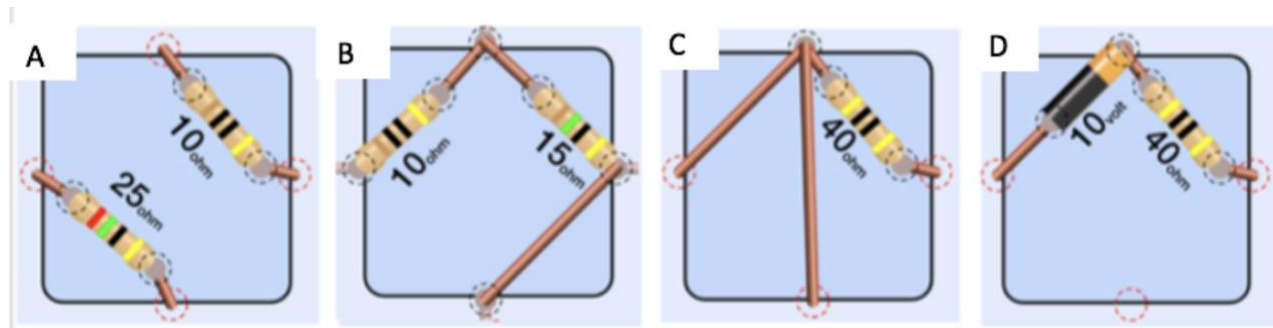


Figure 2.6: Four possible circuits solutions for the hidden mystery circuit in the E&M problem-solving diagnostic tool

2.5.3 Blended diagnostic tool coding

Though data is still being collected for the blended diagnostic tool, we have incorporated findings and feedback from the mechanics and E&M surveys into planning how we will code the responses for this new diagnostic tool. While problem-solving process cannot be completely separated from the context of the problem, we aim to investigate a respondent's problem-solving ability rather than their content knowledge. Similar to the previous two surveys, the blended survey begins by first asking the respondent to write a detailed plan as to how they would determine if the method of hanging the monitor specified in the problem stem would work. This open-ended question requires the respondent to make a number of problem-solving decisions. These decisions can be captured by a subset of Price et al.'s work characterizing 29 specific problem-solving decisions that science and engineering experts make in solving an authentic problem in their respective fields (2021). The subset of decisions that specifically apply to this 'devise a plan' question come from the first three categories in Price et al.'s work: Selection and

Goals of the Problem, Frame Problem, and Plan the Process for Solving. The specific decisions that can be applied in coding this question are decisions 3-15 and are summarized in Table 2.4. We plan to code for which of these decisions respondents make, but based on the data we may need to incorporate additional codes to fully capture features in the responses.

Table 2.4: Summary of problem-solving decisions developed by Price et al (2021) applicable for coding ‘devise a plan’ problem in blended problem-solving diagnostic tool

Category A: Selection and Goals of the Problem	3. Goals, criteria, constraints?
Category B: Frame Problem	4. Important features and info?
	5. What predictive framework?
	6. How to narrow down the problem?
	7. Related problems?
	8. Potential solutions?
Category C: Plan the Process for Solving	9. Is problem solvable?
	10. Approximations and simplifications to make?
	11. How to decompose into subproblems?
	12. Most difficult or uncertain areas?
	13. What info needed?
	14. Priorities?
	15. Specific plan for getting information?

The next question on the blended survey asks the respondent what information they would need to determine if the method of hanging the monitor specified in the problem stem would work. This question is probing one specific expert-level decision (13. What info needed?). We, therefore, plan to code this question based on expert consensus as to what

information they provide as necessary for solving the problem. We will compare novice and intermediary responses to this consensus to investigate how aligned they are with experts.

Similarly, the following question probes one specific expert-level decision by asking what assumptions or approximations would need to be made to solve the problem (10.

Approximations and simplifications to make?). We also plan to code this problem based on expert consensus of the assumptions or approximations that need to be made, and then we will compare novice and intermediary responses to the expert consensus. Additional codes to the subsequent survey questions will be developed following a similar procedure.

2.6 Results

2.6.1 Mechanics diagnostic tool results

The total counts by respondent physics expertise and devise plan codes for the mechanics diagnostic tool may be found in Table 2.5 and Table 2.6. Introductory physics students are categorized as ‘novices’, graduate students in physics are categorized as ‘intermediate’, and college-level physics instructors are categorized as ‘experts’. In Tables 2.5 and 2.6, percentages reported are the percentage of respondents coded as describing a particular strategy or technique to solve the problem. For example, among the 116 novice respondents, 25% of those stated assumptions in their plan for solving the problem. The devise plan codes are divided into two categories: process codes and content codes. We first address differences in expertise by process codes, then by content codes. The p-values noted in the tables indicate that the observed association between the codes and physics expertise is statistically significant, suggesting that novice, intermediate, and expert respondents in the population are not equally likely to engage in

the specific device plan codes indicated.

Table 2.5: Process code counts for respondents' approach to the 'devise a plan' question in the mechanics problem-solving diagnostic tool. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem.

Process codes	Novice (N = 116)	Intermediate (N = 16)	Expert (N = 6)
Holistic approach (versus procedural)*	47%	38%	100%
Detailed versus brief**	26%	38%	83%
States assumptions**	25%	50%	67%
States quantities needed	73%	75%	67%
Math or solving equations	64%	69%	67%
Free body diagram (or mention of drawing a diagram)	53%	44%	33%
Labels problem as a 'torque problem' or a 'statics problem'	34%	19%	0%
Evidence of simplifying the problem**	25%	50%	67%
Defines specific set of axes	12%	13%	17%
Defines what it means for ladder to slide***	23%	63%	67%

* $p \leq 0.05$

** $p \leq 0.01$

*** $p \leq 0.001$

Table 2.6: Content code counts for respondents' approach to the 'devise a plan' question in the mechanics problem-solving diagnostic tool. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem.

Content codes	Novice (N = 116)	Intermediate (N = 16)	Expert (N = 6)
Pivot	39%	44%	17%
Normal force*	32%	50%	67%
Torque	58%	44%	83%
Friction	77%	81%	83%
Static friction	42%	38%	33%
Work energy theorem	6%	13%	0%
Sum forces	31%	13%	83%
Separate problem components into x and y axes*	28%	6%	50%
Mentions constant velocity**	3%	6%	33%
Kinematics	1%	6%	0%
Newton's 2 nd Law	5%	0%	17%
Static equilibrium	20%	19%	17%

* $p \leq 0.05$

** $p \leq 0.01$

The process for devising a plan to solve the problem differed significantly in several ways based on respondents' physics expertise (see Figure 2.7). Experts were more likely to be more holistic and less procedural in their plan for solving the problem (Fisher's exact test, $p < 0.05$), and were more likely to devise a detailed plan compared to novices and intermediates

(Fisher's exact test, $p \leq 0.01$). For example, one novice respondent focused on the steps they would take to solve the problem in their response:

"I would solve this as a statics problem. First, I would draw a free body diagram and choose a pivot point. I would pick a pivot point based on where I have the most unknowns given the information in the problem. Then I would set the sum of the forces and the sum of the torque to 0. I would then solve for the max force of friction between the ramp and the ground to find the threshold point. From that, I would find the corresponding "r" to find the length."

In contrast, an expert respondent focused more on setting up the problem and addressing specific features that would be needed to solve the problem instead of just focusing on a procedure for solving it. They are also more detailed in their response:

"I assume that the ramp is massless, that there is enough friction between man and ramp so that the man doesn't slip, and that the man is strong enough to carry the load without tipping over. The man exerts a force on the ramp equal to the man's weight $m \vec{g}$. If the forces exerted by the surfaces on the ramp at the 2 end points are $\vec{N}_1$ (upper) and $\vec{N}_2$ (lower), then $N_{1y} + N_{2y} = m g$ and $N_{1x} + N_{2x} = 0$. Balancing the torque on the ramp about the lower end point (upper is also fine) gives $N_{1x} L \sin \theta + N_{1y} L \cos \theta = m g R \cos \theta$, where θ is the angle between ramp and ground, and R is the distance of the man's foot measured along the ramp from the ground. The friction limits give $N_{1y} < \mu_1 N_{1x}$ and $-N_{2x} < \mu_2 N_{2y}$, where μ_1 and μ_2 are the coefficients of friction at the two end points respectively. I solve the equations to eliminate N_{1y} , N_{2x} , and N_{2y} , and then enforce the constraints $N_{1x} > 0$, $N_{1y} > 0$, and $N_{2y} > 0$ along with the friction limits. However, since there isn't enough information to eliminate N_{1x} , it can always take on an appropriate value to prevent sliding whenever $1/\mu_2 < L/R(\tan \theta + \mu_1) - \mu_1$. Therefore if $1/\mu_2 < \tan \theta$, the ramp never slides. If $1/\mu_2 > \tan \theta$, the ramp will slide when the man reaches R^* such that $1/\mu_2 = L/R^*(\tan \theta + \mu_1) - \mu_1$."

There was also a statistically significant difference between respondents' expertise and whether they stated assumptions in their plan for solving the problem (Fisher's exact test, $p \leq 0.01$). We note, however, that the instructions for devising a plan specified for respondents to state their assumptions, so this may be more of a distinction between experts and intermediates better following the directions. Expert and intermediate respondents were also more likely to show evidence of simplifying the problem in their plan compared to novices (Fisher's exact test,

$p \leq 0.01$). Evidence of simplifying the problem included breaking the problem into more manageable subproblems or making decisions on how to make the problem more readily solvable. Furthermore, experts and intermediates were more likely to define what it means for the ladder to slide, whereas less than a fourth of novices addressed this (Fisher's exact test, $p \leq 0.001$).

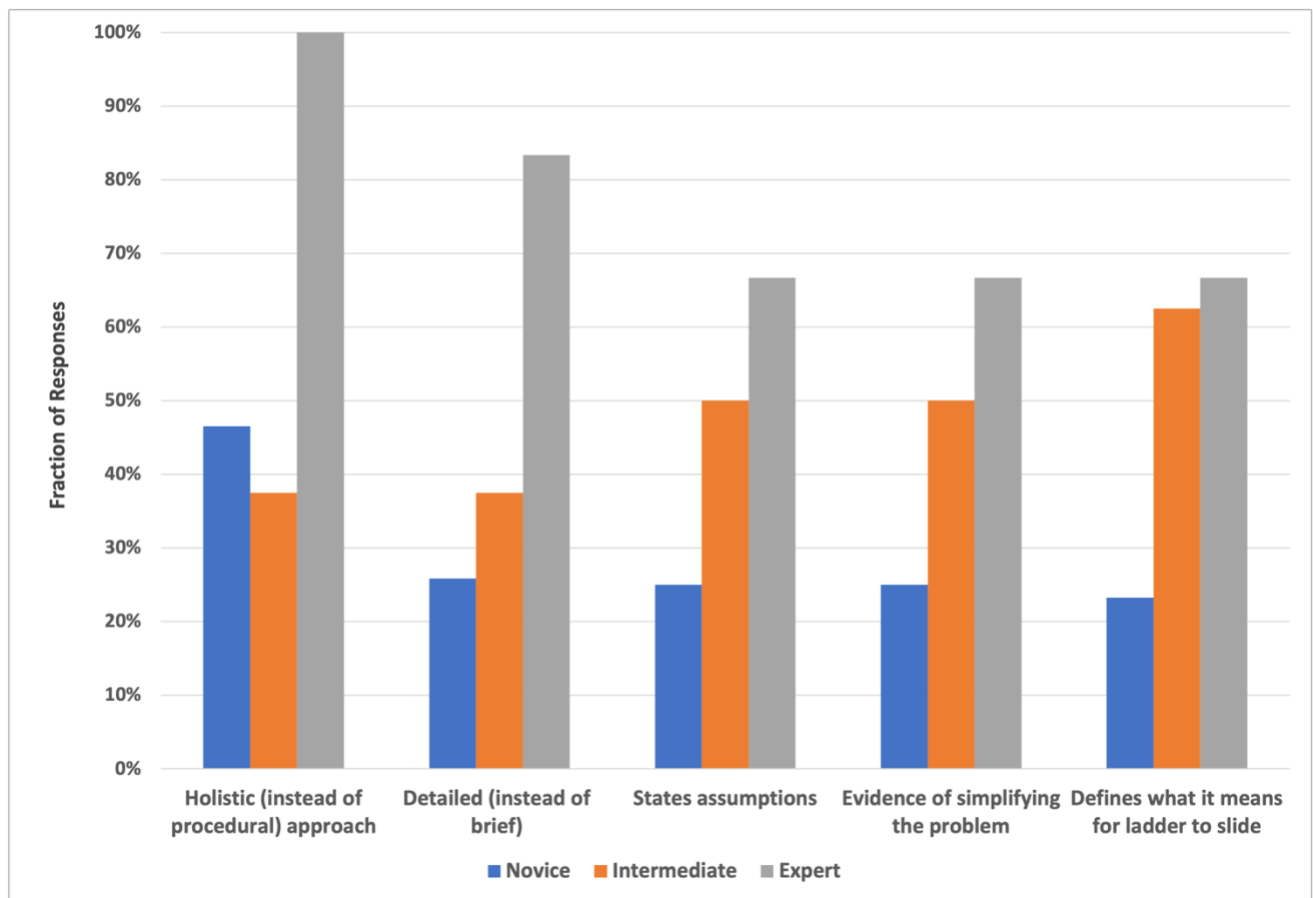


Figure 2.7: Respondents' process for devising a plan to solve the problem by physics expertise for the mechanics problem-solving diagnostic tool

There was also variation in the physics content respondents included in devising a plan to solve the problem by level of physics expertise (see Figure 2.8). Experts were more likely to address the normal force as relevant in the plan for solving the problem while only about one-

third of novices acknowledged this concept (Fisher's exact test, $p < 0.05$). Compared to intermediate respondents, experts and novices were more likely to describe separating the problem components into x and y axes (Fisher's exact test, $p < 0.05$). Experts were also more likely to mention having a constant velocity as part of their plan to solve the problem while very few novices and intermediates included this (Fisher's exact test, $p \leq 0.01$).

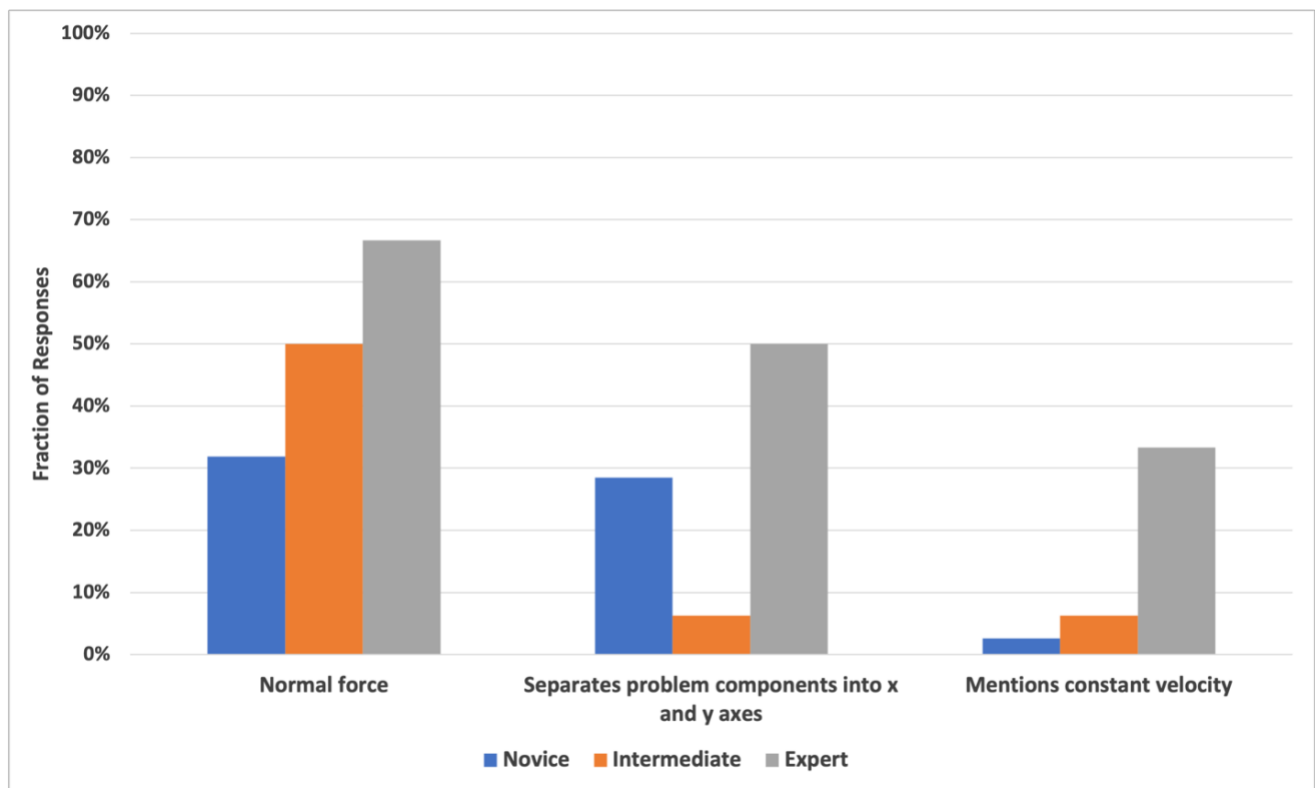


Figure 2.8: Respondents' physics content for devising a plan to solve the problem by physics expertise for the mechanics problem-solving diagnostic tool

In addition to investigating the differences in how respondents devise a plan for solving the problem, we compared the assumptions respondents selected out of a list of 14 (see Figure 2.4) that they would need to make in order to correctly solve the mechanics problem. The total counts by respondent physics expertise and the assumptions they selected as necessary for

solving the problem may be found in Table 2.7. The percentages reported are the percentage of respondents who selected a particular assumption as needed to correctly solve the problem. For example, out of the 114 novice respondents who answered this question, only 8% of them selected the assumption “the pivot point is at the center of mass of the ramp” as a necessary assumption for correctly solving the problem.

Table 2.7: Assumptions needed to solve the problem in the mechanics problem-solving diagnostic tool. Percentages represent the percentage of respondents that selected that assumption as being needed to correctly solve the problem

Assumptions	Novice (N = 114)	Intermediate (N = 16)	Expert (N = 5)
The pivot point is at the center of mass of the ramp	8%	13%	0%
The pivot point is at the top of the ramp	25%	19%	0%
The pivot point is at the bottom of the ramp	27%	38%	0%
The pivot point is at some point along the ramp	16%	13%	60%
The heavy object and the man are in the same position on the ramp	85%	75%	60%
The static friction forces prevent the ramp from sliding until the man gets far enough up the ramp	74%	63%	40%
The ramp will start to slide before the man reaches the truck	28%	31%	0%

Table 2.7 (Continued)

All the forces are initially balanced	67%	88%	40%
Friction between the ramp and the truck is negligible	14%	25%	0%
Friction between the ramp and the floor is negligible	11%	6%	0%
The truck won't roll during the process	93%	94%	100%
The man's feet won't slip on the ramp	83%	75%	100%
The ramp will not bend	90%	88%	100%
The man and the heavy object are not accelerating as they travel up the ramp	46%	69%	80%

When comparing the assumptions selected by respondent expertise, we found no statistically significant difference among respondents. However, as there were possible ambiguities in what assumptions would be necessary for solving the problem, we then focused on the assumptions that experts agreed were needed. These assumptions that at least 4 out of 5 experts agreed upon include “the truck won’t roll during the process”, “the man’s feet won’t slip on the ramp”, “the ramp will not bend”, and “the man and the heavy object are not accelerating as they travel up the ramp”. While there was no statistically significant difference between respondent expertise and whether these four assumptions were selected, we did find that novice and intermediate respondents tend to select additional assumptions which experts do not agree on including “the heavy object and the man are in the same position on the ramp”, the static

friction forces prevent the ramp from sliding until the man gets far enough up the ramp”, and “all the forces are initially balanced”.

The third and final question we analyzed in the mechanics problem-solving diagnostic asked respondents to group a list of 12 criteria according to the importance for evaluating the problem-solving process of a student example attempting the problem. The three categories respondents could group the 12 criteria in are “most important for evaluating problem-solving process”, “medium importance for evaluating problem-solving process”, and “least/no importance for evaluating problem-solving process”. The total counts by respondent physics expertise and how they ranked specific criteria for evaluating the problem-solving process may be found in Table 2.8. The percentages reported are the percentage of respondents who ranked a particular criterion in each category (most important, medium importance, or least/no importance). For example, out of the 114 novice respondents who answered this question, 77% of them ranked the criteria “free body diagram includes all forces in the correct direction” in the group as “most important” for evaluating the problem-solving process. The p-values noted in the table indicate that the observed association between the criteria ranking and physics expertise is statistically significant, suggesting that novice, intermediate, and expert respondents in the population are not equally likely to rank certain criteria for evaluating process as the same level of importance.

Table 2.8: Criteria for evaluating the problem-solving process in the mechanics problem-solving diagnostic tool. Percentages reported are the percentage of respondents who ranked a particular criterion in each category (most important, medium importance, or least/no importance)

Criteria for evaluating process	Ranking	Novice (N = 114)	Intermediate (N = 16)	Expert (N = 9)
Free body diagram includes all forces in the correct direction***	Most important	77%	100%	25%
	Medium importance	13%	0%	50%
	Least/no importance	2%	0%	25%
Assumptions have been clearly stated	Most important	62%	75%	50%
	Medium importance	25%	19%	50%
	Least/no importance	4%	6%	0%
Solution has a free body diagram	Most important	61%	69%	50%
	Medium importance	28%	25%	50%
	Least/no importance	4%	6%	0%
Forces in the x-direction add to zero	Most important	68%	75%	50%
	Medium importance	20%	25%	50%
	Least/no importance	2%	0%	0%
Forces in the y-direction add to zero	Most important	53%	38%	63%
	Medium importance	25%	19%	12%
	Least/no importance	13%	44%	25%
Frictional forces have been calculated	Most important	22%	44%	25%
	Medium importance	54%	50%	50%
	Least/no importance	14%	6%	25%
No algebra mistakes have been made	Most important	28%	38%	12%
	Medium importance	47%	56%	75%
	Least/no importance	13%	6%	12%
Signs of the forces are correct	Most important	10%	0%	12%
	Medium importance	24%	44%	12%
	Least/no importance	54%	56%	75%
Solution is evaluated for reasonableness	Most important	62%	69%	63%
	Medium importance	22%	25%	37%

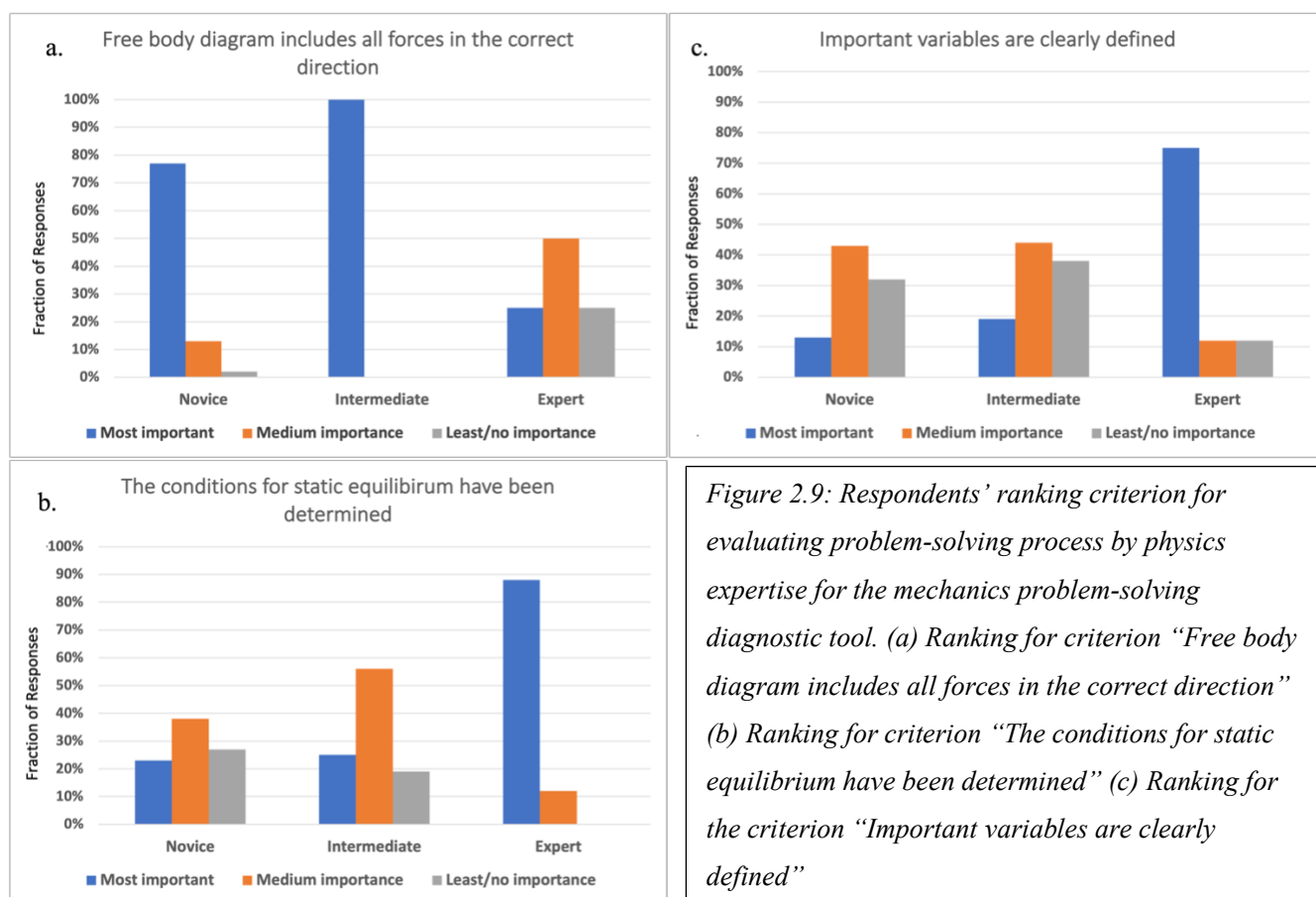
Table 2.8 (Continued)

	Least/no importance	4%	6%	0%
The conditions for static equilibrium have been determined**	Most important	23%	25%	88%
	Medium importance	38%	56%	12%
	Least/no importance	27%	19%	0%
Torques add to zero	Most important	40%	69%	75%
	Medium importance	36%	31%	25%
	Least/no importance	11%	0%	0%
Important variables are clearly defined**	Most important	13%	19%	75%
	Medium importance	43%	44%	12%
	Least/no importance	32%	38%	12%

** $p \leq 0.01$

*** $p \leq 0.001$

How respondents ranked criteria for evaluating the problem-solving process of a student example attempting the problem differed significantly in three ways based on the respondent's physics expertise (see Figure 2.9). Novice and intermediate respondents were more likely to rank the criterion "free body diagram includes all forces in the correct direction" as "most important for evaluating problem-solving process" while experts were more likely to rank this criterion as "medium importance for evaluating problem-solving process" (Fisher's exact test, $p < 0.001$, see Figure 2.9a). On the other hand, experts were more likely to rank the criterion "the conditions for static equilibrium have been determined" as most important for evaluating the problem-solving process while novice and intermediate respondents were more likely to rank this as only medium importance (Fisher's exact test, $p < 0.01$, see Figure 2.9b). Furthermore, experts were more likely to rank the criterion "important variables are clearly defined" as most important for evaluating the process while novices and intermediates tended to rank this as either medium or least/no importance (Fisher's exact test, $p < 0.01$, see Figure 2.9c).



2.6.2 E&M diagnostic tool results

The total counts by respondent physics expertise and devise plan codes for the E&M diagnostic tool may be found in Table 2.9 and Table 2.10. Introductory physics students are categorized as 'novices', and because there were only three advanced graduate students in physics that took the survey, they are combined with the college-level physics instructors who are categorized as 'experts'. As with the devise plan findings from the mechanics survey, percentages reported are the percentage of respondents coded as describing a particular strategy or technique to solve the problem. For example, among the 88 novice respondents, 33% of those stated assumptions in their plan for solving the problem. The devise plan codes are divided into two categories: process codes and content codes. We first address differences in expertise by

process codes and content codes, and then we present differences in process and content codes by whether novices selected the correct solution for the hidden mystery circuit. The p-values noted in the table indicate that the observed association between the codes and physics expertise is statistically significant, suggesting that novice and expert respondents in the population are not equally likely to engage in the specific device plan codes indicated.

Table 2.9: Process code counts for respondents' approach to the 'devise a plan' question in the E&M problem-solving diagnostic tool. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem.

Process Codes	Novice (N = 88)	Expert (N = 9)
States quantities needed	81%	89%
States assumptions	33%	56%
References math or equations	18%	33%
Holistic approach (versus procedural)	50%	44%
Detailed (versus brief)*	18%	56%
Evidence of simplifying problem**	9%	56%
Acknowledges (infinitely) many possibilities	0%	11%

* $p \leq 0.05$

** $p \leq 0.01$

Table 2.10: Content code counts for respondents' approach to the 'devise a plan' question in the E&M problem-solving diagnostic tool. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem.

Content Codes	Novice (N = 88)	Expert (N = 9)
Checks for battery	28%	56%
Uses voltmeter	60%	78%
Uses ammeter	49%	67%
Measures voltage	67%	89%
Measures current	61%	78%
Determines or considers resistance	47%	78%
Connects battery	41%	78%
Connects light bulb	38%	11%

The process for devising a plan to solve the problem differed significantly in only two ways based on respondents' physics expertise (see Figure 2.10). Similar to the findings from the mechanics survey, experts were more likely to be more detailed in their plan for solving the problem while most novices gave a brief description of their plan (Fisher's exact test, $p < 0.05$). Experts were also more likely to show evidence of simplifying the problem in their plan for solving the problem whereas less than 10% of novices showed evidence of simplifying the problem in their plan (Fisher's exact test, $p \leq 0.001$). Unlike the mechanics survey, there was no variation in physics content respondents included in devising a plan to solve the problem for the E&M survey by level of expertise.

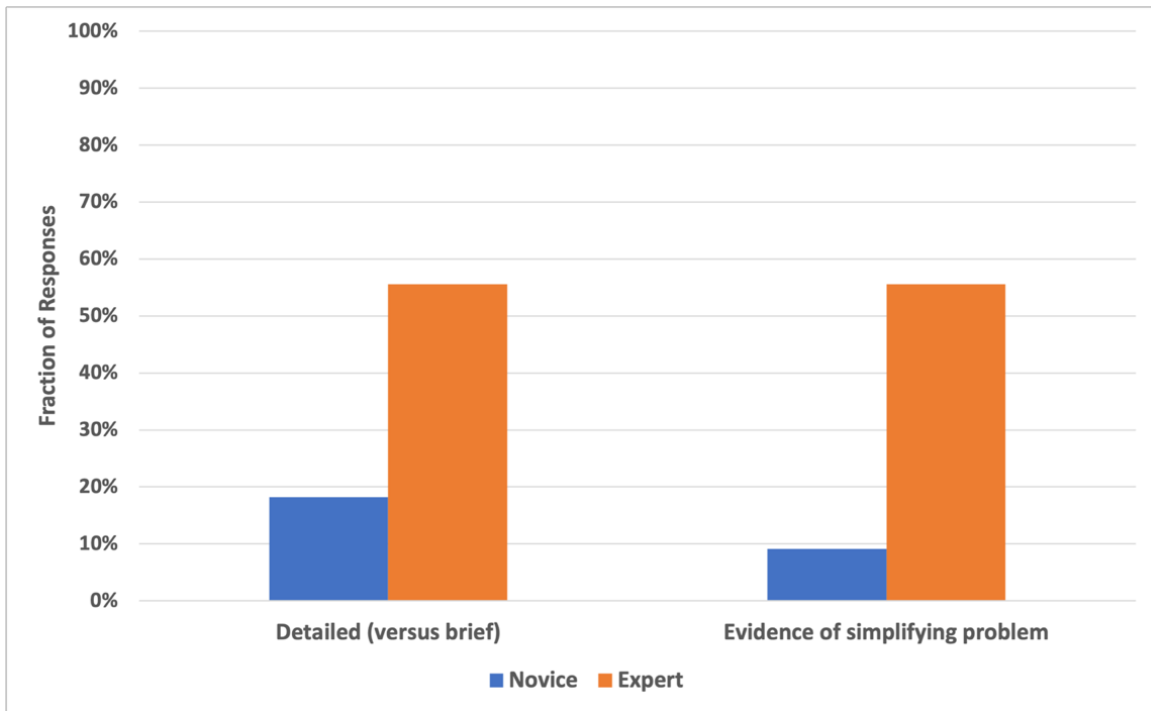


Figure 2.10: Respondents' process for devising a plan to solve the problem by physics expertise for the E&M problem-solving diagnostic tool

In addition to investigating process and content codes by expertise, we also investigated these codes by whether novices selected the correct circuit as a possible solution to the hidden mystery circuit. The total counts for novices by correct solution and devise plan codes for the E&M diagnostic tool may be found in Table 2.11 and Table 2.12. Only 25 novices out of the 88 who took the survey selected the correct solution while the majority of 55 did not. Eight of the novices did not answer this question. The p-values noted in the table indicate that the observed association between the devise plan codes and whether novice students selected the correct solution is statistically significant, suggesting that novice respondents who answered the problem correctly and those who did not in the population are not equally likely to engage in the specific device plan codes indicated.

Table 2.11: Process code counts for novices' approach to the 'devise a plan' question in the E&M problem-solving diagnostic tool by correct solution. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem.

Process Codes	Correct answer (N = 25)	Incorrect answer (N = 55)
States quantities needed	80%	84%
States assumptions	28%	36%
References math or equations*	4%	25%
Holistic approach (versus procedural)	32%	58%
Detailed (versus brief)	12%	24%
Evidence of simplifying problem	8%	9%
Acknowledges (infinitely) many possibilities	0%	0%

* $p \leq 0.05$

Table 2.12: Content code counts for novices' approach to the 'devise a plan' question in the E&M problem-solving diagnostic tool by correct solution. Percentages represent the percentage of respondents describing a particular strategy or technique to solve the problem.

Content Codes	Correct answer (N = 25)	Incorrect answer (N = 55)
Checks for battery	28%	31%
Uses voltmeter	44%	69%
Uses ammeter	36%	58%
Measures voltage	72%	67%
Measures current	64%	62%
Determines or considers resistance	44%	51%
Connects battery	28%	49%
Connects light bulb	32%	40%

The process for novices devising a plan to solve the problem differed significantly in only one way based on whether the correct solution was selected (see Figure 2.11). Novices who chose an incorrect solution for the hidden mystery circuit were more likely to reference using math or specific equations in their plan for solving the problem (Fisher's exact test, $p < 0.05$). However, only 25% of novices who selected an incorrect solution mentioned this in their plan. Surprisingly, there was no variation in physics content novices included in devising a plan to solve the problem for the E&M survey by whether they selected the correct possible solution or not.

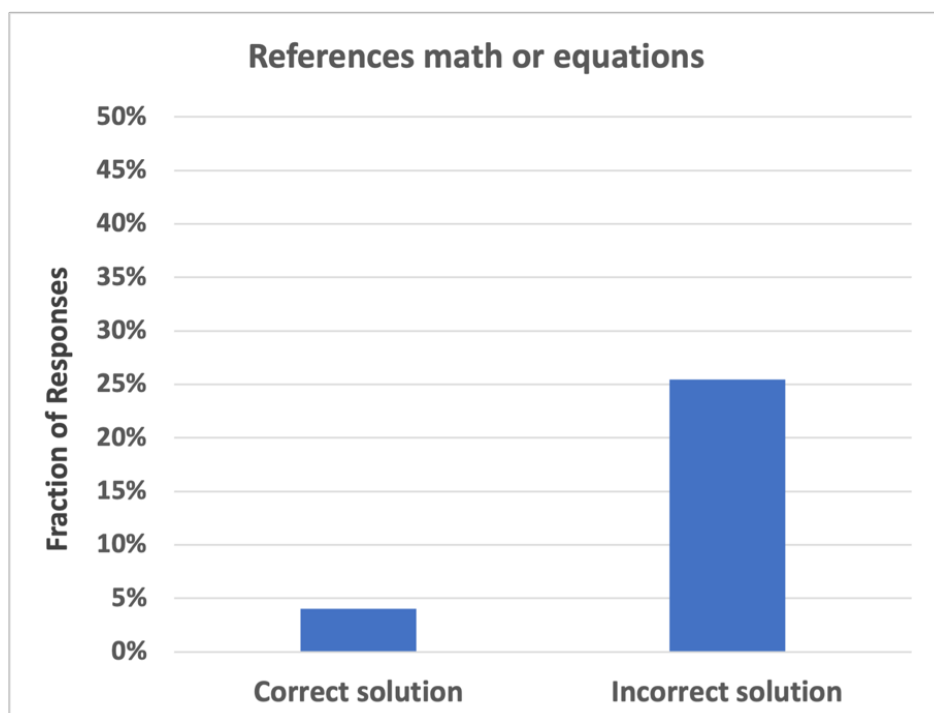


Figure 2.11: Novices referencing math or equations in devising a plan to solve the problem by whether they selected the correct possible solution for the E&M problem-solving diagnostic tool

2.7 Discussion

There are several notable results from the analysis of how respondents of varying physics expertise devise a plan to solve an authentic problem, many of which align with expectations. First, we found features in which novices differ from experts in their process for devising a plan to solve a problem for both the mechanics and E&M problem-solving diagnostic tools. Unsurprisingly, experts are more likely to be more detailed in their descriptions of how they would solve the problems, and they also show evidence of simplifying the problem in their plans. For the mechanics survey, experts were less procedural in their approach compared to novices. This may be a result of how students are instructed to systematically solve problems involving static equilibrium and that experts consider the full scenario before focusing on one specific

procedure to solve the problem. Experts were also more likely to address assumptions they would make to solve this problem in the mechanics survey. However, the problem instructions specify that respondents should state their assumptions in their response, so this may just be an example of experts better following instructions than novices.

We found few differences in the physics content respondents included in devising a plan to solve the problem by level of physics expertise. For the mechanics survey, experts were more likely to address normal force, separate the problem components into x and y axes, and mention constant velocity than novices. While experts are more familiar with the content knowledge for solving such problems, it is possible they were specifically more likely to include these three features in their plan because they also lead to making the problem more readily solvable. Novices may be unable to appropriately leverage their physics understanding to simplify problems. Interestingly there were no differences in physics content respondents included for the E&M survey. This may be a result of novices being familiar enough with the content knowledge needed to solve the problem, regardless of their ability to apply their knowledge to successfully solve the problem.

Furthermore, we investigated the relationship between how novices devised a plan for solving the problem in the E&M survey and whether they selected the correct circuit as a possible solution to the hidden mystery circuit. We found no differences between whether novices choose the correct solution and how they devised a plan besides whether they referenced using math or specific equations. Novices who selected an incorrect solution were more likely to reference math or equations in their devise plan, though this was still only about a fourth of students who selected the wrong answer. Overall, the problem presented in the E&M survey seemed to be challenging for novices both in devising a plan for solving the problem and in

selecting the correct possible solution. How novices initially devised a plan did not lead them to successfully solve the problem. However, it is important for future work to investigate how novices deal with new information or data and whether they are able to reflect and update their plan for solving the problem.

In the mechanics problem-solving diagnostic tool we also investigated the differences in assumptions respondents would need to make to solve the problem and how respondents would rank a set of criteria for evaluating the process for solving the problem by respondent physics expertise. While there were no statistically significant differences in the assumptions respondents selected as necessary for solving the problem by expertise, we did find that novice and intermediate respondents tend to indicate additional assumptions that experts did not agree upon as needed. This may be because novices and some intermediates struggle with focusing on what decisions are actually necessary for simplifying the problem. In analyzing how respondents rank a set of criteria for evaluating the problem-solving process of a student worked example, we found that experts tend to prioritize criterion which focuses on defining conditions and variables whereas novices and intermediates tend to focus more on correctly displaying or visualizing the problem. Accurate problem representation is a particularly helpful strategy for more novice problem-solvers, so this may be why this criterion was prioritized by the non-expert groups.

There is an interesting distinction between the performance of the intermediate respondents for the devise plan question and the assumption selection and ranking criteria questions in the mechanics problem-solving survey. For the devise plan question, the percentage of intermediate responses for their process and content approach to solving the problem tends to fall between that of the novice and expert responses. That is, they are more adept at devising a plan for solving an authentic problem than novices, yet they are not as advanced as experts,

unsurprisingly. However, for the assumption selection and criteria ranking questions, intermediate responses tend to be more closely aligned with novice responses. This is possibly a result of these questions being more challenging, yet it also highlights the nonlinearity of problem-solving ability. While intermediate respondents' performance aligns with expectations in some cases, they do not in others. Thus, it is still unclear the timeline and nature of experiences needed to develop expert-like problem-solving skills.

2.8 Conclusions, limitations, and future work

The analysis of the mechanics and the E&M problem-solving diagnostic tools provide insight into the differences between respondents' physics expertise and their problem-solving ability as well as ways to improve the measurement of these skills. While the differences we found between novice and expert problem-solving abilities align with expectations, future surveys, including the blended problem-solving diagnostic tool, should be structured to better elicit the problem-solving process. The difference in content knowledge between novices and experts is unsurprising and does not provide any insight into their ability to approach an authentic problem. Thus, the blended survey aims to investigate what problem-solving decisions respondents make to solve the problem rather than the physics content knowledge in those decisions. Moreover, this survey gives the respondent opportunities to re-answer questions after being provided with additional information, thus probing how they deal with new data and their ability to reflect. Findings and feedback from the mechanics and E&M surveys have also informed how we will interpret future results.

To better investigate physics problem-solving ability, future data collection will also include a wider range of respondents. In the two previous surveys, the expert respondents were

current college-level introductory physics instructors. This may have influenced how they respond to the questions in that they are possibly framing their answers as model answers for what they would want their students to produce. It would thus be beneficial to include other physics experts who are not teaching introductory physics courses to get a more complete representation of expertise. Likewise, it will be useful to include a range of physics graduate students by year and teaching experience to better investigate the progression from novice to expert-level problem-solving ability. To better understand how an educational setting can affect the development of problem-solving ability, the blended problem-solving diagnostic tool will also be implemented in two separate Harvard introductory physics courses, one of which is a problem- and project-based course.

From our findings and feedback from the mechanics and E&M problem-solving diagnostic tools, we have developed an improved survey better focused on investigating problem-solving ability. Furthermore, by surveying a greater range of respondents, we are better equipped to measure the differences by expertise. While problem solving is challenging for students, by better understanding what problem-solving decisions they struggle to make, we can better support these areas to improve problem-solving ability in an authentic context.

Chapter 3

Measuring and teaching problem-solving practices using an interactive task in PhET simulation

3.1 Introduction

It is recognized that science and engineering education needs to move beyond didactic instruction of knowledge and help students acquire the practices used by scientists and engineers to solve complex real-world problems (NGSS, 2013; Carr, Bennett IV, & Strobel, 2012; Krajcik, 2015). That is, science and engineering education needs to shift away from the traditional teaching method in which an instructor delivers information and knowledge to the students via a one-way flow of information where students are expected to absorb and memorize the information provided. Instead, students should be encouraged to engage in hands-on learning activities, including experiments, design projects, and problem-solving tasks, which allow them to apply the learned concepts in an authentic context. However, one barrier to students learning how to solve authentic problems stems from the overemphasis on efficiency-oriented instructional practices in our education systems (Schwartz, Bransford, & Sears, 2005). Such

instructional practices are designed to optimize the learning process and improve student outcomes by focusing on efficiency and productivity. Typical tasks are designed to assess students for efficient retrieval of facts and procedures rather than systematic, in-depth practices of scientific inquiry and problem solving (Pellegrino, 2014). In this study, we examine how a brief task embedded in an interactive circuit simulation can be used to elicit students' problem-solving performance and how patterns of interactions from the log data are related to different practices adopted to solve the problem. Furthermore, we explore if a worked example (in the form of the E&M problem-solving survey described in Chapter 2) helps students learn problem-solving practices. We asked:

1. What practices are involved in solving complex, authentic problems in science and engineering domains?
2. Can worked examples help students learn these practices?

3.2 Theoretical Framework

Solving complex, novel problems involves a set of multifaceted and interdependent practices with different underlying cognitive processes. However, which aspects of these practices students can learn through worked examples remain a question of research. A problem-solving task embedded in a simulation makes it possible to capture the processes that students go through to solve the problem in the form of log data. This allows us to compare the problem-solving outcomes and the processes that students go through to determine whether and how students' problem-solving trajectories may differ depending on if they have first gone through a partial-worked example demonstrating how to solve the problem first. In this section, we first address features of worked examples and how they can be used to assist learners. We

then highlight how interactive simulations can be used to engage students in problem-solving tasks.

3.2.1 Worked examples as a learning mechanism

A worked example is a “step-by-step demonstration of how to perform a task or how to solve a problem (Clark, Nguyen, & Sweller, 2006). In cognitive psychology, given that a student has studied a problem and its appropriate solution steps, worked examples are accredited for reducing cognitive load and leading to more efficient learning (Sweller, 2006). By providing learners with the worked-out solution steps needed to reach the solution, worked examples allow learners to focus on a specific problem state and its associated solution. This then allows learners to allocate their available cognitive resources to construct solution schemas for specific problem types using the borrowing and reorganizing and the information store principles. However, when dealing with problem-solving tasks, learners need to be involved in cognitively costly search processes in working memory for suitable solution steps using the randomness as genesis and narrow limits of change principles, which may take up all their available cognitive resources (Sweller, 2006). Thus, it is argued that novices learning following a worked example compared to a problem-solving task is dramatically increased because of the reduction of random processes (Sweller, 2006).

Previous studies have demonstrated the benefits of worked examples in learning a variety of subjects (Sweller & Cooper, 1985; Paas, 1992; Bokosmaty et al., 2015). These studies used well-structured problems which have a “well-defined initial state, a known goal state, and constrained set of logical operators” (Jonassen, 1997). Such problems have been described as requiring skills that can be transferred only to similar types of problems. It is thus less

conclusive regarding how worked examples may facilitate learning in complex and novel problem-solving tasks. In one study using circuit troubleshooting problems, Van Gog, Kester & Paas (2011) found that high school students performed significantly better on a test task if they first engaged in a worked-example task before a problem-solving task compared to students who engaged in the tasks in the opposite order. The students who engaged with the worked-example task first were also described to have reached this higher performance on the test task with lower investment of mental effort during the training. In contrast, a study by Schwartz and Martin (2004) found that high school students who were asked to invent a solution to a statistics problem followed by viewing a worked example performed twice as well as those who received a lecture and a worked example instead.

3.2.2 Interactive simulations to engage in problem-solving tasks

In educational settings, simulations are interactive multimedia representations of dynamic models and processes that allow students to manipulate and modify parameters and observe outcomes (de Jong & van Joolingen, 1998; Honey & Hilton, 2011). Simulations are well-suited for discovery learning and can motivate students through challenges and experimentation, yet they are largely underutilized to engage students in problem-solving tasks with real-world contexts (Liu et al., 2011). The open-ended nature of the task environment affords the students the agency to make various decisions along the problem-solving process, while the large amounts of log data recording students' interactions provide potentially key evidence for evaluating the practices adopted to solve the problem (Fischer et al., 2022).

In order to characterize and measure in an interactive simulation the practices used by scientists and engineers to solve authentic, complex problems, we must first elucidate the

characteristics of authentic problems in science and engineering domains. Authentic problems require the application of domain knowledge. The problem solvers need to decide what relevant knowledge or models to apply. These problems are also open-ended and have limited data provided upfront. Problem solvers must decide what data they need to collect as well as how to collect this data. Furthermore, authentic problems are complex, involve multiple steps, require planning, and cannot be solved following a set of rules. Problem solvers need to decide what are the criteria for a successful solution (Salehi, 2018). Previous work by Wang et al. (2021), has shown that simulation-based tasks capture evidence of specific problem-solving practices including decomposition, data collection, data recording, and reflection. Similarly, in this study involving a task embedded in a simulation, we aim to capture these problem-solving practices, specifically how students break down the problem into smaller ones, whether they systematically collect the data needed, whether they keep track of the data collected and whether they can apply their domain knowledge to interpret the data collected.

3.3 Materials

In this study, we investigate students' problem-solving performance using a black-box problem embedded in the PhET Circuit Construction Kit simulation (<https://phet.colorado.edu/en/simulation/circuit-construction-kit-dc>). The problem involves a circuit configuration hidden behind a black box with four wires (terminals) protruding from the box. Students can interact with the black-box configuration by connecting wires, batteries, light bulbs, and resistors to the terminals as well as using the ammeter and voltmeter to take measurements to figure out the hidden circuit (Figure 3.1).

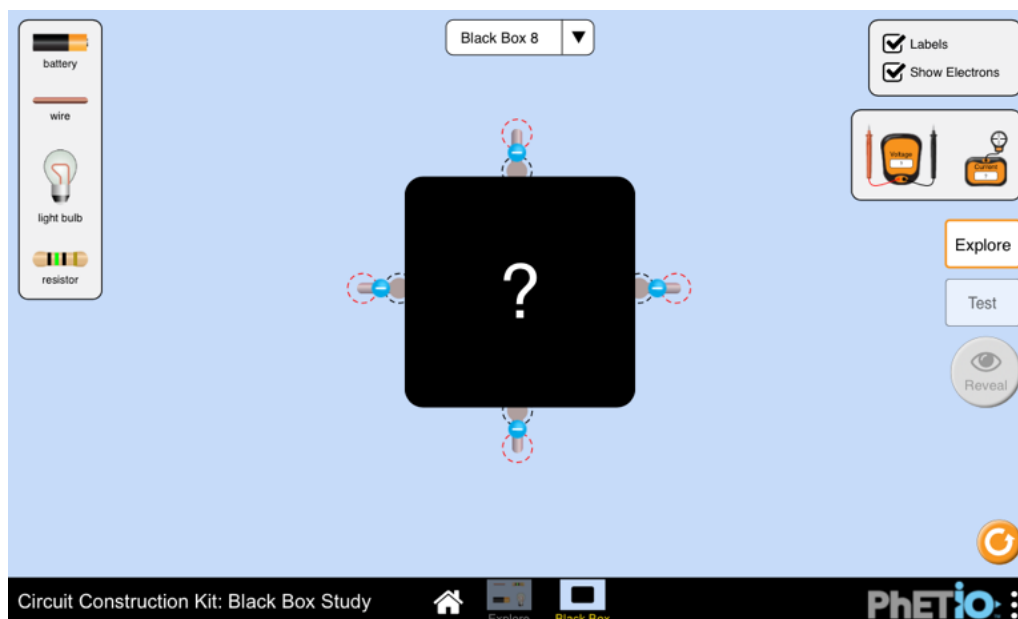


Figure 3.1: The black box problem user interface

Solving the problem requires basic knowledge of electric components and measurement tools as well as Ohm's Law ($V = IR$). Students must also be able to decompose a complex circuit into simpler sub-parts, collect data, and keep track of the collected data to achieve a meaningful solution.

The black-box problem was presented to students in two formats: a guided written-response format (the E&M survey presented in Chapter 2) and an interactive simulation format. Both formats involve inferring a hidden circuit configuration (Figure 3.2), including how the four wires are connected, what components are in the hidden circuit, and the values of the components.

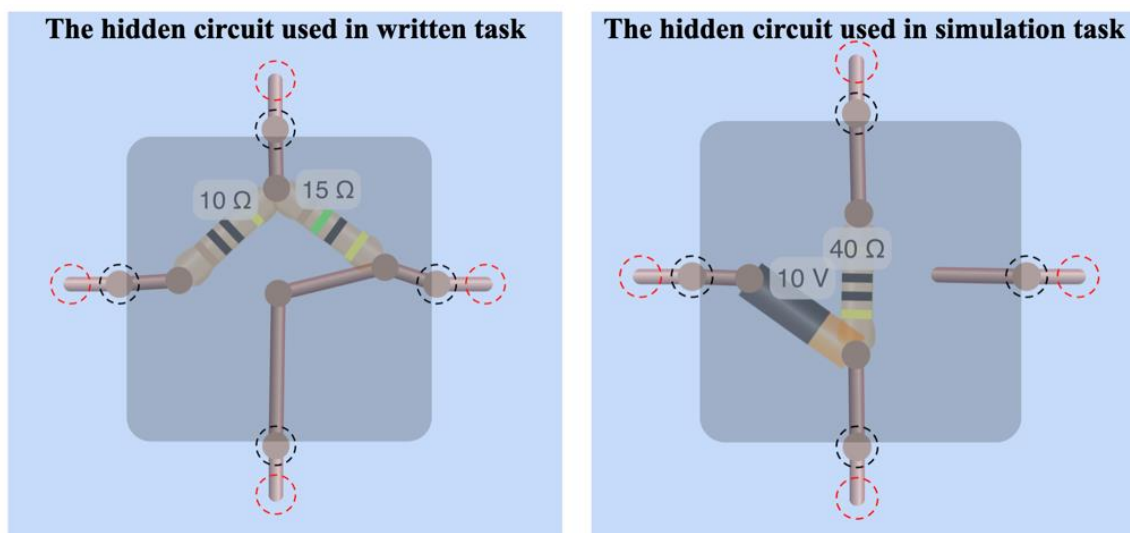


Figure 3.2: The hidden circuits behind the black-box problems used in the written-response (left) and simulation (right) tasks

The two formats differ in one key dimension: the amount of data provided for solving the problem. The guided written-response format serves as a partial-worked example: students are asked to first think about their general approach for solving the problem, then they are presented with progressively more data in the form of step-by-step screencasts from the solution process (Figure 3.3). Students are shown how a circuit can be constructed around the black box and how the voltmeter and ammeter are used to collect the data along with a table of the collected data. In their written responses, students are prompted to describe tentative solutions based on the data provided. Students are next presented with a different circuit constructed around the black box with more data to represent the solution process. After four sets of step-by-step screencasts from the solution process, students are then asked to choose the correct circuit from four options.

In the simulation task, the students need to decide what data to collect and how to collect the data by interacting with the simulation. As a result, there are notable features that distinguish the guided written-response task from the simulation task. In the guided written-response task,

the decomposition and data collection practices are provided to the students. The students are presented with simple circuits constructed around the black box with the voltmeter and ammeter connected to collect data. Furthermore, data recording practices are demonstrated to the student in the form of tables with the voltage and current measurements between each terminal.

However, the guided written-response task does not justify the data collection decisions made, like what circuits to construct around the black-box configuration and why use the voltmeter and ammeter. This task also does not include data interpretation practices, such as how Ohm's law can be used to solve the problem. Lastly, the task does not incorporate problem definitions and reflection practices. We hypothesized that students who go through the written-response task first may pick up important cues on effective problem-solving practices, such as decomposition, data collection, and data recording practices, which would help them perform better in the subsequent simulation task.

You connect a 9 V battery between protruding wires 1 and 2. You place a voltmeter between pairs of protruding wires to measure the voltage. You use the ammeter to measure the current flowing through each of the protruding wires. The data you collect is presented in Table 2 (below).

The diagram below shows how you connect the voltmeter to measure the voltage between wires 1 and 2 and how you connect the ammeter to measure the current in wire 2.

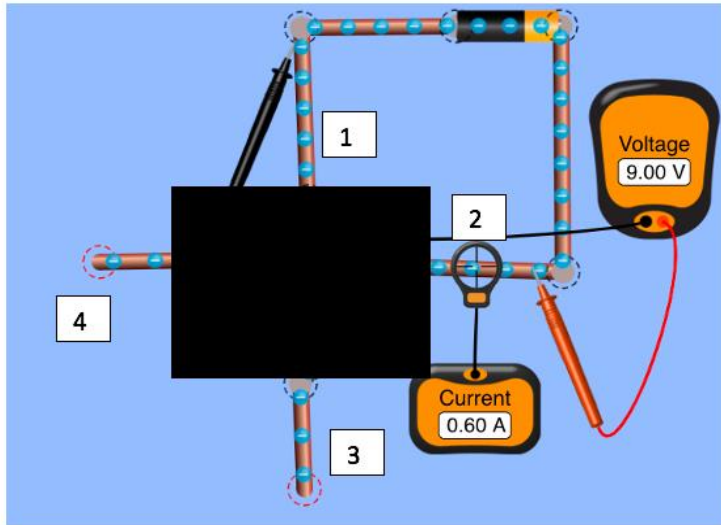


Table 2

Wires	Voltage (V)
1 & 2	9
2 & 3	0
3 & 4	0
4 & 1	0
1 & 3	9
2 & 4	0

Wire	Current (A)
1	0.6
2	0.6
3	0
4	0

Based on these new data what can you conclude? Be specific.

Figure 3.3: A data collection example presented to students in the written-response task

3.4 Method

For this study, 83 students were recruited from an introductory college-level physics course in electromagnetism from Harvard Extension School. These students were mostly pre-med concentrators taking the course as a requirement for their concentration and not as an elective. The majority of the participants (83%) were college students along with some non-physics graduate students (15%) and high school teachers (2%). Students in this study had already covered Ohm's law and DC circuits in the course, and this study was used as an optional review for their final. They received course credit for participating in the study, and they were able to do so online at a time and place of their convenience. The participants were randomly assigned to one of the two orders of the study: guided written-response task first and simulation task second ($n = 43$) or just the unguided, simulation task ($n = 40$) (see Fig 3.4). This is to allow us to investigate to what extent the partial-worked example can help students learn authentic problem-solving procedures. The same problem statement was used in both tasks as shown in Figure 3.5.

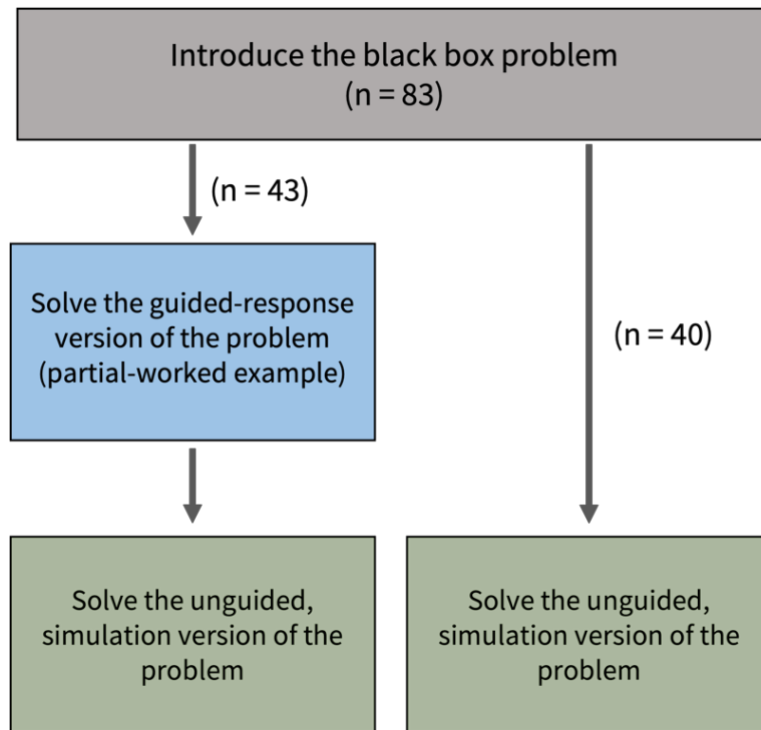
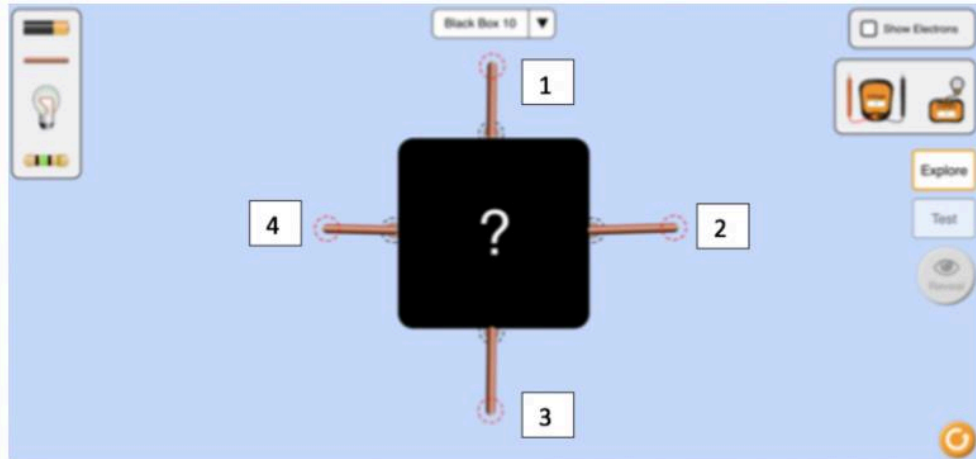
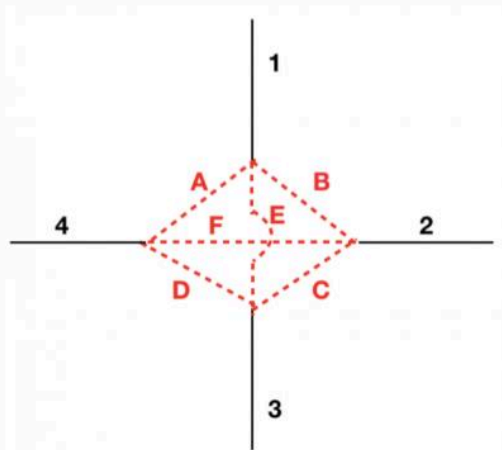


Figure 3.4: Study design for black box problem: $n = 43$ participants assigned to the guided written-response task first and the simulation task second, $n = 40$ assigned to just the unguided, simulation task

The black box hides a mystery electrical configuration. Your job is to figure out a possible electrical configuration that is hidden behind the black box by collecting data from the four wires protruding from the box. The mystery circuit could be comprised of the following components: wires, batteries, resistors. The circuit does not necessarily have all 3 types of components.



Each branch (depicted below as A, B, C, D, E, F) could be a wire, a battery, a resistor, or nothing. In other words, **you could place at most 1 circuit element directly between any 2 wires.** It is OK to connect a battery to the wires without any resistance.



You will be given the following items to use to figure out the mystery circuit:

- wires (as many as you need)
- ammeter
- voltmeter
- resistors (you can use as many resistors as you need)
- 9V battery
- light bulb

Figure 3.5: The problem statement used in both the simulation and written-response tasks

In the guided written-response task, students followed along with the worked example for how to collect and record data, and they proposed tentative solutions in their written responses. The final section of the written-response task asks students to choose the correct hidden circuit from four options. In the unguided, simulation task, students were instructed to work on the problem for a maximum of fifteen minutes and then draw a diagram representing the hidden circuit as their proposed solution. Students then uploaded their diagrams to the online survey and answered a question about whether and how they used any physics formula when solving.

3.5 Results

A total of 83 students completed this study, 43 of whom had gone through the guided written-response task before the unguided simulation task. Students' responses recorded in the online survey and log interaction data recorded in the simulation were analyzed to understand their problem-solving processes and evaluate their problem-solving outcomes. When comparing the time spent working on the unguided, simulation task, we found that there was no difference between whether students had gone through the guided written-response task first or not (mean time-on-task: 12.7 and 13.8 minutes, Wilcoxon test, $p = 0.35$).

Students' performance in the interactive task was evaluated by a pre-developed rubric scoring their submitted solutions for the hidden circuit in three dimensions: structure, components, and values. The score of each dimension ranges from 0 to 2, making the total solution score range from 0 to 6. The rubric for scoring students' submitted solutions can be found in Table 3.1 along with two examples in Figure 3.6.

Table 3.1: Rubric for scoring participants' solutions for the hidden circuit

Score	0	1	2
Structure	No evidence of how the circuit is connected	Partial structure identified	Full structure identified
Component (resistor, battery, wire)	None of the components correctly identified	At least one but not all components identified	All component correctly identified
Value (10V and 40 ohms)	None of the values correctly identified	At least one but not all values identified	All values correctly identified

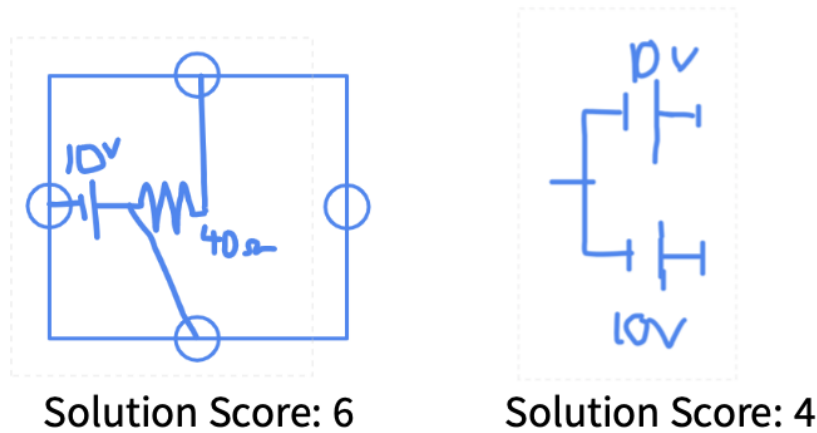


Figure 3.6: Two examples of solutions for the hidden circuit and their score

The researcher was blind to the group assignment when scoring students' solution drawings.

Only 8% of all students reached a solution score of 5 or higher in the simulation task ($n = 7$), and students' solution scores did not significantly differ between the group (mean solution score of the group who had the guided written-response task: 2.24; mean solution score of the group who only had the unguided, simulation task: 1.78, Wilcoxon test, $p = 0.10$).

Although no significant difference was found in students' solution scores in the simulation task depending on whether they went through the written-response task first, we saw

that the worked example did affect their solving practices by looking at the log data. Students who had the guided written-response task were more effective at decomposing the problem and collecting the data. This was seen in how students constructed circuits around the black box. Simpler circuits demonstrate an understanding of decomposing the problem into smaller, more manageable components while complex circuits, which involve connecting more than two wires of the black box at a time in a circuit, show an ineffective strategy, as the measurement readings from a complex circuit are hard to interpret and do not help determine the hidden circuit segment between two wires. Examples of simple vs complex circuits constructed around the black box can be seen in Figure 3.7. The mean percentage of complex circuits built by the group that had the guided written-response task was 18% and the mean of the group that only had the unguided simulation task was 46% (Wilcoxon test, $p < 0.001$, see Figure 3.8).

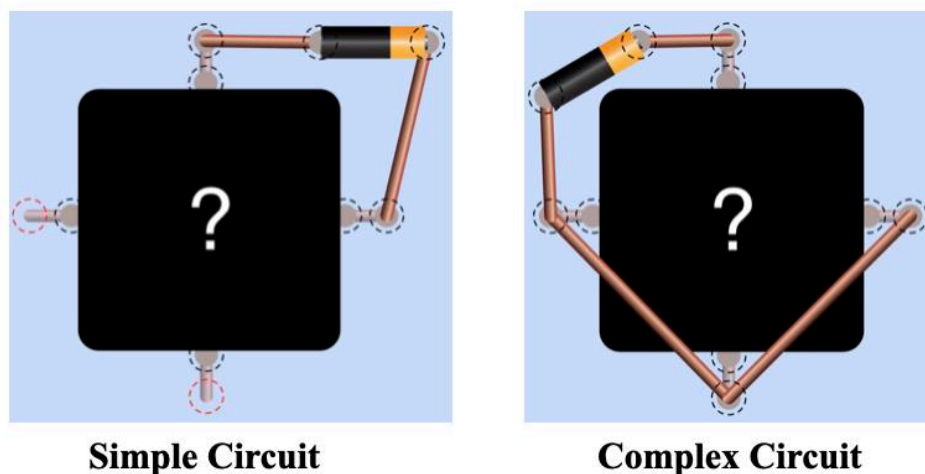


Figure 3.7: Example of simple vs complex circuits constructed around the black box

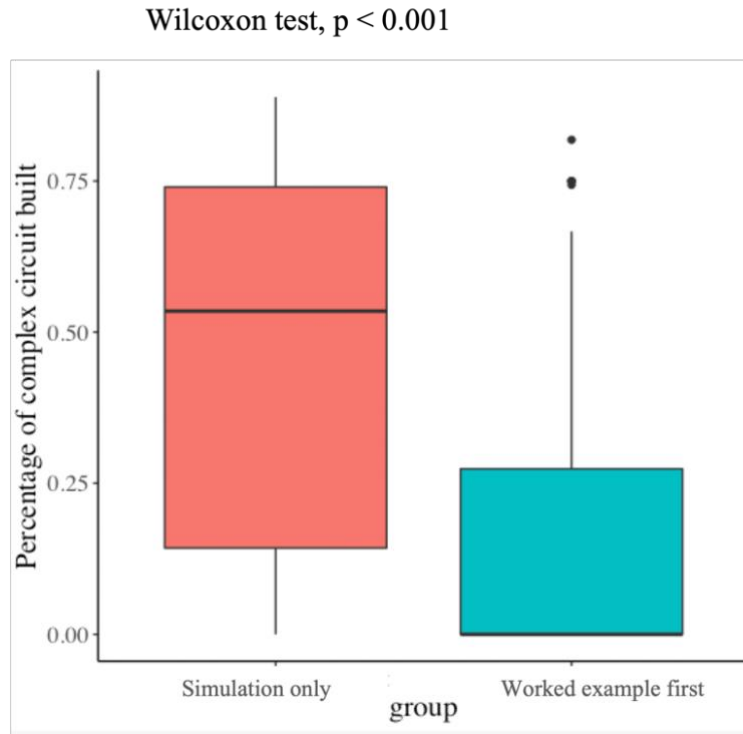


Figure 3.8: Percentage of complex circuits built by group

Greater use of measurement tools (voltmeter and ammeter) to collect data is also a desirable practice for identifying the hidden circuit. Students in the group that engaged in the guided written-response task before the simulation task used the voltmeter more often than those in the group that only had the simulation task. The mean voltmeter usage count of the written-response task group was 35 while the mean voltmeter usage count of the simulation-only group was 19, and the difference between the two groups was marginally significant (Wilcoxon test, $p = 0.06$, see Figure 3.9). However, there was no difference in ammeter usage count by the two groups (Wilcoxon test, $p = 0.67$).

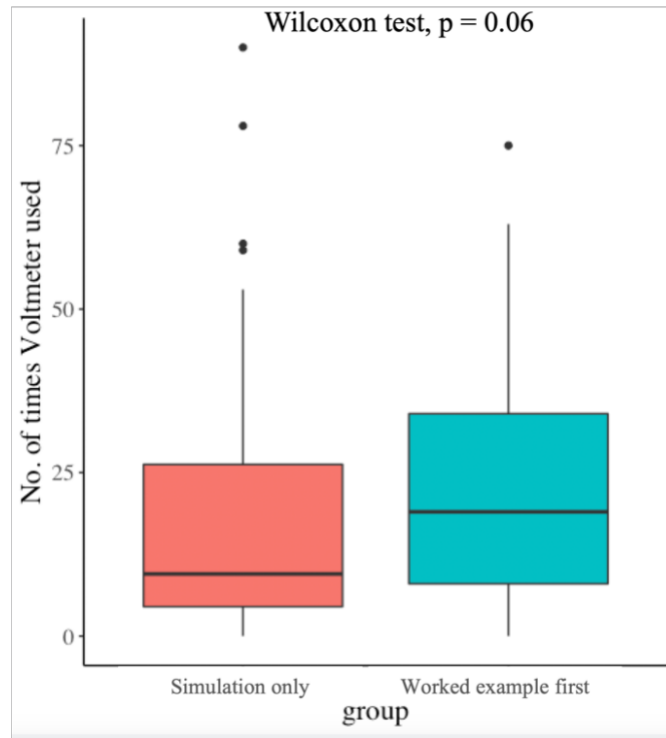


Figure 3.9: Number of times voltmeter was used by group

Another practice that was key to students' problem-solving success is data recording. Keeping track of all the data collected in a systematic manner makes it easy to perform calculations and synthesis as well as to review the progress. Though the log data does not directly monitor students taking notes using pen and paper, the data recording practice can nonetheless be inferred from the pauses of on-screen interactions after measurement events. Surprisingly we did not see a difference in pause time after students used the voltmeter (Wilcoxon test, $p = 0.59$) or after they used the ammeter (Wilcoxon test, $p = 0.66$) for students in the two groups.

After students uploaded the diagrams of their solutions for the hidden circuit for the simulation task, they were asked about whether and how they used any physics formula when solving the problem. Specifically, we investigated whether students referenced applying Ohm's

Law, and a summary of students' responses by group may be found in Table 3.2. We found that the proportion of students who reported applying Ohm's Law to the simulation task did not differ by group assignment ($\chi^2(1, N = 82) = 0.001, p = 0.97$).

Table 3.2 Summary of whether students referenced applying Ohm's Law to solve the hidden circuit problem by group

	Did not apply Ohm's Law	Applied Ohm's Law
Simulation only	n = 29	n = 11
Guided written-response task first	n = 30	n = 13

3.6 Discussion

The study showed the value of technology-based assessment tasks coupled with data mining techniques, as the difference in problem-solving processes between the two groups only became evident after mining the log data of students' interactions with the task. The study also contributes to our knowledge of the affordances and limits of worked examples in facilitating the learning of science and engineering practices such as problem solving. We found that the worked example in the written-response task helped students adopt more effective decomposition and data collection practices, as reflected by the smaller percentage of complex circuits built and the increased usage of the voltmeter. However, this improvement in problem-solving practices did not translate into more students successfully solving the hidden circuit in the simulation task. Several factors might have contributed to this. First, the proportion of

students who reported applying Ohm's Law to the simulation task did not differ by group assignment. In addition, we did not find a significant difference in the number of times that the ammeter was used for students in the two groups or in the pause time after ammeter or voltmeter usage. These results suggest that merely showing the data collection procedures in a worked example did not help students grasp the importance of flexible application of domain knowledge or keeping a detailed record of the data for solving the problem. It remains unclear to what extent students see connections between procedures for solving novel problems and specific problem-solving practices or whether they understand the decisions behind the selection of these procedures.

Worked examples that only show students the multi-step solution process may risk resembling "cookbook labs" where students carry out the procedures without much mental effort or learning (Holmes & Wieman, 2018). Thus to engage students in learning the practices used by scientists and engineers to solve authentic problems, we need to make both the solution steps and the decisions behind each step visible to students. Future studies will involve iteratively designing and studying different versions of worked examples aimed at helping students acquire effective problem-solving practices.

Chapter 4

Epistemological framing and strategy use during answer checking

4.1 Introduction

Reflection is widely recognized as an important part of the problem-solving process. Indeed, many problem-solving frameworks (Price et al., 2021; Salehi & Wieman, n.d.; Heller & Reif, 1984; Adams & Wieman, 2015; Wieman, 2015; Čančula et al., 2015) and templates (Van Heuvelen, 1996; Heller & Heller, 2000; Young & Freedman, 2015; Knight, 2017) include reflection on problem-solving processes and strategies (metacognition) or reflection on the solution to the problem itself as important final or ongoing steps of the problem-solving process.

Salehi's problem-solving framework draws a distinction between the “experimentation” practices of problem-solving and the “reflective” practices (Salehi & Wieman, n.d.). The experimentation practices are as follows: defining the problem and decomposing it into more tractable sub-problems, collecting the information/data necessary to solve the problem, representing and recording that data, and then interpreting the data to make some conclusions. This framework was developed for solving “authentic” problems, i.e., problems that a practicing scientist might encounter in the workplace but can also be applied to a traditional physics problem (Burkholder et al., 2020). Equally important to the experimentation practices are the

reflective practices: reflecting on the problem definition and assumptions made, reflecting on the solver's own knowledge as it applies to the problem, reflecting on the strategy used to arrive at the solution, and reflecting on the solution itself. Price et al. (2021) identify the same reflective practices in their framework of problem-solving, though those are instead framed as decisions to be made rather than actions to carry out.

As these theories illustrate, reflection during problem-solving is not a singular action but can be decomposed into different components. Of interest to us is the final reflective practice or decision: reflecting on the solution itself. Price et al. found that skilled scientists and engineers would reflect on the solution by exploring potential failure modes and limitations of the solution, determine whether the solution “makes sense,” often by passing discipline-specific tests for solutions to similar problems, and determine whether the solution meets the goals identified at the beginning of the problem-solving process (2021). Studies have shown that physics majors appear to have this ability, but often choose not to do so (Sikorski et. al, 2018).

There has been substantial work looking at sense-making in physics education research. A seminal study is Sherin's finding that students make sense of equations using symbolic forms (2001). These are similar to p-prims but posit that there is always a conceptual schema associated with the patterns of symbols in an equation. Eichenlaub and Redish discuss some examples of how students blend physical and mathematical reasoning, characterizing sensemaking in terms of epistemic games (2019). They showed that introducing some typical sense-making strategies can help novice students access more expert-like thinking.

Price et al. also note that, while the decision to reflect on one's solution is universal across science and engineering disciplines, how that decision is made is specific to a particular discipline or domain. Some existing problem-solving frameworks for introductory physics posit

that reflection on the solution to a physics problem consists of well-defined actions: e.g., checking the units of an expression, and checking limiting cases of the expression (Burkholder et. al, 2020). Others are less explicit and direct students to simply evaluate their work without providing guidance as to how to do so (Van Heuvelen, 1996; Heller & Heller, 2000; Young & Freedman, 2015; Knight, 2017). In any case, Warren’s study (2010) provides evidence that practicing reflecting on solutions to physics problems can improve future problem-solving performance, indicating the importance of this problem-solving practice for learning.

A seminal work on reflection comes from Dewey (1933) and was built upon by Schön (1983). Dewey characterizes reflection as a process of arriving at an idea of what is missing based upon what is known; he also characterizes it as an active and deliberate process. Schoen elaborates on this and argues that reflection allows practitioners to make tacit knowledge explicit; he also dichotomizes reflection into “reflection-in-action” and “reflection-on-action”. The framework of Price et al. identifies both reflection-in-action and reflection-on-action as critical parts of the problem-solving process. For the purposes of this study, students will be engaging in “reflection-in-action,” as they are being asked to process information in real-time and actively construct arguments about physical phenomena.

In a previous study (Burkholder et. al, 2020), E.B. and L.B. characterized how physics students decide to check whether an answer makes sense. Students were given two static equilibrium problems in two different formats and asked to verify the given answers to the problems without solving the problem themselves. From this prompt, they identified categories of strategies students employed to make sense of their answers, some of which reflected more expert-like thinking than others:

1. Identifying dependency. This group of strategies involves identifying what factors (e.g., variables) should affect the solution to a problem. In many textbook problems, this strategy consists of simply processing the problem statement and identifying what variables are specified and which are not. This is often not a cognitively demanding task, though there are common and notable exceptions, such as when students must decide whether the mass of an object will matter in a dynamics problem.
2. Identifying functional relationships. This group of strategies relies on either a mathematical or physical sense of how two quantities should be related. For example, identifying that the orbital velocity of a satellite should be proportional to the mass of the object it is orbiting. The cognitive demands of the strategies can be variable. The strategy could be as simple as identifying the covariational relationship described above or could be as sophisticated as carrying out dimensional analysis to determine the precise power of the proportional relationship. The commonly prescribed strategy of checking the units of an expression would fall into this category.
3. Making predictions. These strategies involve using mental models of a system to determine how it will behave under certain conditions. For example, identifying what happens when the angle of an incline is zero or 90 degrees. The most common strategy in this category is to evaluate limiting cases of an expression. These strategies are often the most cognitively demanding because they rely on simultaneous physical and mathematical reasoning.
4. Rederiving the solution. In the laboratory environment, students were given the solution to a problem and would often attempt to derive the solution themselves from first principles. We hypothesize that, in a more ecologically valid environment in which the

student derives the answer themselves, to begin with, this would be replaced with checking their math. The cognitive difficulty associated with this strategy is highly variable depending on the features of the problem.

This previous work provides one proposed ontology for reflecting on a solution in the context of introductory physics, though they hypothesize that these categories of strategies are likely to manifest in other contexts and be quite general. For example, if one is reviewing a historical argument, one needs to identify important individuals, locations, and events (Identify dependency), how those variables are related (Identify functional relationships), and predict how these factors might influence current events, or predict how changing one of the historical factors would have altered the outcomes to assess the quality of the argument. These categories are similar to the strategies for sensemaking identified by Hahn et. al (2017) but we believe that these strategies are more general than the strategies they identified. As discussed above, one can map our strategies into different disciplines outside of quantitative physics problems. The strategies identified by Hahn et. al were more specific to physics (e.g., find a “limiting case” or identifying “fundamental dimensions”).

It is known that novice students tend not to reflect on their solutions automatically (Ambrose et al., 2010), and even have difficulty checking their answers to problems (Warren, 2010; Burkholder, 2010). From a cognitive resources perspective (diSessa, 2017), this poses the question of whether this “difficulty” is because students lack the appropriate mental models and content knowledge required to carry out various solution reflection strategies (they lack the resources), or whether students simply do not use these strategies because they do not perceive them as germane to the task of reflecting on one’s solutions (the resources are not activated). Investigating this question thoroughly is complex due to the dynamic nature of epistemology

within an individual as well as across individuals within a system (Hammer & Elby, 2002).

Previous studies have investigated how assessment features activate local coherences of epistemic resources which influence how students frame a particular task (Shar, 2020).

We are interested in the epistemology of reflection on a solution: moving beyond what actions characterize this reflection to why students may approach reflection tasks the way that they do. We are interested in variation within individuals, as well as how it manifests across individuals with similar physics backgrounds. This study presents a pair of analyses that examine in detail student epistemologies while reflecting on the solution to a physics problem. In the first analysis, we focus on the prevalence of different answer-checking strategies across student levels and different problem types. This is to provide some evidence as to what common factors might impact student epistemology of solution reflection. The second analysis then investigates the epistemological framing associated with answer checking to provide a more detailed picture of this epistemology both within and across individuals.

4.2 Analysis 1

In a previous study (Burkholder et al., 2020), an experiment was carried out to explore what types of solution reflection strategies “transitioning novice” students used. In that work, a transitioning novice was defined as someone who had some explicit training in expert problem-solving practices but was far from expert in their behaviors. Students were given two static equilibrium problems in two different formats and asked to verify the answers to the problems without solving the problem themselves. The problems were chosen because they involved multiple concepts (both force and torque balances) and were posed in a way that is more authentic than typical physics textbook problems. For example, the problems asked about design

specifications for the systems instead of just calculating a force or a distance. Additionally, the problems were chosen because their solutions would have multiple unspecified quantities that one could conceivably vary in a real system (e.g., angle, mass, length, friction coefficient) and thus there would be multiple opportunities for students to engage in sensemaking even if they could not reason about a specific feature of the problem. From this prompt, we identified a limited set of strategies students employed to make sense of their answers, some of which reflected more expert-like mental models than others (see above).

Two strategies that we identified, used almost ubiquitously by experts when checking their answers, were evaluating limiting cases of an expression and checking the units of an expression. In a previous student, we found that 22- 38% of the 78% of students who completed the study checked the units of an answer without being prompted, compared with 73-80% of students who checked their units when explicitly told to. In both cases, most of the students were able to use this strategy correctly. Checking limits, however, was far less prevalent and much more difficult. Only 11% of students checked limits without prompting (8.1% correctly) and 27% percent checked when told to check the angular dependency of a solution (12% correctly).

In that work, we hypothesized that students did not check limiting cases because students lacked the appropriate cognitive resources to execute that strategy. Checking limiting cases of an expression requires the student to have a sophisticated enough mental model of a problem to make predictions about how it will behave under certain conditions. This requires deep conceptual understanding of the physics and mathematical understanding of the relationships between quantities. However, we also found evidence that students' strategy choices may be linked to their perceptions of the task. Most of the students in this study were able to

successfully check whether an answer was correct using simpler strategies like checking units or calculating force components. It was thus also possible that students did not check limiting cases because they did not see that as a necessary tool in verifying the answer to the problem (the resources were not activated).

In this first analysis, we aimed to investigate barriers to reflecting on one's solution by systematically varying assessment features (Shar et al., 2020) as well as students' physics backgrounds. Variation of assessment features was chosen to investigate the activation of cognitive resources, while students' physics background was used as a rough proxy for the availability of cognitive resources. We asked:

1. Does the mathematical and physical complexity of a problem scenario affect students' choice of solution reflection strategies?
2. Do we see variations in students' strategy use by prior physics experience?

We hypothesized that, if the assessment features were affecting students' strategy choice, we would see a difference in students' strategy use between the two problems. This would provide evidence to support the idea that strategy choice is affected by the activation of certain resources. If the barriers to checking answers in a more expert-like way were only because students lacked the cognitive resources to execute those strategies, we would likely see variation across student physics backgrounds but not across problem types. Variation across both student physics backgrounds and problem types would provide evidence that both resource possession and activation are important for strategy selection.

4.3 Methods

Data were collected in the fall quarter of 2019 and winter quarter of 2020 from undergraduate and graduate students enrolled at Stanford University (all interviews were conducted prior to the start of the COVID-19 pandemic). Graduate students (hereafter: Grad) in physics and physics majors (hereafter: Major) were recruited through emails to department mailing lists. Students who had previously taken an introductory physics course (hereafter: Intro) were recruited through emails from the instructors of those courses. Students who had not yet completed a physics course and had not taken physics in high school (hereafter: None) were recruited from Physics 41E, a course taught by E. W. B (Burkholder et al., 2021). All students were offered \$25 for an hour of their time. In all, we recruited 37 students (the breakdown by physics level is given in Table 4.1).

We note that there was a smaller number of physics majors and graduate students as those populations were smaller at Stanford University than the other groups of students. This means that the conclusions that can be drawn about behaviors of physics majors and graduate students may be somewhat limited as these samples may not be completely representative.

Table 4.1: Number of students by level and type of problem. The number of students at each level who completed each problem was not statistically different (Fisher's exact test, $p = 0.77$)

	Chandelier	Ladder	Total
No Prior Physics (None)	6	5	11
Former Introductory Student (Intro)	6	11	17
Physics Major (Major)	1	4	5
Graduate Student (Grad)	2	2	4
Total	15	22	37

In the interview, students were given 6 multiple-choice questions from the PIQL (Brahmia et al., 2021), as well as 6 open-ended questions probing their ability to use and interpret math in physics concepts. The open-ended questions were all written by E. W. B. Students were asked to think aloud as they solved all of the problems, and all the interviews were transcribed by L. F. B. and E. W. B. For the purposes of this study, we only considered the portions of the transcripts where students were solving open-ended problem 4. In this problem, students were presented with a situation where two fictional students, Sarah and Jessica, were studying for a midterm exam and came across a particular static equilibrium problem. They were presented with the text of the problem, as well as the two students' answers (one of which was correct, and one of which was incorrect). They were then asked to determine which of the solutions, if either, was correct. There were two versions of this problem, the Chandelier problem and the Ladder problem, which are shown in Figures 4.1 and 4.2. Students were randomly assigned one version of this problem in the interview (see Table 4.1). There was no statistically significant difference in which

version of the problem students saw by their physics background (Fisher's exact test, $p = 0.77$).

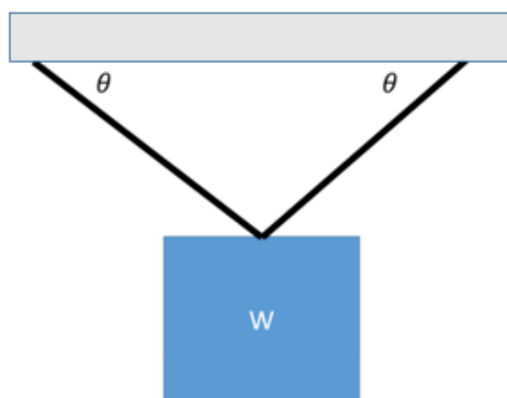
In both cases the “correct” student was Sarah.

We chose to use the language “correct” as we found this to be a clearer prompt for students than other sensemaking tasks we attempted during the same interviews. For the ladder problem, one cannot really determine which solution is correct without attempting to rederive the answer. Indeed, a limiting case of the “correct” answer makes a prediction that makes little physical sense. We found that only the graduate students noticed this, however, and so we gave students credit for both identifying this issue or saying that Sarah was correct.

The first version of the problem is called the “Chandelier problem,” and is mathematically and physically simpler than the ladder problem (see Figure 4.1). In this problem, the fictional students were asked to find the required strength of a rope to hang a chandelier from the ceiling. The chandelier is attached with two ropes to the ceiling at equal angles. Solving this problem simply requires the students to calculate the vertical component of tension and then do a one-dimensional force balance. The second and more difficult problem is the “Ladder problem” (see, Figure 4.2). In this problem the hypothetical students are asked to determine the maximum height off the ground you can stand on a ladder before it slides out from underneath you. Solving this problem requires multiple force and torque balances in two dimensions, as well as some more sophisticated trigonometry.

Two students, Sarah and Jessica, are working together on studying for their physics midterm. On the practice exam, they come across this question:

Consider a chandelier of weight W that is attached to the ceiling by two ropes of equal length, as shown below. One end of each rope is attached to the center of the chandelier, and the other end is attached to the ceiling. The ropes make an angle θ with the ceiling on either side of the chandelier. How strong does the rope have to be as a function of the weight of the chandelier?



They decide to solve the problem independently, and then compare their answers. Their answers are:

Sarah

$$F_{\text{rope}} = \frac{W}{2 \sin \theta}$$

Jessica

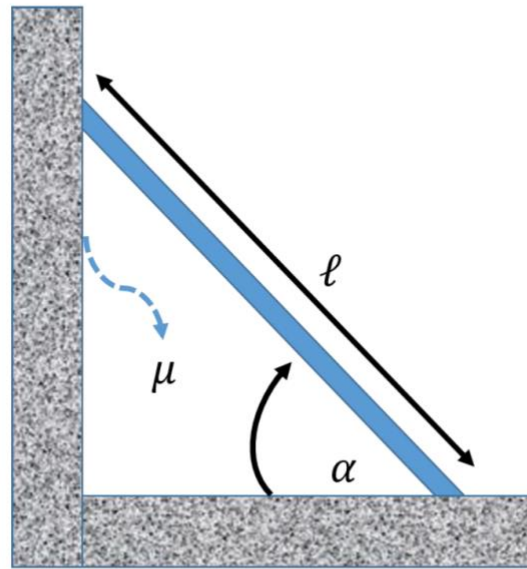
$$F_{\text{rope}} = \frac{W}{2 \cos \theta}$$

Without solving the problem yourself, determine if Jessica's answer, Sarah's answer, or neither is correct.

Figure 4.1: The Chandelier Problem

Two students, Sarah and Jessica, are working together on studying for their physics midterm. On the practice exam, they come across this question:

You need to change a lightbulb in your basement. To reach the lightbulb, you lean a lightweight aluminum ladder of length ℓ against the wall such that it makes an angle α with the floor. The floor and wall are both made of concrete, which has a coefficient of friction μ with the aluminum ladder. What is the maximum height off the ground h to which you can climb before the ladder starts to slide down and away from the wall?



They decide to solve the problem independently, and then compare their answers. Their answers are:

Sarah

$$h = \frac{\mu \ell}{(1 + \mu^2)} \left(\frac{\sin^2 \alpha}{\cos \alpha} + \mu \sin \alpha \right)$$

Jessica

$$h = \frac{\ell}{(1 + \mu^2)} \left(\frac{\cos^2 \alpha}{\sin \alpha} + \mu \cos \alpha \right)$$

Figure 4.2: The Ladder problem

The interview transcripts were all coded according to an a priori coding scheme.

Interviews were broken down into units determined to be discrete chunks of a thought. This often included multiple sentences. We coded each chunk for what strategy students were using to check their answers (Burkholder et al., 2020). The interviews were independently coded by E.

W. B. and O. C. M. in batches of 3 interviews. If during the coding process we felt that a chunk reflected multiple strategies simultaneously, we broke the chunk up into smaller pieces.

Students' solution reflection strategies were coded according to the findings of Burkholder et al. (2020); see Table 4.2 for examples and definitions. The two raters met after each group was coded and discussed and resolved all disagreements. In early groups, Cohen's kappa for interrater reliability was typically between 0.6 – 0.65. Disagreements between raters typically probed new edge cases of how the codes should be defined. As the definitions of codes were refined at each discussion point, previous coding was adjusted to fit the new rubric. In latter batches of coding, Cohen's kappa approached 0.8 – 0.85, indicating less variation between the two raters. As Hammer and Berland point out (2013), it is important not to treat this coding as error-free. To provide the reader with more context we present the full text of several example students to illustrate the findings found by the quantification of qualitative data in Appendix B.

We note that the organization of our strategies is different from that presented by Hahn et al. (2017). Their "special case" and "limiting case" would correspond to our "Making a prediction," provided there was physical sensemaking that accompanied it. Their "functional dependence" we have broken down into identifying physical dependence and identifying mathematical dependence. Their fundamental dimensions code would typically fall under "identify functional relationship." We note that our coding scheme is likely organized somewhat differently because of the different nature of the tasks employed in this study versus their study. However, the codes employed in that study can be mapped onto the codes used in this study.

Table 4.2: Coding scheme for strategy use including code name, code definition, and an example of each code from the data

Strategy	Definition	Example
Making a Prediction	A statement about how the equation/system should behave in a certain instance based on both physical and mathematical sense	Um, and then we have the weight and each of the ropes needs to be able to hold half the weight.
Identify Functional Relationship	Specifically identifying how two quantities should be related based on either physical or mathematical sense	As far as sine or cosine, let's see. Or adjacent over...well it doesn't have to hold the full weight so it makes sense that they're dividing it.
Identify Dependency	Identify that a quantity should depend on another quantity with no indication as to specific relationship	Okay so look at the sine squared over cosine. The height would matter. Would the height matter more or the distance away? Height would matter more.
Solve Problem	Trying to re-calculate the answer to make sense of it	So, this one has weight mg going downward and then I guess there's these two tensions holding it up. They're going to be equal.
Memory	Trying to remember what the answer is	I think it might be Sarah. Cause I remember, I remember using $mg \sin$ for things that are like falling because we want the y component instead of the x .

We were primarily interested in how strategy use varied by the assessment features and students' level of background physics knowledge. To investigate this, we thus counted instances of each strategy and computed what fraction of the responses used that strategy. We also counted the total number of times across all students within a group/question type that the strategy shifted (Hammer & Elby, 2002) to investigate the within subjects variability in epistemology of solution reflection.

4.4 Results

The total code counts by student physics background level and question difficulty may be found in Table 4.3. Percentages reported are the percentage of total coded lines in each category. For example, across the 4 graduate students, there were 23 codable lines, and 57% of those lines were associated with making a prediction. Note that we have included a code indicating whether a student arrived at an answer that was correct because in both problems, one of the alternative solutions we provided was correct and one was wrong. This indicates what percentage of students chose the correct answer from the two alternatives. In the following sections, we first address differences by problem type, then by student level.

Table 4.3: Code counts for students' strategy use during solution reflection. Percentages represent the total percentage of lines within each column that students were using a particular strategy. The total number of lines is given in the header

	Student Level				Problem Type	
Strategy	None (N = 40)	Intro (N = 97)	Major (N = 24)	Grad (N = 23)	Chandelier (N = 70)	Ladder (N = 114)
Making Prediction	0%	25%	79%	57%	22%	39%
Identify Functional Relationship	48%	54%	21%	35%	37%	50%
Identify Dependency	0%	2%	0%	4%	0%	3%
Solve Problem	50%	19%	0%	0%	38%	7%
Memory	3%	1%	0%	4%	3%	1%
Arrives at Correct Answer	38%	67%	100%	100%	67%	50%

4.4.1 Differences by question type

There was a statistically significant difference in strategy use across the two questions (Fisher's exact test, $p < 0.0001$, see Figure 4.3). Students were more likely to make predictions and identify functional relationships in the ladder problem, while they were more likely to attempt and re-derive the answer in the chandelier problem. We note that most students tried to make a prediction, identify a quantitative relationship, or solve the problem in both the ladder and chandelier problems, and that students from all levels were able to execute the more sophisticated strategies of making a prediction. Essentially zero students tried to simply identify the relevant variables or remember the answer. We also see that fewer students solved the ladder problem correctly compared with the chandelier problem.

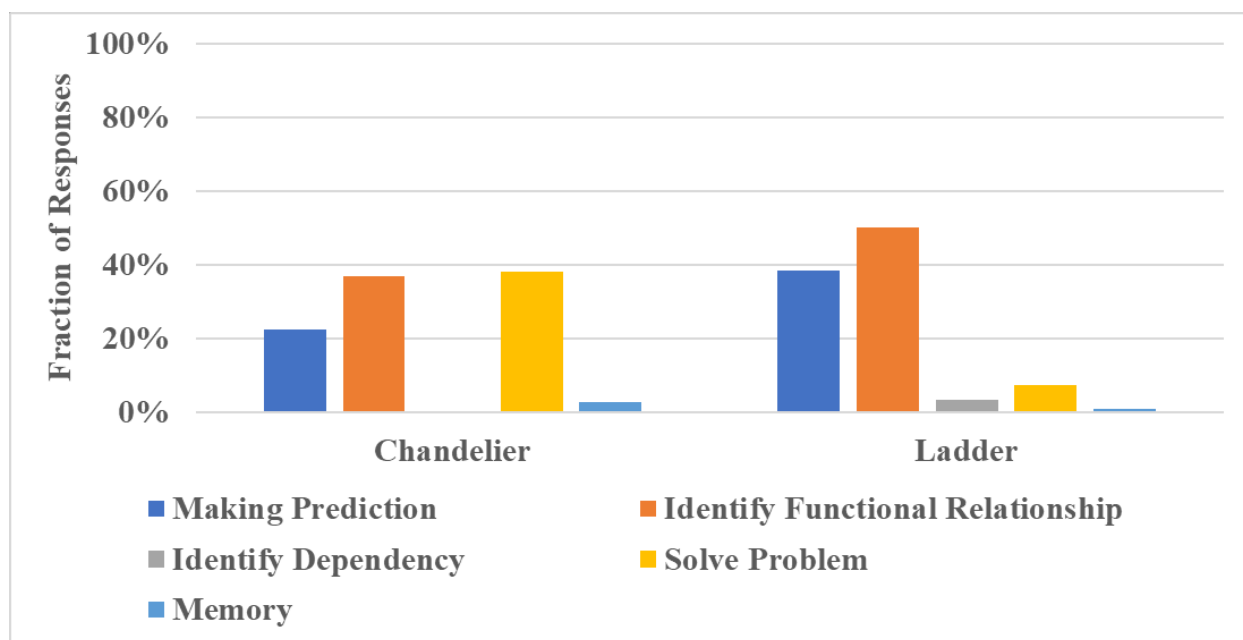


Figure 4.3: Student strategy use by problem type. The different colors indicate different strategies defined in Table 4.2

To illustrate these differences, we include two examples. Consider the following student attempting to check the answer to the Chandelier problem:

“Okay, well, ooh...so, we have one component that is going downwards. Okay, we have tension force going up at an angle. Correct. Now, if we're using this angle now the only question we have to break it into its x and y components, right?”

This student sees the problem and immediately starts drawing a force diagram and breaking forces into components. This is a procedure that they have committed to memory as what you are supposed to do when you see any problem involving forces. Prior research has shown that training students to do this can interfere with their overall problem-solving ability (Shar et al., 2020). To contrast this, consider this student solving the ladder problem:

“In Jessica's answer if the coefficient of friction is zero, then the height is equal to $\frac{\cos^2 \alpha}{\sin \alpha}$...which I'm pretty sure is not always zero, and so I'm gonna say Jessica can't be correct, so maybe Sarah is correct.”

This student has predicted that the answer should go to zero when the coefficient of friction is zero and is checking that prediction. This is a more sophisticated strategy that students in the Intro and None groups typically do not use (Sherin, 2001). Though the physics and math used to derive the answer to the ladder problem are far more complex, this illustrates that the task is not necessarily very cognitively demanding. Rather, the features of the problem statement may affect resource use and activation among students.

4.4.2 Differences by student level

There was significant variation in strategy use by student groups ($\chi^2(12) = 72.9, p < 0.0001$, see Figure 4.4) as well as the fraction of students arriving at a correct answer. Graduate students and physics majors spent most of their time attempting to make predictions about how the system would behave in various limiting cases, and some of their time identifying quantitative relationships between variables. Students in the Intro and None groups split their time between identifying quantitative relationships and attempting to rederive the answer to the problem themselves. However, some of the former introductory students did attempt to make predictions about limiting cases of the expressions. The percentage of introductory students making a prediction is almost identical to Hahn et al.'s number from their techniques of theoretical mechanics course.

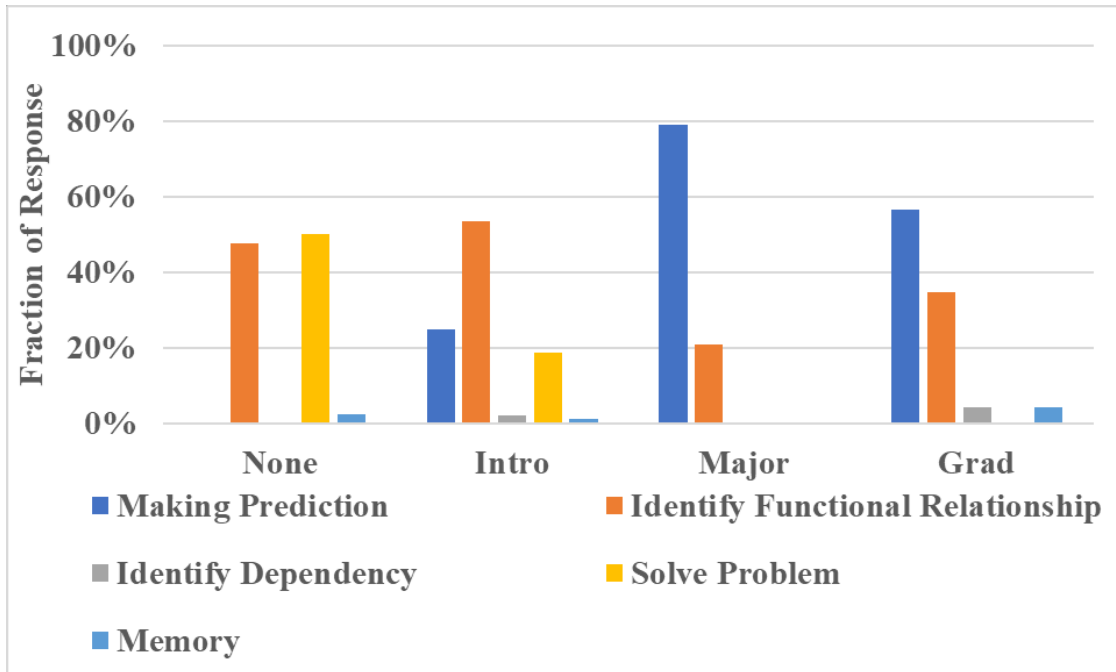


Figure 4.4: Student strategy use by students' prior physics experience. The colors correspond to the same strategies as Figure 4.3

4.4.3 Shifts in strategy

In Table 4.4 we count the number of strategy shifts by student level and problem type. We see that, overall, students do not shift strategies very often. We see small number of students who shift strategies often (all rates are 0.6-1 strategy shift per student with large standard deviations), but most students stay with the same strategy.

Table 4.4: Mean and standard deviation (in parentheses) of strategy shifts per student by student level

	Student Level			
	None (N = 40)	Intro (N = 97)	Major (N = 24)	Grad (N = 23)
Strategy Shifts	0.9 (1.4)	0.6 (1)	1 (2)	1 (2)

4.5 Discussion

There are several notable results from this first analysis, many of which align with expectations. First, we found that students used more sophisticated strategies in solving the ladder problem compared with the Chandelier problem. This aligns with our previous finding that students often adopt the simplest productive strategy to check their answers. The chandelier problem is quite simple to solve (a single force balance), and indeed it is simpler to re-derive the answer than predict limiting behavior. The ladder problem, however, is much more difficult to solve, and so it is simpler to check limiting cases than attempt to re-derive the answer. This also explains why we saw essentially no students trying to use “identify dependency” in the ladder problem. Though this is an easier strategy to use, both answers that students were given contained all the relevant variables, so it was not productive in terms of determining which expression was correct.

It also makes sense that more advanced students used more sophisticated strategies to check their answers. Physics majors and physics graduate students have spent a lot of time practicing checking the limits of expressions throughout their coursework and research and are more comfortable with this task. Introductory physics students are sometimes told to check limits, but it is often not required of them and thus they do not get as much practice with this

skill. They rely on skills they do practice often, such as identifying components of forces and solving force and torque balances.

We note that many of the introductory students did not exactly follow directions and attempted to solve the problem themselves. We do not consider this a failure of methodology or failure to follow instructions on the part of these students. Rather this reflects the unusual nature of a sensemaking task for novice students. Many of them attempted to start sensemaking, but in the process of doing so, resorted to the more familiar algorithmic strategies (see below) they associate with physics.

Based on our hypotheses, these results cannot rule out that strategy choice in solution reflection is affected by both the availability and activation of the necessary cognitive resources. Students are more likely to use more sophisticated strategies regardless of prior physics knowledge when it is required of them by the features of the problem – supporting the idea that students must perceive the task to require these more sophisticated strategies. At the same time, however, more advanced students are more likely to use these sophisticated strategies. This does not necessarily suggest that lower-level students lack the resources necessary to make predictions. It is possible that more advanced students use these strategies because they have been conditioned to believe that this is part of physics problem-solving, while more novice students have not been conditioned that way. This is in line with the findings of Gupta and Elby (2011) who showed that a first-year student had all the resources necessary for sense-making, but they were not activated due to epistemological barriers.

The results of the first analysis strongly suggest that epistemology plays a role in students' strategy selection while checking answers. Students' strategy selection – what it means to reflect on the solution to a physics problem – is clearly affected by both features of the

assessment as well as students' experiences with physics. We thus want to investigate the epistemological framing while checking answers to physics problems to provide a more detailed picture of what information we communicate to students when we ask them to reflect on the solution to a problem. That is the purpose of analysis 2.

4.6 Analysis 2

In this analysis, we adopt the theoretical framing of student epistemologies presented by Shar et al. (2020). In that work, Shar et al. investigated how assessment features impacted student engagement with the assessment problems. From think-aloud interviews with undergraduate students in introductory physics courses, they specifically explored how students frame assessments and what knowledge resources regulate those frames. Within the context of assessment features, Shar et al. further documented the stabilities and dynamics in resources and frames. To evaluate the ways in which student understandings of knowledge and learning are present in their engagement in assessments, Shar et al. developed a theoretical framework on epistemology. This framework involves epistemological framing, which answers 'what is it that is going on here', and epistemological resources, the smaller elements that make up the framing.

Rather than a stable model of epistemology, Shar et al. adopted a model of epistemology that is contextual and dynamic to characterize the approach students take toward learning-based activities. Particularly when engaging in learning physics, undergraduate students have been found to adopt a variety of frames depending on whether they are drawing on mathematics or physics knowledge or whether they are engaged in algorithmic or conceptual thinking (Burkholder et al., 2021). Moreover, students may transition between multiple epistemological frames during an assessment rather than adopting a single epistemology (Hammer & Elby,

2002). The epistemological frames that we adopt from Shar et al.'s work can be identified by student behavior and have been recognized in the literature by Chari et al. (2019). These epistemological frames include Conceptual Physics, Algorithmic Physics, Conceptual Math, and Algorithmic Math (Table 4.5).

In order to examine and document the dynamics of epistemological framing, Shar et al. identified epistemological resources as the behavioral clusters that are activated together in context to make up epistemological frames as defined in Hammer and Elby's work (2002). With frames defined as local coherences of resources, we can expect a fairly consistent set of epistemological resources to be associated with a particular epistemological frame as identified by behaviors. Likewise, we can expect a set of associated behaviors, or frames, to be associated with a group of resources (Hammer & Elby, 2002). The epistemological resources Shar et al. coded for included nature of knowledge, source of knowledge, epistemic activity, and epistemic source, which we also adopted to our work.

In our work, we apply this theoretical framework of epistemological framing as described in Shar et al.'s work. With this framework, we aim to examine student epistemological framing when making a specific problem-solving decision, i.e., deciding how well a solution holds (or solution reflection). Our focus is to thus examine how assessment features or prompts and student physics background change epistemological framing.

Our research question for this study was how epistemologies of solution reflection vary within students and across students of similar physics backgrounds.

4.7 Methods

We used the same interview questions from the same student population and the same interview transcripts. Interviews were broken down into discrete chunks (see above) and were coded for students' epistemic frames (Modir et al., 2017).

Students' epistemic frames were coded according to the scheme established in the work by Shar et al. (2020). The interviews were independently coded by E. W. B. and O. C. M. in groups of 3. The two raters met after each group was coded and discussed and resolved all disagreements. Interrater reliability was similar to that found in Analysis 1, and we acknowledge, not free of bias or error. Following the results of Shar et al. (2020), we also investigated instances in which students shifted frames. They found essentially no frameshifting in their investigation, but as we detail below, we find many instances of frameshifting. During the coding process, the two raters felt that math and physics frames frequently blended together, and it was often difficult to distinguish. Though we present examples in Table 4.5 for instances in which we could distinguish between math and physics frames, we only present the results in terms of conceptual versus algorithmic thinking.

Table 4.5: Epistemic frames and resource definitions adapted from Shar et al. (2020), including examples of each code from the current data set

Category	Name	Definition	Example
Epistemic Frame			
	Algorithmic Physics	Recalling or applying physics equations or knowledge to derive expressions	So mu sine... so... I'll pay attention to the orientation of the angle... this is mu cosine alpha and then sine alpha...
	Conceptual Physics	Discuss scenarios and properties of physics quantities	Assuming that there's no friction, like mu is 0, shouldn't these answers tell you that you can't climb any length l?
	Algorithmic Math	Performing mathematical computation	Uh, zero. I know. It's 0 over 1, that's great.
	Conceptual Math	Exploit properties of mathematical constructs to quickly obtain a result	Which...which leads me initially to think of like Sarah's answer as the better one because she has mu on the numerator and Jessica does not have that.

4.8 Results

4.8.1 Differences by problem type

The results detailing students' epistemic frames are listed in Table 4.6. Interestingly, there were no statistically significant differences by problem type in students' overall epistemic framing (Chi-squared test, $p > 0.10$). Though students were more likely to use more sophisticated strategies in the ladder problem compared to the chandelier problem, that did not

seem to translate into differences in epistemological framing that we were able to measure from the think-aloud protocol.

Table 4.6: Counts of epistemic frames by student level and problem type. N is the total number of codable lines within each column and category, and the percentage represents the fraction of responses that were coded as each frame/resource within that category by column count.

Category	Student Level				Problem Type	
	None (N = 40)	Intro (N = 97)	Major (N = 24)	Grad (N = 23)	Chandelier (N = 70)	Ladder (N = 114)
Algorithmic Math/Physics	90%	68%	29%	48%	66%	61%
Conceptual Math/Physics	10%	32%	71%	52%	34%	39%

4.8.2 Differences by problem student level

There was more variation in student epistemic framing by students' background physics experience. Intro and None students were more likely to frame the solution reflection task as an algorithmic physics or math procedure, while physics majors and graduate students were more likely to view it as a conceptual task ($\chi^2(3) = 28.0, p < 0.0001$). Consider two students both solving the chandelier problem:

“Another way of seeing this is that if you take theta to zero you do something like this which is like really hard to hold, so it should blow up and that's what happens.” – Grad

“Okay, so If we're trying to determine the force of the rope, the net force would be something like the force in the rope plus the weight, minus the weight.” – None

The graduate student is constructing a physical model in their head and making predictions about the system – they are in the conceptual physics frame. The student who has not yet completed a physics course, on the other hand, immediately tries to construct equations based on a force body diagram and is in the algorithmic physics frame.

4.8.3 Epistemic frame and strategy shifts

We identified a number of instances of students shifting frames in our investigation, averaging between 0.7 and 3 frame shifts per student across the different groups. Interestingly, graduate students and the most novice students were the least likely to shift frames, while intermediate students were more likely to shift frames.

This stands contrary to Shar et al.'s findings of students seldom switching frames (2020). It is possible that shifting frames is more inherent to the task of reflecting on one's solution as compared to the problem-solving task presented in the work of Shar et al. If a student is struggling to solve a problem in one frame, they should ideally be able to smoothly switch to another frame, as experts do (Bing & Redish, 2012). Alternatively, the lowest level students may not shift frames when it is required of them, resulting in difficulties problem solving, whereas the most advanced students do not shift frames because they are able to select an appropriate frame from the beginning of their problem-solving. This would seem to align with the percentage of novice and advanced students who arrived at a correct answer.

Table 4.7: Mean (and standard deviation) of epistemic frame shifts per student by student level

	Student Level			
	None (N = 40)	Intro (N = 97)	Major (N = 24)	Grad (N = 23)
Frame Shifts	0.7 (0.25)	1.2 (1.2)	3 (1)	1 (1.2)

4.9 Discussion

It is not surprising that less advanced students frame solution reflection as a task that relies on algorithmic procedures, while more advanced students frame it as something requiring deep conceptual knowledge and that more advanced students use qualitative reasoning more often. Indeed, typical introductory physics instruction often does not result in students acquiring adequate conceptual understanding of the content covered (e.g., Hake, 1998), and thus it is unsurprising that these students would not perceive physics problems as conceptual problems.

The finding that students more often shifted epistemological frames when engaging in solution reflection tasks contrasts with Shar et al.'s findings of students mostly working within the same frames. However, the nature of solution reflection could require students to engage in different frames depending on what strategy they use. For example, making a prediction can involve both algorithmic procedures (calculating the limit) and conceptual thinking (explaining what that limit means). However, based on the strategies the advanced students use and their overwhelmingly conceptual framing, it seems that frame shifting can be associated with answer checking, but is not always necessary.

The highest-level students did not often shift frames or strategies. This suggests they were able to select an approach to the problem that worked for them and resulted in a satisfactory answer. The lowest level students shifted strategies often, but not frames, which could suggest some fundamental difficulty with the task: they perceive physics problems in a certain way and have certain strategies they use, but those strategies do not work for checking the answer to a problem. Thus, they shift strategies often without fundamentally changing the way they think about the problem. The physics majors appear to shift frames often but not strategies. This suggests they have appropriate strategies to check their answers but may have to shift their

thinking about how those strategies apply to the problem at hand because they are still learning to use them.

We note that frame shifts do not correspond one-to-one with strategy shifts. As detailed above the hypothesis is that certain strategies may require shifting frames. We counted only two instances of frame shifts that coincided with strategy shifts, which supports this explanation. This would add to the literature suggesting that sense-making involves the coordination of mathematical and physical resources.

4.10 Conclusions, limitations, and future work

These two analyses provide rich data describing students' epistemologies while reflecting on the solution to a physics problem. Broadly speaking, the data support (1) the influence of both the availability and activation of cognitive resources on students' strategy choice in solution reflection and (2) the instability of epistemology for intermediate students, whereas more seasoned physics students and the most novice physics students exhibit more stable epistemologies.

We saw significant variations in student strategy use across both two different problems and different levels of physics background. If strategy choice was only affected by the availability of cognitive resources, we would not expect to see differences in strategy use for different answer-checking problems. The variation in strategy use across student levels could be an indicator of both the availability and activation of resources. More advanced students typically use more sophisticated strategies, but this could be either because they have been conditioned to use those strategies in that context, or that they have more strategies available to them to reflect on solutions. To check this, we analyzed the strategy use on a different sense-

making task from the same interview and found no difference in strategy use between different student backgrounds, suggesting that in different contexts, more “novice” strategies can be activated in the more advanced student groups as well. Overall, this seems to align with findings that strategy choice is determined by the activation of resources (e.g., Hammer & Elby, 2002).

Our more detailed investigation of the epistemology of answer checking showed that advanced students are more likely to perceive solution reflection as a conceptual task. They also associate solution reflection with particular strategies, like checking limits and identifying functional behavior, or making predictions about how a system will behave. Based on the small number of participants, it appears that the correlation between strategy use and problem difficulty is biased by more novice students choosing easier strategies for the easier problem. They have a certain perception of solution reflection as an algorithmic task that involves checking their force components and their calculations. When these strategies fail, they can turn to other ones, but still largely view those as algorithmic procedures to apply in order to arrive at a solution. The data seem to suggest that the resources that instructors want them to use are not often activated by the problem context.

The intermediate students show some variation in strategy use by problem difficulty, but also a larger variation in epistemological framing of the solution reflection task. This seems to represent a transition between a stable, but often unproductive, epistemology of solution reflection as an algorithmic procedure to be performed using basic problem-solving strategies they have acquired during physics courses, to another stable epistemology where solution reflection is more about conceptual understanding and sensemaking.

In all, the data seem to show that it is often difficult to distinguish between physical and mathematical sense-making during the answer. There is also evidence of many students shifting

between algorithmic and conceptual thinking during answer checking. We presented an example of how this might manifest in the context of checking limits of an expression above. We note, however, that there are some advanced students who are able to check their answers without shifting frames. For example – some of the graduate students are able to reason about an answer entirely conceptually, without mechanically calculating the limit of an expression. This did not, however, make them more successful than the physics majors who often shifted frames.

Though we believe these results are interesting, there are some methodological limitations to this work. First, we use think-aloud protocols as measures of epistemology. We are thus inferring how students perceive a task based on verbal reports (Ericsson & Simon, 1980). Another measure of epistemology would be to allow students to complete the think-aloud task, and then ask follow-up questions about why they made certain strategy choices or what they thought the question was asking them to do. This would have been impractical with this protocol, as we were attempting to characterize other aspects of mathematical reasoning in physics at the same time and did not want to influence students' behavior by asking probing questions.

The second limitation is that our solution reflection task is limited to reflection in action and does not probe reflection on action (Schön, 1983): it asks students to make sense out of a solution given to them rather than a solution they derived themselves. Future work should examine whether strategy use and epistemology change when checking your own answer as compared to someone else's answer because, as seen here and by Shar et al. (2020), assessment features are important in activating various cognitive resources that students draw upon in problem-solving.

Though qualitative in nature, this study provides some directions for future instruction. First, it is important to recognize that intermediate physics students have unstable epistemologies. These students, as they are starting to engage in more conceptual thinking and sensemaking, can be easily provided with feedback by an instructor when they are engaging in productive sensemaking strategies. This may speed the development of more expert-like sensemaking strategies.

An obvious direction of future inquiry to the authors, aside from solidifying our understanding of the epistemology of student solution reflection in more authentic contexts, is to study ways we can shape not only the cognitive aspects of problem-solving for students, but also the epistemological aspects. It is clearly not enough to tell introductory students to check limits or units of an expression and still expect them to engage in the kind of sense-making which physicists associate with those activities. We hypothesize that a targeted reflection activity (e.g., Salehi, 2018) could have some impact on how students perceive the solution reflection task if repeated frequently throughout the introductory course. Learning science suggests that instructors can model “doing physics” for students in order to communicate expectations. If we are to communicate to students that physics is about building and understanding simplified models of real systems, those ideas must be integrated into our lessons and assessments.

Chapter 5

Student metacognitive ability on introductory physics assignments

5.1 Introduction

Metacognition in science education is garnering more attention because it is widely acknowledged that cultivating metacognitive abilities is a fundamental aspect of the learning journey (Baird, 1986; Pintrich, 2002; Georghiades, 2004; Rickey & Stacy 2002; Zepeda et al., 2015; Greene & Azevedo, 2009; Fisher, 1998). Specifically, studies have indicated that metacognition plays a vital role in students' acquisition of physics knowledge (Schunk, 2008; Zimmerman, 2006; Zimmerman & Campillo, 2003). However, there is limited research on how metacognition specifically affects students' ability to solve physics problems.

Metacognition is defined as “any cognitive activity that takes as its object, or regulates, any aspect of any cognitive enterprise” (Flavell, 1985). Similarly, metacognitive ability is described as the ability to reflect upon, understand, and control one's learning (Schraw & Dennison, 1994). Metacognitive skills enable individuals to monitor their knowledge and skill levels, plan and allocate learning resources efficiently, and evaluate their learning state. Studies have shown that students who receive instructions and training on metacognitive strategies perform better on self-regulated learning activities and exhibit increased motivation (Zepeda et

al., 2015). These learners use cognitive strategies to construct knowledge, while also engaging in planning, monitoring, and evaluating their own thinking. By employing these metacognitive strategies, they are able to optimize their learning.

While research has shown the positive effect of metacognition on physics knowledge, there is a lack of studies examining the role of metacognition in physics problem solving. This is perhaps due to the absence of an inventory that measures metacognition for science problem-solving whereas most of the research investigating metacognition in this context uses verbal interviews or a small set of research-developed items (Rozenkwajg, 2003). These studies, though, have found that students who are more metacognitive during physics problem-solving are more likely to correctly solve the problems (Neto & Valente, 1997; Rozenkwajg, 2003). The absence of research pertaining to the impact of metacognition on problem solving in physics is concerning, particularly considering the importance of problem solving in terms of achievement and advancement in the field (Chi, 2006). To address this absence, it is necessary to separately investigate the role of specific metacognitive subcomponents in physics problem solving.

Metacognition can be divided into two main components which include knowledge about cognition and regulation of cognition (Brown, 1987; Flavell, 1987; Jacobs & Paris, 1987). Knowledge about cognition pertains to an individual's understanding of themselves as cognitive agents, various methods applicable to learning and problem solving, and the specific requirements of a given learning task (Brown, 1987). There are three subcomponents of knowledge about cognition: declarative knowledge, procedural knowledge, and conditional knowledge. Declarative knowledge refers to the knowledge about oneself as a learner or a problem solver. Procedural knowledge relates to the knowledge about how to perform a task or an activity. And conditional knowledge includes knowing when and why to use declarative and

procedural knowledge (Taasobshirazi & Farley, 2013). The other main component of metacognition is regulation of cognition, which refers to adjustments or actions individuals make to control or manage their own learning (Pintrich, 2002; Schraw, Crippen & Hartley, 2006; Schraw & Moshman, 1995). There are five subcomponents of regulation of cognition: planning, monitoring, debugging, information management, and evaluation. Planning includes goal setting, activation of past information, and arranging time, while monitoring involves the continuous assessment of one's goals, work, and performance. Debugging encompasses the elimination of unnecessary information or strategies used to correct learning and problem-solving mistakes. Information management incorporates strategies used by an individual to help them solve a problem more effectively. An example of this in physics includes creating and using free-body diagrams. Lastly, evaluation refers to the assessment of one's work after it is completed or the reviewing of one's learning and associated products and process (Taasobshirazi & Farley, 2013). This subcomponent of metacognition will be the focus of this study.

Evaluation is an essential component of the physics problem-solving process (Heller et al., 2001). It helps to verify whether solutions obtained are correct as well as identify mistakes made during the problem-solving process, which can help avoid similar mistakes in the future. Furthermore, evaluation helps assess the quality of the solution obtained and whether it is reasonable or makes sense in the context of the problem. Through critical analysis of solutions and identification of weaknesses or areas of improvement, evaluation can additionally help develop problem-solving skills. Exercising such reflective practices helps problem solvers process and decide how to both adapt and transfer their practices to the context of a given problem (Glaser et al. 1992; Schoenfeld, 1992). The ability to effectively apply reflective

practices and evaluation techniques in problem solving is a characteristic of expert problem solvers (Price et al., 2021) while novice problem solvers tend to view problem solving as a more linear process (Heller et al., 2001).

In this study, we aim to investigate the trends in students' metacognitive ability on homework assignments in an introductory physics course and whether their ability changes over the semester. More specifically, we explore whether their ability to evaluate their assignment solutions improves over the semester or whether their ability to evaluate their solutions depends more upon the given problem. We also examine student characteristics that contribute to more successfully evaluating solutions and the correlation of successful evaluation on the final course score.

5.2 Methods

Data were collected in the fall 2021 semester of the undergraduate introductory physics course Applied Physics 50 (AP50): Physics as a Foundation for Science and Engineering at Harvard University. This course is a one-year, calculus-based introductory physics course focused on the application of physics to real-world problems to teach scientific reasoning and problem-solving skills. AP50 is a project-based course that aims to develop skills useful in a student's future career through emphasizing learning by doing rather than listening, exchanging ideas with others, seeing how science applies to the real world, and working in teams to solve problems and build things.

The fall 2021 semester of AP50 included 103 Harvard College students who consistently completed the homework assignments. Of these students, 57% identified as female and 43% as male. Almost half of the students in this course were life science concentrators (49%), and 29%

were engineering concentrators. The remaining 14% were grouped as ‘other’ concentrators, which did not fall into the life science or engineering categories (e.g. economic and history concentrators), and 8% had not yet declared a concentration (‘undeclared’). Most of these students were in their sophomore (43%) or junior (38%) years, while 10% were first-year students and 9% were seniors. Table 5.1 summarizes the student population by concentration and college year. Prior to taking this course, the majority of students had taken high school physics (89%) while a small portion had taken college-level physics (6%) or had no previous physics (5%). Furthermore, at the beginning and at the end of the course, students were given the Force Concept Inventory (FCI), which is a multiple-choice assessment to measure their understanding of basic concepts in Newtonian physics (Hestenes, Wells, & Swackhamer, 1992). The mean FCI pre-score for the course was 54.7% with a standard deviation of 23.6%, and the mean FCI post-score was 68.1% with a standard deviation of 23.0%. Students’ FCI normalized gain was calculated using the following formula as introduced by Hake (1998):

$$(1) \quad \langle FCI_{gain} \rangle = \frac{\langle FCI_{post} \rangle - \langle FCI_{pre} \rangle}{100\% - \langle FCI_{pre} \rangle}.$$

The normalized gain is the difference between a student’s pre- and post-score, normalized with their pre-score. In other words, this can be described as the amount a student learned divided by the amount they could have learned. The normalized gain allows for a consistent analysis over diverse student populations with varying initial knowledge states.

Table 5.1: Number of students by undergraduate year and concentration

	Life science	Engineering	Other	Undeclared	Total
First year	1	7	1	1	10
Sophomore	15	21	3	6	45
Junior	29	2	7	1	39
Senior	6	0	3	0	9
Total	51	30	14	8	103

5.2.1 Weekly homework assignments

The AP50 course consists of three main components: pre-class reading assignments and activity preparation, weekly in-class activities, and projects. The in-class (or synchronous) activities first require students to work individually (asynchronously) on the activity, and then in class, students engage in the activity with their assigned groups. These activities include the tutorials, which are aimed at addressing common misconceptions and deepening understanding of the course material using Desmos, and homework assignments, which are focused on developing problem-solving and estimation skills. This study focuses on students' individual work on their weekly homework assignments.

Students are assigned three problems per week for their homework assignments. These problems consist of two textbook-like problems and one estimation (or Fermi) problem, which will be described in more detail later. Students first work on these assignments individually and asynchronously, and they submit their work to the online grading and assignment tool Gradescope before the class section on Thursday. During the Thursday class time, students work

with their assigned groups to discuss the three problems and mark up their work in a different color pen. Throughout this process, groups also discuss their problem-solving approach and solutions with a member of the course's teaching staff and thus receive prompt feedback on their work. At the end of the class section, students receive detailed solutions to the problems and have the opportunity to continue to mark up their work. By the end of the day, students are required to submit their marked-up work along with a thoughtfully completed reflection form to Gradescope. In addition to the in-class feedback, students receive feedback on their work within a few days of submission. Rather than on solution accuracy, student work is graded on their individual (asynchronous) effort that clearly demonstrates thoughtfully attempting all problems following the four-step problem procedure described below. Student work is also graded on clearly and accurately addressing any conceptual and/or calculation errors made during their individual attempt. Furthermore, a portion of their grade includes completing the reflection sheet by accurately evaluating their own work and articulating how they will review to improve. There are a total of 12 assignments per semester, and this study focuses on a subset of 6 of these assignments throughout the semester (2 at the beginning, 2 in the middle, and 2 at the end of the semester).

In order to receive full credit on the assignment problems, students must clearly follow the four-step problem-solving procedure described in their course text. Most introductory physics students do not walk into a classroom with a robust and logical general problem-solving framework. An explicit problem-solving framework helps students make connections between physics concepts and the rest of their knowledge. The framework students are instructed to follow is adapted from George Polya's framework detailed in his 1945 book *How to Solve It*. This framework includes four basic principles of problem solving: understanding the problem,

devising a plan, carrying out the plan, and looking back. Similarly, the problem-solving framework in AP50 includes the following four steps: getting started, devise plan, execute plan, and evaluate solution. The getting started step consists of organizing the given information and understanding what is asked for the problem. This step also includes addressing any assumptions that need to be made to solve the problem. The next step, devise plan, focuses on spelling out what must be done to solve the problem. This includes acknowledging what physical relationships or equations can be applied and how the problem can be broken down into smaller, more manageable segments. The execute plan step requires the problem solver to follow the steps outlined in the previous section and articulate the thought process at each step. Without an explicit problem-solving framework, students typically just engage with this step to solve a problem. Lastly, the evaluate solution step involves reflecting on the work in the previous steps, examining the result, and determining whether the solution is reasonable. There are many ways the student can justify their solution such as checking the symmetry of the solution, evaluating limiting or special cases, relating the solution to situations with known solutions, checking units, using dimensional analysis, and/or checking the order of magnitude of the answer. In this study, we are particularly interested in how students engage with this reflective last step. Some researchers and theorists suggest that metacognitive processes including planning, monitoring, and evaluation, may not be conscious or explicit in many learning situations (Butler & Winne, 1995; Pressley, Borkowski, & Schneider, 1989). However, by specifically instructing students to follow the four-step problem-solving procedure we can elicit the specific metacognitive subcomponent of evaluation.

The problem-solving framework is not a formula for solving a problem, as one does not exist for true problem solving, but it does help students to establish links between physics

concepts and their existing knowledge, especially when solving more authentic problems. As previously mentioned, each homework assignment consists of three problems: two textbook-like problems and one estimation (or Fermi) problem. We are particularly interested in investigating how students evaluate their solutions on the third problem of each assignment as these problems are more like authentic, real-world problems. Such authentic problems are unrestricted, as there are multiple possible solution methods and thus a range of possible correct answers, whereas typical textbook problems simply require students to apply a known procedure to find an unknown answer. These estimation problems, therefore, require the solver to make more expert-level decisions, as highlighted in Price et. al (2021), including determining when and how to apply relevant physics knowledge to approach the problem. We focused on a subset of 6 estimation problems throughout the semester (2 at the beginning, 2 in the middle, and 2 at the end of the semester). These estimation problems along with their corresponding assignment number can be found in Table 5.2.

Table 5.2: Six estimation problems

Assignment	Problem
2	What is the combined distance walked by all the humans in the history of the world?
3	A penny is dropped from the top observation deck of the Eiffel Tower in Paris, France. Estimate its speed upon hitting the pavement. Please state your assumptions clearly and describe how you arrive at your estimation.
6	The Druzhba pipeline is one of the biggest oil pipeline networks in the world. It carries oil 4,000 kilometers (2,500 mi) from the eastern part of Russia to various countries in Eastern Europe. To reduce pollution, an engineer suggests that the entire pipeline should be emptied and raised 1 m above the ground. Estimate the amount of energy this would require.
7	While helping your young niece build a birdhouse you demonstrate how to use a hammer. Unfortunately, you miss the nail and bring the hammer down on your thumb which is holding the nail. Estimate the impulse the hammer delivers to your thumb.
10	Estimate the change in angular velocity of the earth if the entire human population simultaneously faced west and took one step forward.
12	Estimate the number of seats in Sanders Theater where you would hear a reduced sound level if two clarinet players on center stage standing directly next to one another played the exact same tone at 440 Hz.

5.2.2 Evaluation framework

To better understand the trends in how students evaluate their solutions to the estimation problems over the semester, we developed a framework to capture the different methods for evaluation as summarized in Table 5.3. This type of framework involved systematically labeling

and grouping similar techniques to generate themes in how students evaluate their solutions and make analyzing the data possible.

Table 5.3: Framework for how students evaluated their solutions on the estimation problems

Evaluation technique	Description
Reflection using problem definition and assumptions	States how assumptions or problem definitions affect solution more than “if my assumptions are correct, then I think my answer is correct”
Reflection using knowledge	Uses previous knowledge or experience to reflect on reasonableness of solution
Reflection using strategy	Reflects on how strategy or approach affects solution more than just strategy “making sense”
Reflection using solution	Includes using the answer itself to justify solution: Uses magnitude of solution to justify answer Googles or looks up solution
Mathematical answer-checking strategies	Checks bounds to solution, Converts units of answer to something more familiar, Uses solution to calculate another value to justify answer, or Calculates another value to compare solution to
No substantial evaluation	Justifies answer through approach “making sense” Statement of “answer seems reasonable” without justification Just restates solution

For the six estimation problems, students’ fourth step of the problem-solving framework, evaluate solution, was coded according to the evaluation framework above. The first four items

of this framework (Reflection using problem definitions and assumptions, Reflection using knowledge, Reflection using strategy, and Reflection using solution) are inspired by reflection codes developed by Price et al. (2021) and Salehi (2018). However, adjustments have been made to emphasize the reflection of the solution *using* a specific area of the problem-solving process rather than *on* these areas of the process as to better align with the instructions provided to students to specifically evaluate their solution in some way.

The evaluation technique of reflection using problem definitions and assumptions involves the problem solver reflecting upon whether their solution makes in terms of the goal of the problem and what assumptions they made to solve the problem. For this to count as a successful evaluation technique, students need to be specific as to how they defined the problem space or how their assumptions affect their solution more than stating that if “the assumptions make sense, then the solution makes sense”. Reflection using knowledge is when the solver justifies their solution through their content knowledge or their experience with the problem situation such that they have a familiarity or expectation of what the solution should be. The next evaluation technique of reflection using strategy includes reflecting on their process for solving the problem and how it resulted in the solution. This could involve reflecting on the effectiveness of the strategy for producing the solution, but this needs to be more specific than simply stating that if “the approach makes sense, then the solution makes sense”. The last reflection technique is reflection using solution which encompasses using the value of the solution itself to verify that it makes sense. For example, the problem solver may justify the solution’s reasonableness by interpreting its order of magnitude relative to the problem or by looking up a value to compare their solution to.

In addition to these reflection codes, students can also evaluate their solutions by engaging in some mathematical procedures to check or compare their solutions. This can include checking the upper and lower bounds to the solution, converting the units of the answer to a unit the solver is more familiar with, using the solution to calculate another value, or calculating a separate value to compare to the solution. Students do not always incorporate reflection or mathematical answer-checking techniques into their evaluation, though. Instead, students will simply state something like “the solution seems reasonable”, reiterate the final solution, or leave this section blank. Such responses are coded as “no substantial evaluation”.

Along with coding for specific evaluation techniques, we also recorded whether students engaged in more than one evaluation technique when evaluating their solution. To develop a fuller understanding of students’ ability to evaluate their solutions on the estimation problems, we recorded how many orders of magnitude they were from the correct answers provided by a group of experts on the teaching team for AP50. These solutions can be found in Appendix C. Furthermore, we recorded whether students correctly applied correct physics concepts and equations when solving the problem.

5.3 Results and discussions

5.3.1 Importance of evaluation on course score

Achieving good academic performance in a course is an important goal for many students, and there are many factors that can contribute to students’ final grades. In the course AP50, we found that evaluating estimation problem solutions successfully, that is using one of the five evaluation techniques described in the framework, was an important factor in students’ overall course scores. The mean (raw) course score for the class was 65% with a standard

deviation of 11%, and the course score is normally distributed about the mean. We investigated the relationship between successful evaluation and course score by assignment, a summary of which may be found in Figure 5.1

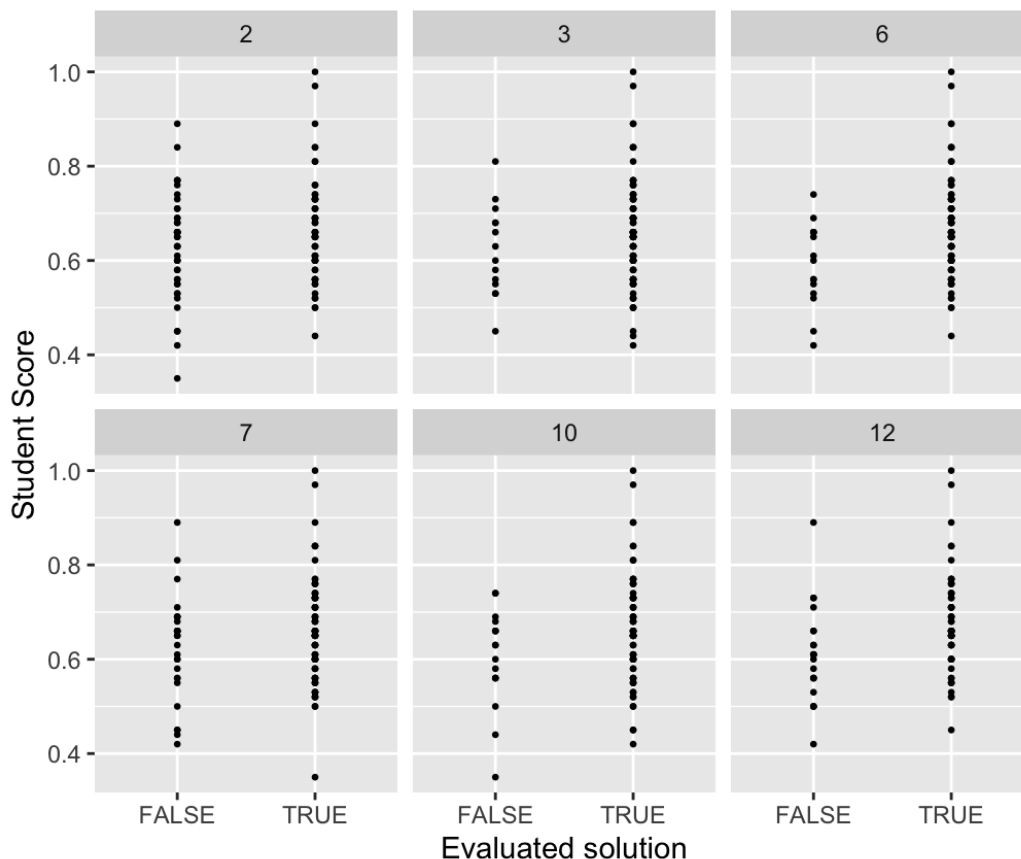


Figure 5.1: Trellis plot depicting the course scores for students who did not successfully evaluate their solutions and students who did by assignment

Furthermore, we conducted a linear regression analysis to investigate the relationship between students who evaluated their solution successfully and their final course score, controlling for concentration, college year, previous physics experience, and FCI normalized gain for each assignment. We fit the following population model for each assignment:

$$\begin{aligned}
(2) \quad score_i = & \beta_0 + \beta_1 evaluated_soln_i + \beta_2 concentration_LifeScience_i \\
& + \beta_3 concentration_Other_i + \beta_4 concentration_Undeclared_i \\
& + \beta_5 year_Sophomore_i + \beta_6 year_Junior_i + \beta_7 year_Senior_i \\
& + \beta_8 previous_physics_NoPhysics_i + \beta_9 previous_physics_HSPysics_i \\
& + \beta_{10} FCI_normalized_gain_i + \varepsilon_i
\end{aligned}$$

where engineering is the reference category for concentration, first year is the reference category for college year, and college physics is the reference category for previous physics experience. The fitted model for assignment 2 is shown below:

$$\begin{aligned}
(3) \quad \widehat{score} = & 0.62 + 0.06 evaluated_soln - 0.0001 concentration_LifeScience \\
& - 0.01 concentration_Other + 0.14 concentrataion_Undeclared \\
& + 0.01 year_Sophomore + 0.01 year_Junior - 0.01 year_Senior \\
& + 0.01 previous_physics_NoPhysics \\
& - 0.02 previous_physicss_HSPysics - 0.001 FCI_normalized_gain
\end{aligned}$$

We tested a null hypothesis that there is no difference in mean scores for those who did and those who did not successfully evaluate their solutions when controlling for concentration, college year, previous physics experience, and FCI normalized gain in the AP50 class population for assignment 2. We found sufficient evidence of a mean difference ($t(df=85) = 2.241, p < 0.05$) in final course score in the class population for those who did and those who did not successfully evaluate their solutions when controlling for concentration, college year, previous physics experience, and FCI normalized gain for this assignment. Therefore, we reject the null hypothesis of no mean difference. When controlling for the aforementioned variables, those who successfully evaluated their solutions on assignment 2 have a mean difference in final score that

is 6% higher than those who do not successfully evaluate their solutions. Plausible values for the association in the class population range from 1.50% and 11%, meaning we can be confident of a mean difference. Furthermore, this linear regression model explains about 2.8% of the variability of the final score. This is unsurprising as there are many course components that contribute to students' final scores and this is only for one assignment.

When investigating subsequent assignments, we found there was also a statistically significant relationship between whether students evaluated their solution successfully and final course score for assignments 6 and 12 when controlling for concentration, college year, previous physics experience, and FCI normalized gain. For assignment 6, we fit a linear regression model as shown below:

$$\begin{aligned}
 (4) \quad \widehat{score} = & 0.58 + 0.07\textit{evaluated_soln} - 0.01\textit{concentration_LifeScience} \\
 & - 0.04\textit{concentration_Other} + 0.08\textit{concentration_Undeclared} \\
 & + 0.04\textit{year_Sophomore} + 0.05\textit{year_Junior} + 0.07\textit{year_Senior} \\
 & + 0.01\textit{previous_physics_NoPhysics} \\
 & - 0.02\textit{previous_physics_HSPysics} + 0.004\textit{FCI_normalized_gain}
 \end{aligned}$$

As before, engineering is the reference category for concentration, first year is the reference category for college year, and college physics is the reference category for previous physics experience. We tested a null hypothesis that there is no difference in mean scores for those who did and those who did not evaluate their solutions successfully for assignment 6 when controlling for concentration, college year, previous physics experience, and FCI normalized gain in the AP50 class population. We found sufficient evidence of a mean difference ($t(df=80) = 2.417, p = 0.01$) in final course score in the class population for those who did and those who did not

evaluate their solutions successfully on assignment 6 when controlling for concentration, college year, previous physics experience, and FCI normalized gain. Therefore, we reject the null hypothesis of no mean difference. When controlling for the aforementioned variables, those who evaluated their solutions on assignment 6 have a mean difference in final score that is 7% higher than those who do not evaluate through reflection using knowledge. Plausible values for the association in the class population range from 1.3% and 13.7%, meaning we can be confident of a mean difference, and this linear regression model explains about 5% of the variability of the final score.

Similarly, we modeled a linear regression model for students who evaluated their solution successfully for assignment 12 controlling for concentration, college year, previous physics experience, and FCI normalized gain as shown below:

$$\begin{aligned}
 (5) \quad \widehat{score} = & 0.62 + 0.08evaluated_soln - 0.02concentration_LifeScience \\
 & - 0.04concentration_Other + 0.07concentration_Undeclared \\
 & + 0.04year_Sophomore + 0.04year_Junior + 0.09year_Senior \\
 & - 0.01previous_physics_NoPhysics - 0.06previous_physics_HSPhysics \\
 & + 0.01FCI_normalized_gain
 \end{aligned}$$

As with the previous models, engineering is the reference category for concentration, first year is the reference category for college year, and college physics is the reference category for previous physics experience. We tested a null hypothesis that there is no difference in mean scores for those who did and those who did not evaluate their solution for assignment 12 when controlling for concentration, college year, previous physics experience, and FCI normalized gain in the AP50 class population. We found evidence of a statistically significant mean difference ($t(df=66) =$

2.645, $p = 0.01$) in final course score in the class population for those who did and those who did not evaluate their solution successfully for assignment 12 when controlling for concentration, college year, previous physics experience, and FCI normalized gain, and thus, we reject the null hypothesis of no mean difference. When controlling for the aforementioned variables, those who evaluated their solution have a mean difference in final score that is 8% higher than those who do not evaluate their solution. Plausible values for the association in the class population range from 2.0% and 14%, meaning we can be confident of a mean difference. This linear regression model explains about 4% of the variability of the final score.

Furthermore, we examined how often students evaluated their solutions successfully using one of the five techniques outlined in the framework out of the six estimation problems we analyzed. We found that the mean frequency for students evaluating their solutions over the six assignments is 4.19 and the median is 4. Only one student did not evaluate any of their solutions successfully while 20% of students successfully evaluated all six of the assignments (see Figure 5.2).

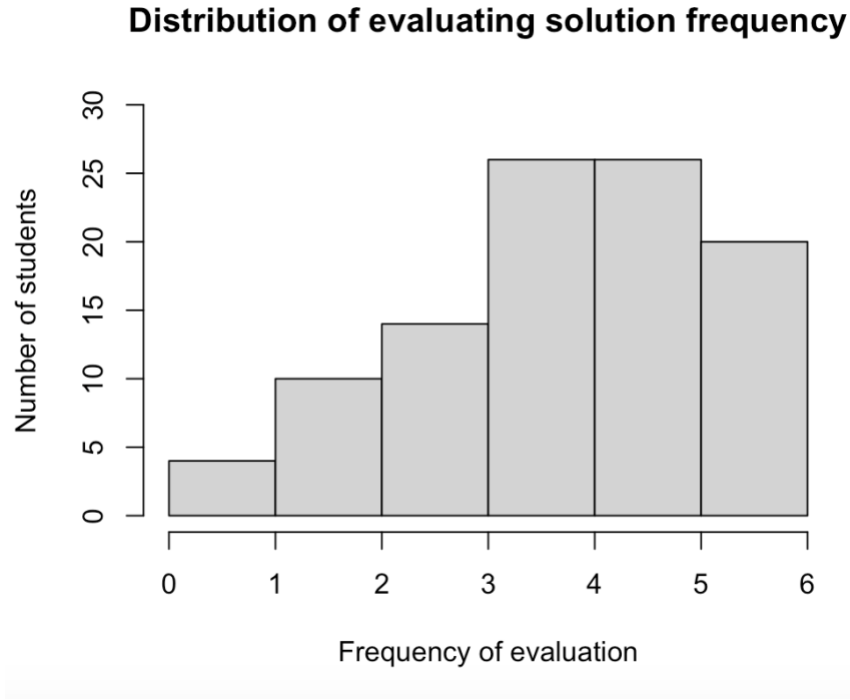


Figure 5.2: Histogram depicting the distribution of the frequency students successfully evaluated their solutions for the six assignments

To investigate the relationship between the frequency of students who evaluated their solutions successfully over the six assignments and their final course score we conducted a linear regression analysis controlling for concentration, college year, previous physics experience, and FCI normalized gain for each assignment. We fit the following population model:

$$\begin{aligned}
 (6) \quad score_i = & \beta_0 + \beta_1 freq_evaluated_soln_i + \beta_2 concentration_LifeScience_i \\
 & + \beta_3 concentration_Other_i + \beta_4 concentration_Undeclared_i \\
 & + \beta_5 year_Sophomore_i + \beta_6 year_Junior_i + \beta_7 year_Senior_i \\
 & + \beta_8 previous_physics_NoPhysics_i + \beta_9 previous_physics_HSPysics_i \\
 & + \beta_{10} FCI_normalized_gain_i + \varepsilon_i
 \end{aligned}$$

As with the previous models, engineering is the reference category for concentration, first year is the reference category for college year, and college physics is the reference category for previous physics experience. The fitted model is shown below:

$$\begin{aligned}
 (7) \quad \widehat{score} = & 0.51 + 0.03freq_evaluated_soln + 0.01concentration_LifeScience \\
 & + 0.01concentration_Other + 0.12concentration_Undeclared \\
 & + 0.03year_Sophomore + 0.02year_Junior + 0.01year_Senior \\
 & - 0.01previous_physics_NoPhysics - 0.03previous_physics_HSPhysics \\
 & + 0.01FCI_normalized_gain
 \end{aligned}$$

We tested a null hypothesis that there is no difference in mean scores for students with different frequencies of successful evaluation when controlling for concentration, college year, previous physics experience, and FCI normalized gain in the AP50 class population. We found sufficient evidence that frequency of successfully evaluating solutions was a significant predictor of course score ($t(df=85) = 4.365, p < 0.0001$) in final course scores in the class population, and therefore, we reject the null hypothesis of no difference in the mean scores. When controlling for concentration, college year, previous physics experience, and FCI normalized gain, we found that the frequency of successfully evaluating solutions is positively related to course score. Specifically, for every one-unit (assignment) increase in frequency of successfully evaluating solutions, course scores are expected to increase by 3%, holding all other variables constant. Plausible values for the association in the class population range from 1.8% and 4.8%, meaning we can be confident of a mean difference. The linear regression model explains about 15% of the variability of the final score.

We have found evidence that in AP50 there is an association between whether students successfully evaluate their solutions and their final score for certain assignments. Particularly, we have found that students who successfully evaluate their solutions for assignments 2, 6, and 12 are associated with a higher final score, and this association holds true even when controlling for concentration, college year, previous physics experience, and FCI normalized gain. It is possible that the successful evaluation of these specific assignments positively correlated with course scores for a few different reasons. Assignments 2, 6, and 12 may have necessitated a higher level of understanding of the course material or how to solve the problem so that students who were able to successfully evaluate their solutions may have a better understanding of the material and may perform better on other areas of the course as well. It is also possible that the individual work ethic played a role. Students who successfully evaluated their solutions consistently throughout the semester, as highlighted in assignments 2, 6, and 12, may be more conscientious with their performance in other areas of the course as well, leading to a higher overall course score. Moreover, we found that there is a positive relationship between the frequency of successfully evaluating solutions and course scores, indicating that as the frequency of successfully evaluating solutions increases, course scores also tend to increase when controlling for the aforementioned variables. This finding further suggests that a student's perseverance and consistency in their course work is, unsurprisingly, a key contributing factor to their course performance. Though, it is also possible that successful evaluation for certain assignments and frequency of successful evaluation are likewise important for academic performance in the course regardless of concentration, college year, previous physics experience, and FCI normalized gain.

It is important to note that we cannot address whether these associations are causal. While it is possible that students received credit on their assignments for following the instructions of applying the four-step problem-solving procedure, which includes the evaluation step, grades and the feedback students were given were effort-based and often only considered whether the evaluation step was included rather than the quality of it. It is possible, though, that the skills students developed in successfully evaluating their solutions transferred to other components of the course, leading them to be more successful in their overall academic performance in the class. Evaluation is an essential part of the physics problem-solving process, especially because physics is a subject that requires a deep understanding of abstract concepts and mathematical principles. Our findings suggest that students who are able to critically evaluate their own reasoning and problem-solving decisions are more likely to be successful in the course.

5.3.2 Trends in evaluation over the semester

To better understand how students evaluate their solutions over the semester, we first investigated if there was a change in whether students successfully evaluated their solutions by using one or a combination of any of the five evaluation techniques identified in the framework. We found that overall, there was no trend in whether students evaluated their solutions over the semester (see Figure 5.3). Though there is an increase in the percentage of students who successfully evaluate their solutions from assignment 2 to the subsequent assignments. We conducted a chi-square test to examine this trend and found that there was a statistically significant association between assignment and students who evaluated their solutions successfully ($\chi^2(5, N = 568) = 28.5, p < 0.001$). However, when assignment 2 is removed

from the data, we find that this statistically significant association goes away ($\chi^2(4, N = 467) = 9.09, p = 0.06$). These findings suggest that aside from assignment 2, there is no association between the assignment and students who evaluated their solutions successfully. It is possible that for assignment 2 students were still learning how to incorporate the four-step problem-solving procedure in their assignments and how to apply evaluation strategies in the final step. Moreover, for assignments after 3 there is even a slight decrease in the percentage of students who successfully evaluate their solutions, further suggesting that students are not more likely to incorporate successful evaluation strategies as the semester progresses.

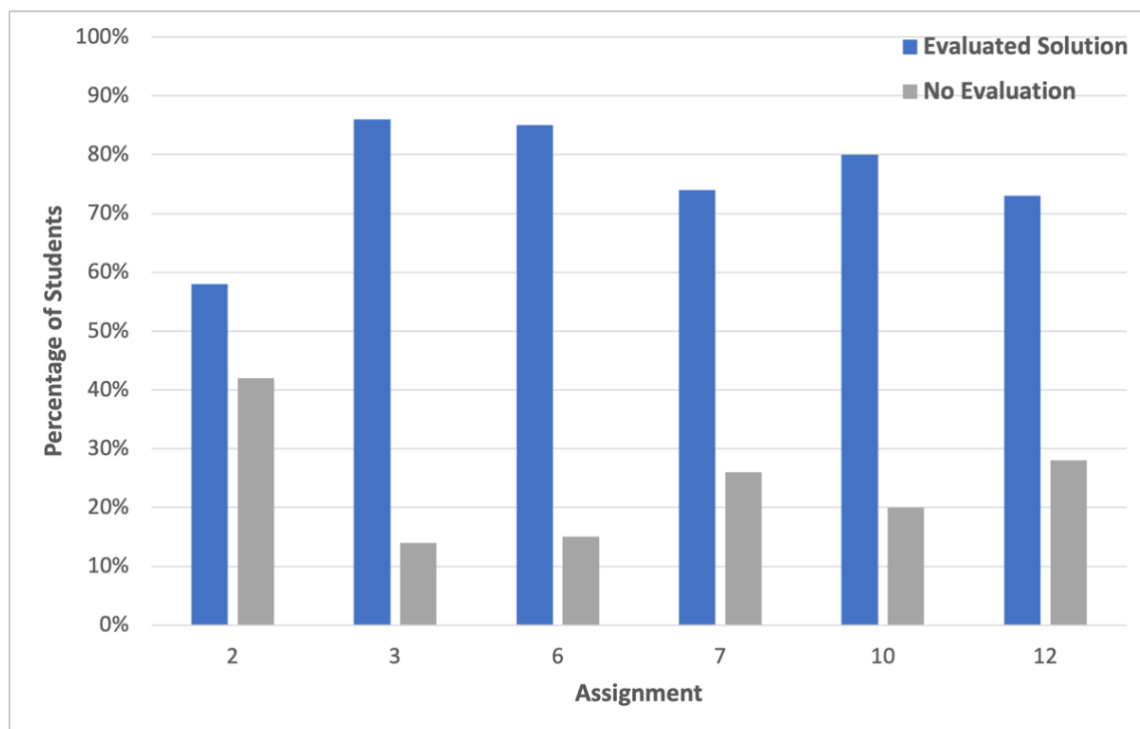


Figure 5.3: Percentage of students who successfully evaluated their solution by assignment.

We also found that there was no association between whether students came to a solution within plus or minus one order of magnitude from the solutions agreed upon by a set of experts

and whether they successfully evaluated their solutions ($\chi^2(1, N = 568) = 1.52, p = 0.22$).

And when we analyze this by assignment, there was also no association. Furthermore, there was no association between whether students applied valid physics concepts and equations to solve the problems and whether they had a substantial evaluation ($\chi^2(1, N = 568) = 1.15, p = 0.28$). Similarly, there was no association when analyzing this by assignment.

It is possible that we did not see any trends in students successfully evaluating their solutions over the semester along with no association in answer correctness or valid physics process because of the lack of specific feedback on students' evaluation step on the assignments. The assignments were also graded by different teaching fellows throughout the semester, so there may have been an inconsistency in grades and feedback given dependent on the teaching fellow. Moreover, the ability to effectively apply evaluation techniques in problem solving is characteristic of expert problem solvers and is thus challenging for introductory physics students. This is particularly true when students are expected to evaluate their solutions in new content material each week. To increase the instances of successful evaluation over the semester students should engage in additional scaffolded activities to practice this particular subcomponent of metacognition along with targeted feedback for improvement.

While we did not find any trends in whether students successfully evaluated their solutions over the semester, we did find trends in the specific evaluation techniques students used for the particular estimation problem. We first investigated whether there was an association between assignments and students who evaluated their solutions through reflection using problem definitions and assumptions. We conducted a chi-square test to examine this trend and found that there was a statistically significant association between assignment and students reflecting using problem definitions and assumptions ($\chi^2(5, N = 568) = 37.9, p <$

0.0001). While less than one third of students used this evaluation technique for each assignment, more students used this reflection strategy for assignments 2 (29%) and 3 (25%) whereas no students did for assignment 10 (see Figure 5.4).

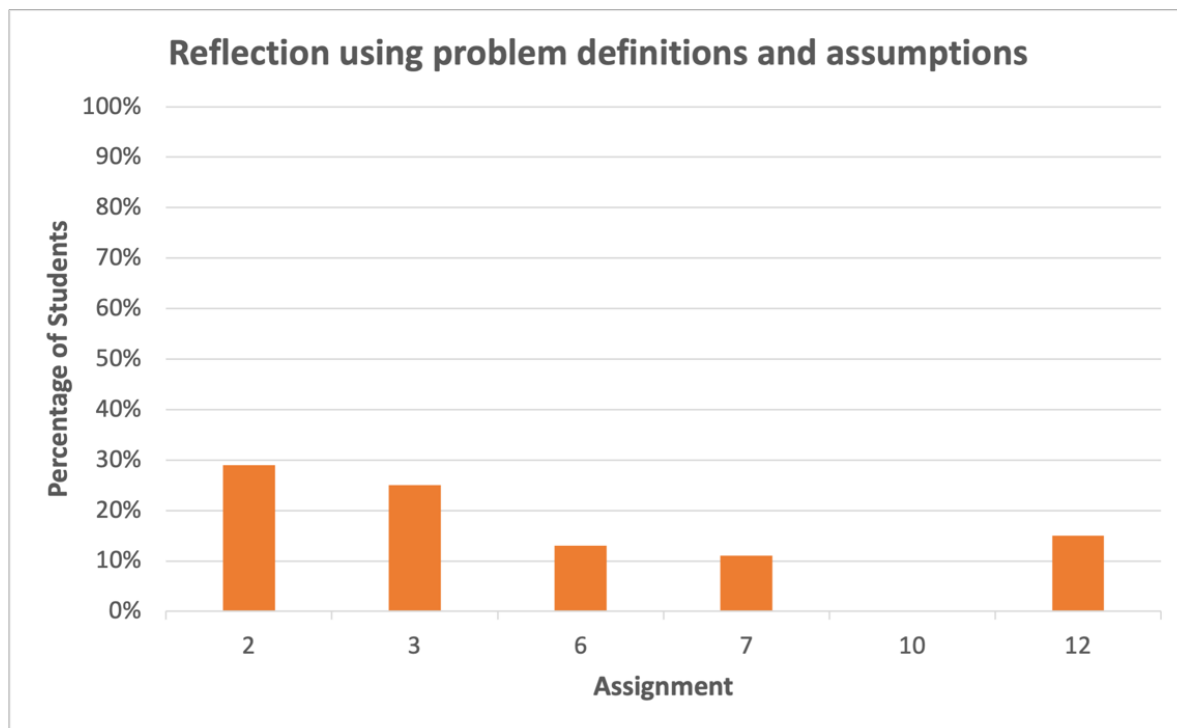


Figure 5.4: Percentage of students who evaluated their solution through reflection using problem definitions and assumptions by assignment

The estimation problem for assignment 2 required students to calculate the total distance all human population in history have ever walked. To do so requires students to approximate how many humans have ever lived as well as make assumptions about how far each human has walked in a day and how many days they have lived. For example, one student stated in their evaluation,

“Honestly, my answer seems kind of low given that it’s supposed to account for every human in history. However, realizing every human has a different lifespan length and could walk substantially more or substantially less than my estimate value, my answer

isn't too far-fetched. The presence of alternative modes of travel now can serve to offset the more walking-heavy lifestyles of early human hunter-gatherers.”

The estimation problem for assignment 3 involved students finding the final speed of a penny dropped from the Eiffel Tower. In their evaluations, many students reflected on the assumption that they did not include air resistance, which affected their final answer. For example, one student stated that they were ignoring air resistance in their getting started step of the four-step problem-solving procedure and defined the gravitational acceleration to be 10 m/s^2 , and they reflected on this and how it impacted their result in their evaluation:

“This answer makes sense, and my assumptions are reasonable. However, I could have used 9.8 m/s^2 rather than 10, and the penny may be affected by air resistance, both of which would result in a decreased final speed.”

For the estimation problem for assignment 10, however, no students used the evaluation strategy of reflection using problem definitions and assumptions. This problem asked the solver to find the change in angular velocity of the earth if every human on it took a step in the same direction at the same time. Solving this problem requires the values of the mass of the earth, the radius of the earth, the current human population, and the velocity for one step. Most students looked up these values or referenced knowing them from previous course experiences, which is a possible reason why they did not reflect using problem definitions and assumptions in the evaluation for this problem.

We also investigated whether there was an association between assignment and students who evaluated their solutions through reflection using knowledge. We conducted a chi-square test to examine this trend and found that there was a statistically significant association between assignment and students reflecting using knowledge ($\chi^2(5, N = 568) = 22.9, p < 0.001$). Less than a fourth of students used this evaluation technique for each assignment, but more students

used this reflection strategy for assignments 12 (20%) and 7 (18%) whereas no students did for assignment 2 (see Figure 5.5).



Figure 5.5: Percentage of students who evaluated their solution through reflection using knowledge by assignment

The estimation problem for assignment 12 requires students to determine the number of seats in a theater that would hear a reduced sound level from two clarinets playing at the exact same tone. In the evaluation, many students referenced their experience playing an instrument or listening to a concert, and they reflected on the reasonableness of their solution based on this knowledge. One student specifically references their expectation of the solution based on their experience:

“This answer makes sense because honestly I would not expect too much variations of sound in the theater. From experience if a band plays together standing close to each other, there is little reduced sound around a theater. The source separation has to be way

smaller to experience more seats with reduced sound (due to more nodal lines).”

Similarly, another student addresses the practicality of the solution by expressing that a larger portion of seats experiencing reduced sound would be unfavorable:

“Should be a low proportion of seats, or else Sanders would be a bad place to listen to performances.”

As with assignment 12, the estimation problem for assignment 7 involves a practical situation that most students have experience with or are familiar with. For assignment 7, the estimation problem involves calculating the impulse a hammer accidentally delivers to a thumb after a missed swing to a nail. Students are able to evaluate their solution based on their knowledge of a similar level of pain or through comparison to a similar type of impact. For example, one student compares the impulse the hammer delivers to a thumb to that of a baseball bat to a baseball:

“My solution seems to be within a reasonable range. I looked up the impulse of a baseball and bat interaction, and it is about 13 Ns or 30% larger than my answer. I would expect that a hammer delivers less than a baseball and bat, but my answer is surely within an order of magnitude.”

For the estimation problem of assignment 2, many students struggled to conceptualize the large order of magnitude of the solution for the total combined distance ever walked in human history. Many students do not have experience or knowledge of such a large value they can compare their solution to and thus were unable to evaluate their solution through reflection using knowledge. An example of such struggle is as follows:

“Frankly, this final answer is so large I can’t quite comprehend whether or not it seems reasonable (given that the scale of the problem is also huge).”

This response in the evaluation was echoed by many students, and it is possible that since it was one of the first assignments, they struggled to successfully evaluate their solution using any of the strategies outlined in the framework.

Furthermore, we investigated whether there was an association between assignment and students who evaluated their solutions through reflection using strategy. We conducted a chi-square test to examine this trend and found that there was no statistically significant association between assignment and students reflecting using strategy ($\chi^2(5, N = 568) = 6.30, p < 0.28$). This evaluation technique was rarely used on any of the assignments (see Figure 5.6), with 7% of students on assignment 10 being the most this technique was used. It is possible that most students did not evaluate the solution through reflection using strategy as it was challenging for them to specifically identify how their strategy contributed to the reasonableness of their solution beyond stating that their “strategy makes sense, so the solution makes sense”. It is also possible, though, that if in the evaluate solution step in their problem solving, they released their strategy was incorrect, then they reattempted the problem with a new approach and a different evaluation.



Figure 5.6: Percentage of students who evaluated their solution through reflection using strategy by assignment

As with the other reflection techniques, we investigated whether there was an association between assignments and students who evaluated their solutions through reflection using solution. We conducted a chi-square test to examine this trend and found that there was a statistically significant association between assignment and students reflecting using solution ($\chi^2(5, N = 568) = 36.3, p < 0.0001$). Students were more likely to engage in this evaluation technique than any of the other four techniques with students particularly using this technique for assignments 10 (73%) and 6 (68%) (see Figure 5.7).

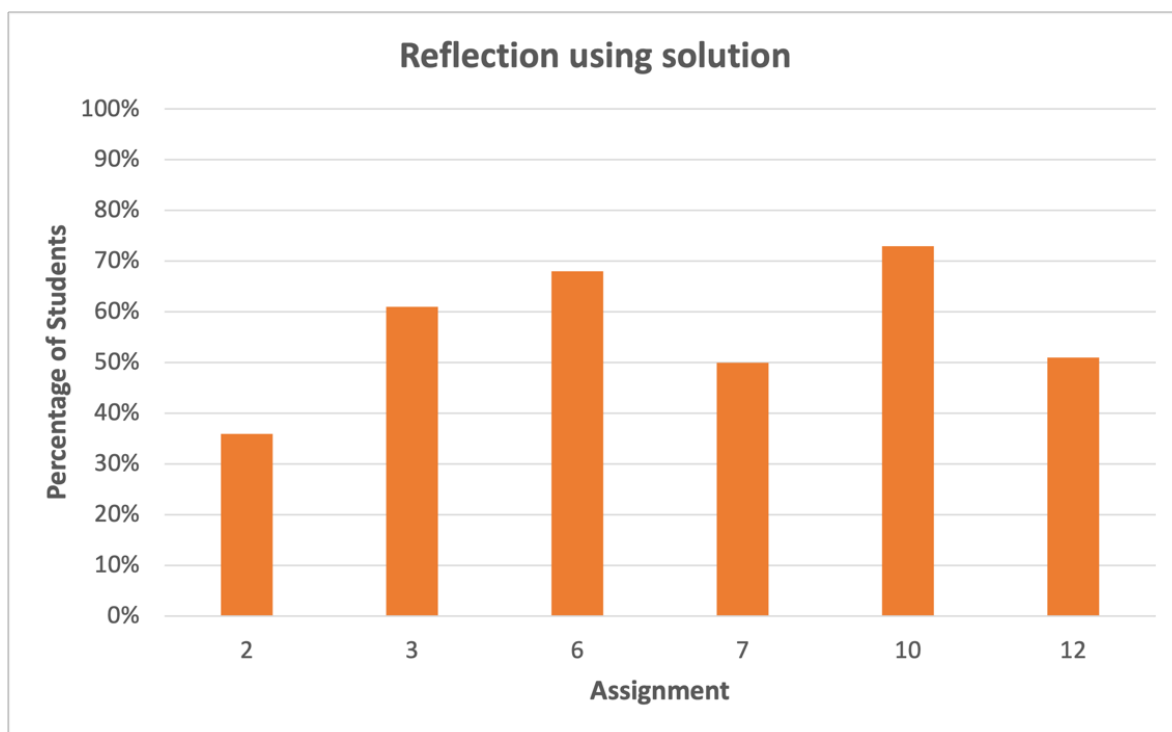


Figure 5.7: Percentage of students who evaluated their solution through reflection using solution by assignment

Reflection using solution seemed to be the most straightforward evaluation technique for students, especially if they did not have experience with the problem content or struggled to connect how their assumptions or strategy specifically affected the solution. For assignment 10

where students are instructed to find the change in angular velocity of the earth if everybody took a step in the same direction at the same time, the final answer is of the order of magnitude of -20 . Such a small value is very challenging for many students to conceptualize or compare to something familiar. Many students thus evaluate the reasonableness of the solution by recognizing the relative order of magnitude, that is, the value should be very small because the earth and its inertia is so large. For example, one student stated,

“This is an insignificant change in velocity, which makes sense because the Earth’s mass/inertia is so great.”

For assignment 6, students are asked to find the amount of energy required to raise a 4,000-kilometer oil pipeline 1 meter above the ground. The final answer is of the order of magnitude of 10, which, like assignment 10, is difficult for students to conceptualize. Similarly, students tended to evaluate their solution by identifying that the answer should be large because the pipeline is very long. Many students provided evaluations similar to the following:

“[Answer] appears to be a reasonable order of magnitude for the energy required to move up the pipeline 1 m because it is a significant amount of energy and demonstrates the significant undertaking needed to raise the pipeline.”

While this strategy of reflection using solution is not the most insightful evaluation technique, it is more thoughtful than students simply stating that the answer “makes sense” or “seems reasonable” without further justification. Moreover, reflection using solution may be used as an evaluation technique for more novice students as a starting place to later include more advanced techniques as they become more proficient at problem-solving.

In addition to the reflection techniques, we investigated whether there was an association between assignments and students who evaluated their solutions through mathematical answer-checking strategies. We conducted a chi-square test to examine this trend and found that there was a statistically significant association between assignment and students using mathematical

answer-checking strategies ($\chi^2(5, N = 568) = 28.4, p < 0.0001$). Less than one-fourth of students used this evaluation technique for each assignment, but more students used this strategy for assignments 3 (24%) and 7 (13%) whereas few students did for assignment 10 (3%) (see Figure 5.8).

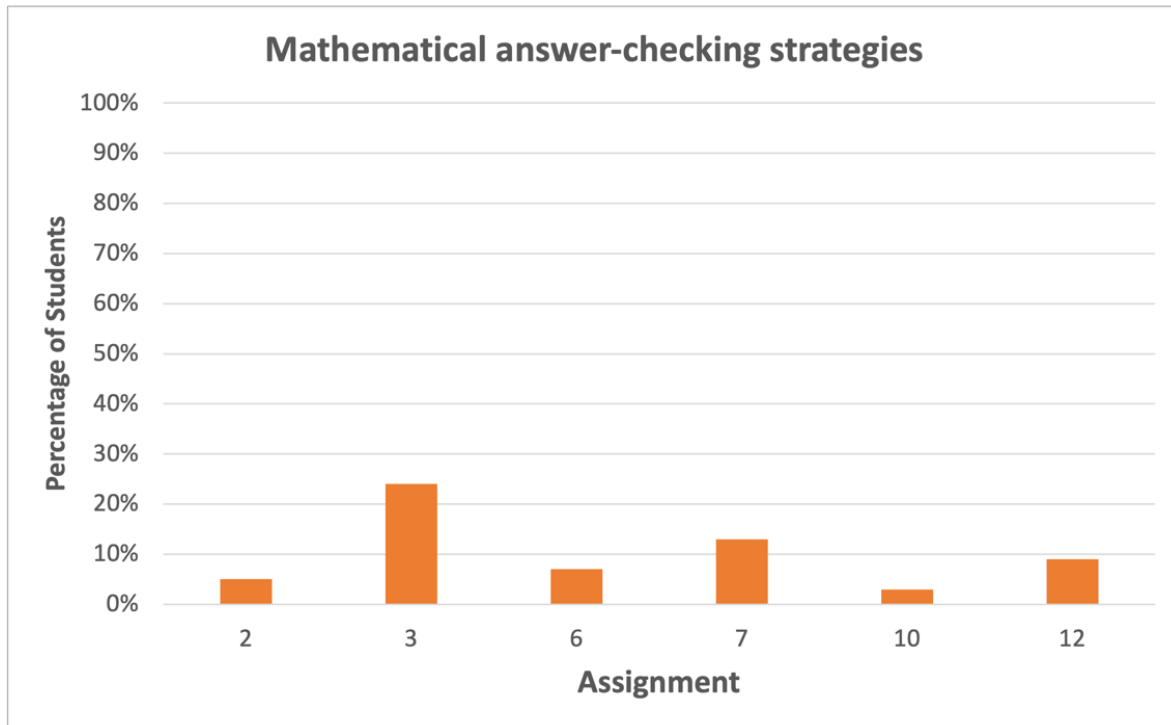


Figure 5.8: Percentage of students who evaluated their solution through mathematical answer-checking strategies by assignment

The estimation problem for assignment 3 involves determining the final velocity of a penny dropped from the Eiffel Tower. To check for the reasonableness of their solution, some students calculated the time it would take to reach this final velocity, which was a calculation students were already familiar with. Similarly, for the estimation problem for assignment 7, students used math answer-checking strategies in their evaluation by calculating the impulse using a different method than their original attempt, that is, using $J = F\Delta t$ if they originally

calculated impulse with $J = m\Delta v$ or vice-versa. For both assignments, using mathematical answer-checking strategies was reasonably straightforward as it involved incorporating a single additional equation that students were familiar with. For the estimation problem for assignment 10, however, only 3% of students used this evaluation technique. It is likely that students did not engage in mathematical answer-checking strategies in their evaluation for this problem because it would involve many additional steps, and the procedure for determining the energy for lifting an oil pipeline was already not straightforward for many students in their original attempt.

Furthermore, in the estimation problem for assignment 3, some students evaluated their solution through the math answer-checking strategy of converting the units of their solution to something more familiar. In this problem students typically solve for a final answer with the S.I. units of m/s. To better understand this value, many students convert their answers to mph. For example, one student stated this as a way to check the reasonableness of the solution and included the unit conversion calculation:

“To evaluate my solution, I will convert the velocity to miles/hr to see the plausibility of my answer

$$\frac{77.5 \text{ m}}{\text{s}} \cdot \frac{1 \text{ mile}}{1609.344 \text{ m}} \cdot \frac{3600 \text{ s}}{\text{hr}} = \frac{173.36 \text{ miles}}{\text{hr}} .”$$

This student added a check mark after their unit conversion, suggesting they were satisfied with the reasonableness of the converted solution. While converting the units of the final answer to something that is more readily understood is a useful technique for evaluation, it is more successful when used in conjunction with a statement as to why this converted value does or does not make sense.

Often a more insightful evaluation involved using more than one evaluation technique outlined in the evaluation framework. Thus, we investigated whether there was an association between assignments and students who used more than one evaluation technique. We conducted

a chi-square test to examine this trend and found that there was a statistically significant association between assignment and students using more than one evaluation technique ($\chi^2(5, N = 568) = 16, p < 0.01$). Using more than one evaluation technique was most common for assignment 3 (32%) (see Figure 5.9). This is unsurprising as this was the assignment where students were more likely to use mathematical answer-checking strategies. Math answer checking-strategies included calculating another value to compare their solution to and converting the units of the final answer to something more familiar. This evaluation is more meaningful when coupled with a statement of why this new value is reasonable, and thus many students also included a statement justifying the new value through reflection using knowledge or solution when using math checking strategies.

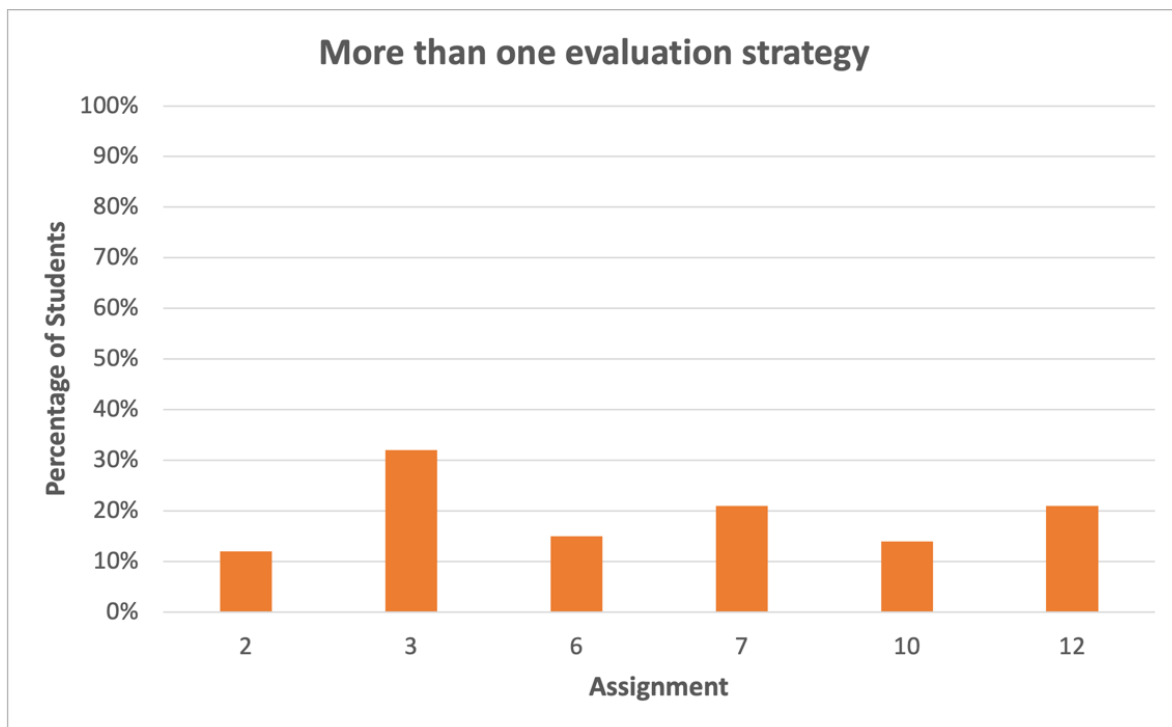


Figure 5.9: Percentage of students who evaluated their solution using more than one evaluation technique by assignment

We have found that there are statistically significant associations between the assignments and the evaluation techniques of reflection using problem definitions and assumptions, reflection using knowledge, reflection using solution, and mathematical answer-checking strategies. That is, students are more likely to use certain evaluation techniques for certain estimation problems. This suggests that evaluation is task-dependent, and certain problems lend themselves to be evaluated using a particular strategy. Furthermore, we found that there is a significant association between the assignments and students using more than one evaluation technique. This is likely the result of students using reflection techniques along with mathematical answer-checking strategies in assignment 3 to evaluate their solution successfully. It is possible, though, that students are limited to applying certain evaluation techniques to a particular problem depending on their familiarity or experience with the physics content as compared to how experts would evaluate their solutions. We thus further investigate whether how students evaluate their solutions is associated with the difficulty of the problem.

To define the difficulty of the six estimation problems, we first binned students' answers by the order of magnitude (OoM) into three groups for each assignment: 0 to ± 1 OoM from the correct answer, ± 2 to ± 4 OoM from the correct answer, and $> \pm 4$ OoM from the correct answer. This order of magnitude breakdown of students' solutions by assignment is summarized in Figure 5.10.

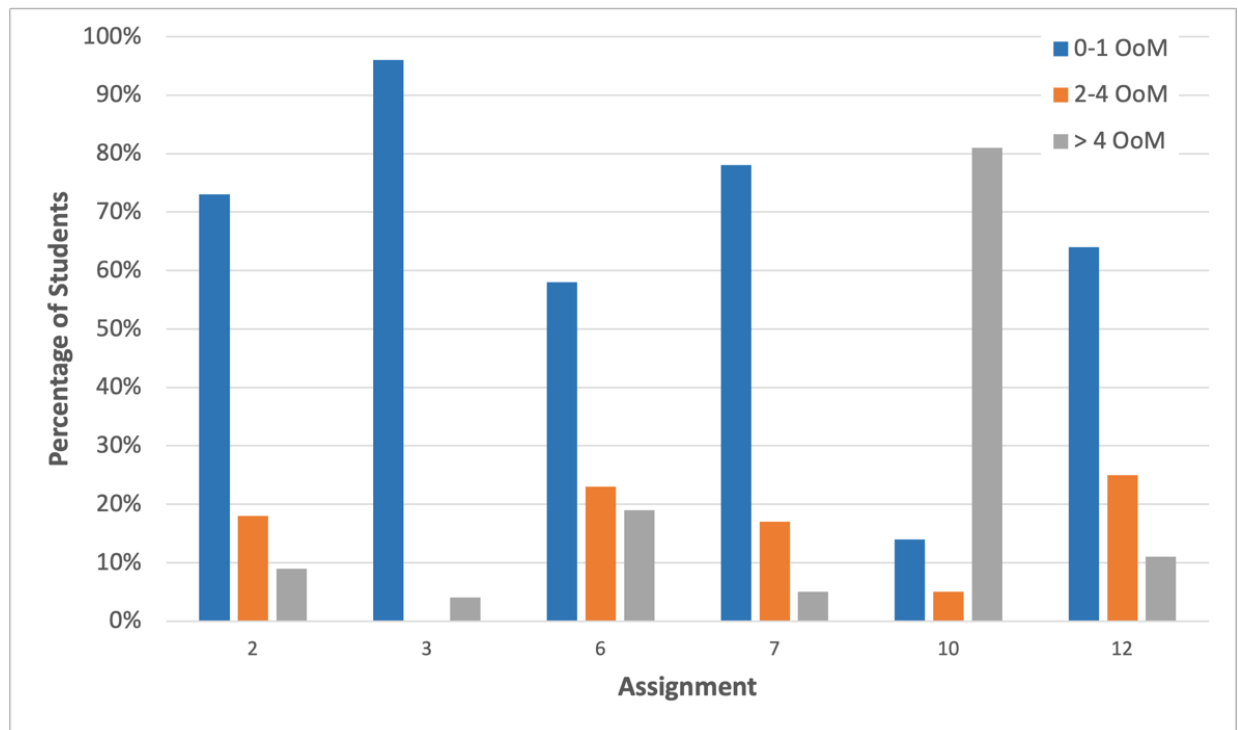


Figure 5.10: Percentage of students whose answers were 0 to ± 1 OoM, ± 2 to ± 4 OoM, or $> \pm 4$ OoM from the correct answer for each assignment

From this, we grouped the assignments by problem difficulty based on how many students' answers were in the 0 to ± 1 OoM from the correct answer group. For the estimation problem for assignment 3, most students answered within ± 1 OoM from the correct answer, so this problem is categorized as "easy". For assignment 10, most students' answers were $> \pm 4$ OoM from the correct answer while less than 15% of student answers were in the 0 to ± 1 OoM from the correct answer group, thus this problem was as categorized as "hard". While the remaining assignments (2,6,7, and 12) were categorized as "medium" difficulty.

Using this categorization of problem difficulty, we investigated whether there was an association between problem difficulty and students who successfully evaluated their solutions by using one or a combination of any of the five evaluation techniques identified in the framework. We conducted a chi-square test to examine this trend and found that there was a

statistically significant association between problem difficulty and students who successfully evaluated their solutions ($\chi^2(2, N = 568) = 9.87, p = 0.001$). Unsurprisingly, more students were able to successfully evaluate their solutions using one or a combination of the five evaluation techniques if the problem was categorized as easy (86%) compared to medium (72%) or hard (80%) (see Figure 5.11). However, it is surprising that students were more likely to evaluate their solutions for hard problems compared to medium problems. As previously mentioned, it is possible that whether students successfully evaluate their solutions or not is affected by the lack of or inconsistency in feedback, resulting in students having unclear expectations as to what constitutes as successful evaluation strategies. Moreover, this difference in whether students evaluate their solutions by problem difficulty may be affected by the specific content of the problems and students' level of familiarity with it.

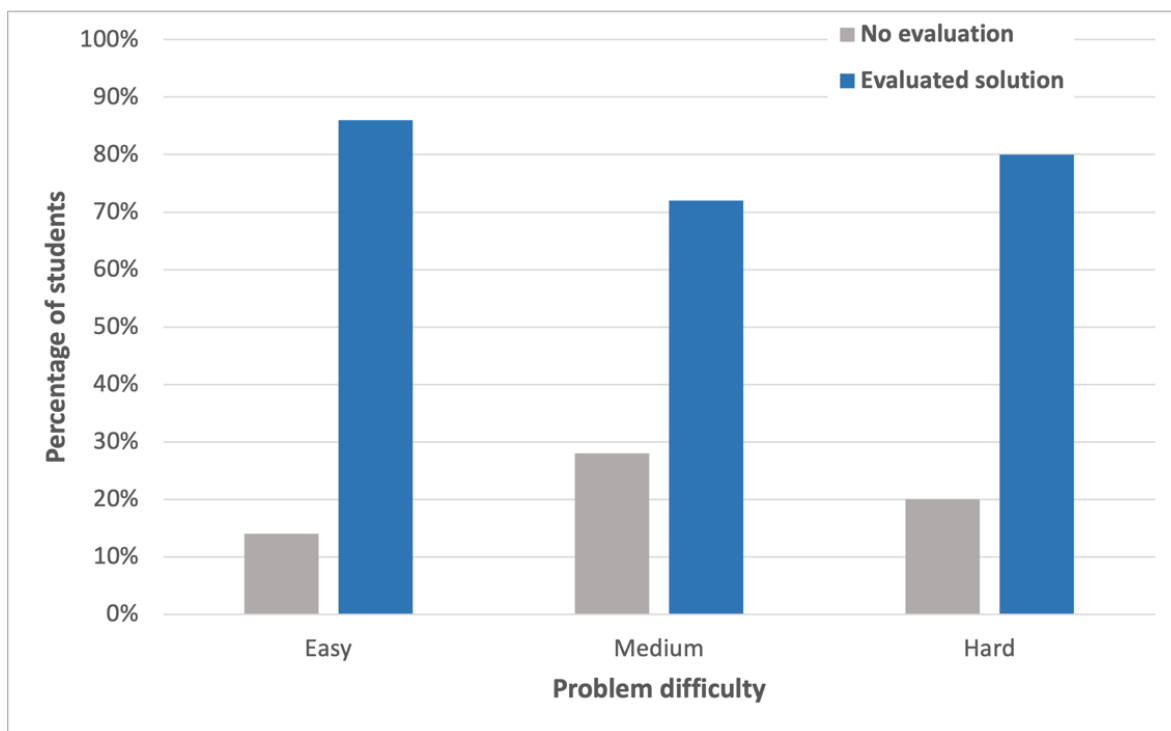


Figure 5.11: Percentage of students who successfully evaluated their solution by problem difficulty

To further understand trends in student evaluation by problem difficulty, we analyzed how students specifically evaluated their solutions by problem difficulty. We first investigated whether there was an association between problem difficulty and students who evaluated their solutions through reflection using problem definitions and assumptions. We examined this trend using a chi-square test and found that there was a statistically significant association ($\chi^2(2, N = 568) = 23.5, p < 0.0001$). Though this evaluation strategy was not widely used by students, we found that more students evaluated their solution through reflection using problem definitions and assumptions for the easy problem (25%) whereas no students used this evaluation technique for the more difficult problem (see Figure 5.12).

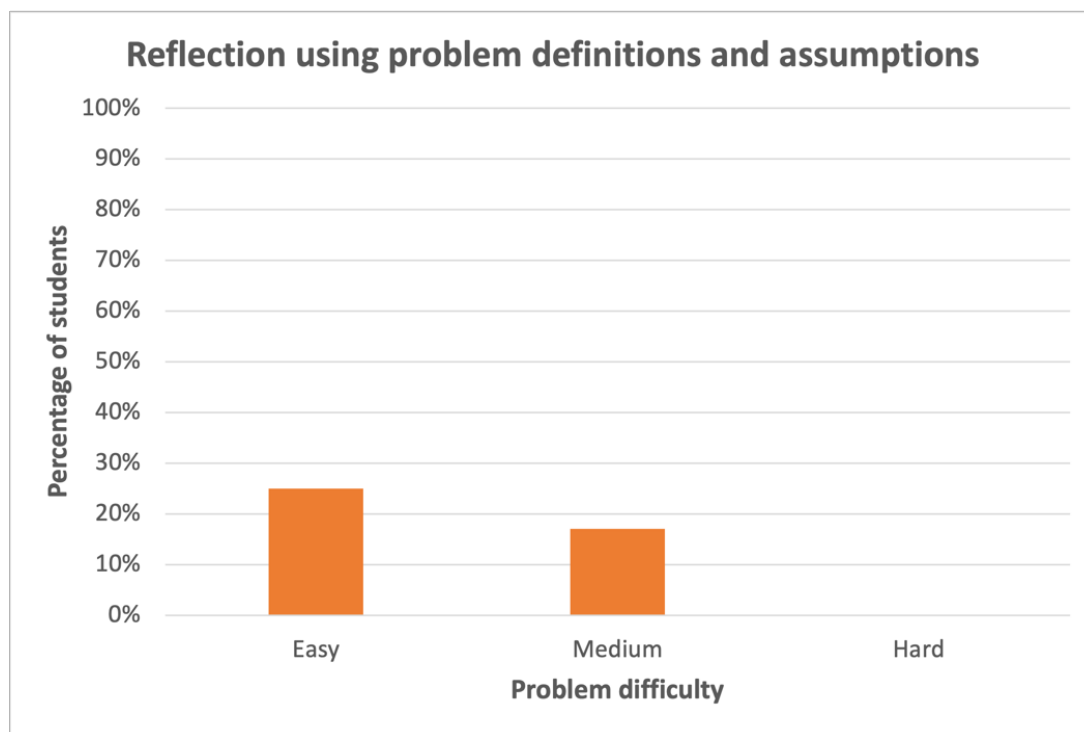


Figure 5.12: Percentage of students who evaluated their solution through reflection using problem definitions and assumptions by problem difficulty

Similarly, we found that the association between problem difficulty and the evaluation technique of using mathematical answer-checking strategies was statistically significant when conducting a chi-square test ($\chi^2(2, N = 568) = 24.9, p < 0.0001$). This evaluation strategy was also used by students more on the easy problem (24%) compared to the medium (8%) and hard (3%) problems (see Figure 5.13). Furthermore, we investigated whether there was an association between problem difficulty and students who evaluated their solutions using more than one evaluation technique. We found that this association was statistically significant using a chi-square test ($\chi^2(2, N = 568) = 12.3, p < 0.01$), and more students evaluated their solution using more than one technique for the easy problem (31%) than the medium (17%) and hard (14%) problems (see Figure 5.14).

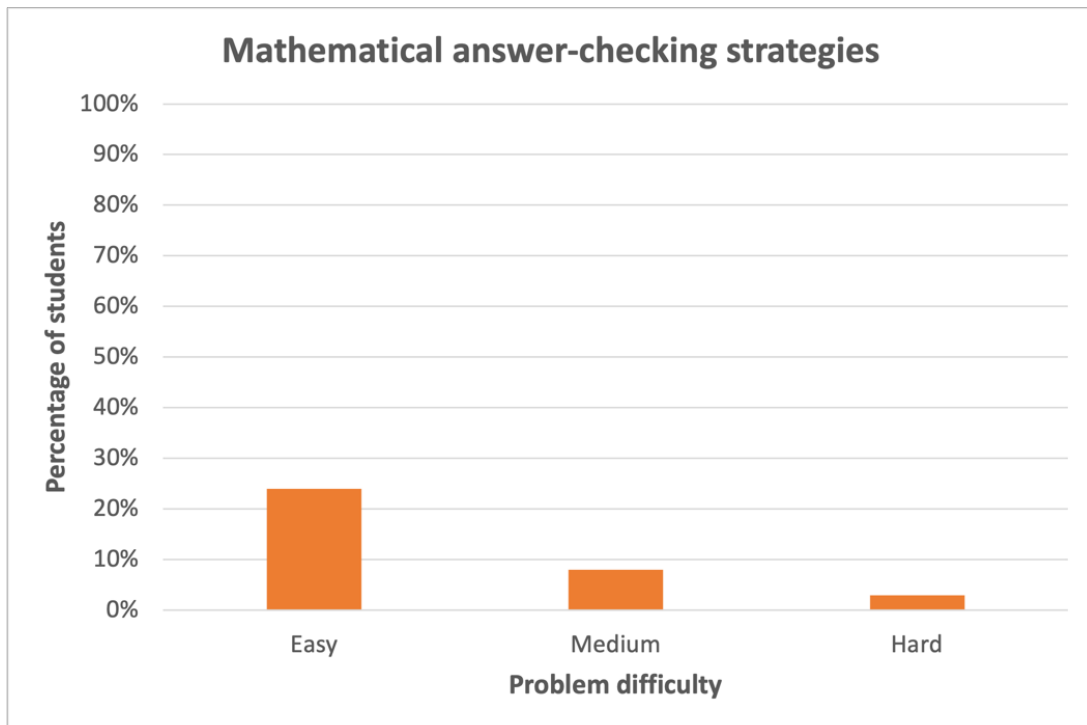


Figure 5.13: Percentage of students who evaluated their solution through mathematical answer-checking strategies by problem difficulty



Figure 5.14: Percentage of students who evaluated their solution through more than one evaluation strategy by problem difficulty

When investigating whether there is an association between problem difficulty and the evaluation technique of reflection using solution, through a chi-square analysis we found that there was a statistically significant association ($\chi^2(2, N = 568) = 15.1, p < 0.001$). Opposite to the previous findings by problem difficulty, we found that more students used this evaluation strategy for the hard problem (73%) than for either the easy (61%) or medium (51%) problems (see Figure 5.15). No statistically significant association was found between problem difficulty and whether students evaluated their solutions through reflection using knowledge ($\chi^2(2, N = 568) = 0.61, p = 0.74$) or reflection using strategy ($\chi^2(2, N = 568) = 4.87, p < 0.10$).

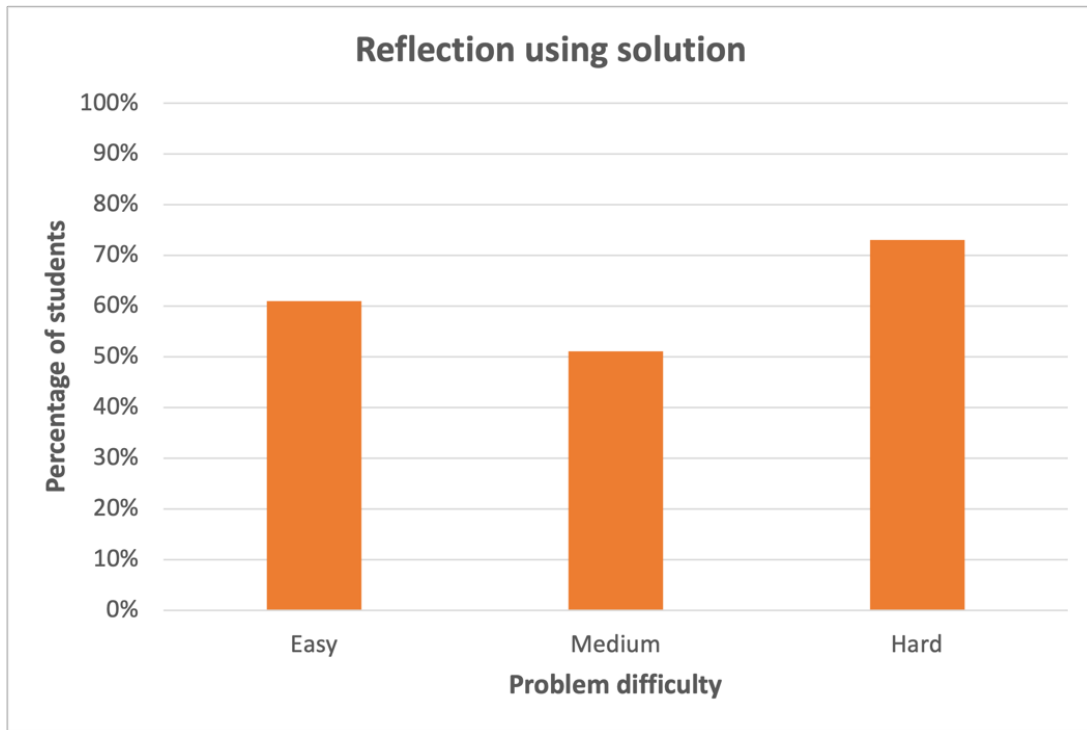


Figure 5.15: Percentage of students who evaluated their solution through reflection using solution by problem difficulty

When we categorize the estimation problems by difficulty level, we found that more students successfully evaluate their solutions for the easy problem, and when breaking down the evaluation by specific techniques, we found that more students evaluate their solutions through reflection using problem definitions and assumptions as well as through mathematical answer-checking techniques for the easy problem. Similarly, we found that for the easy problem, more students evaluate their solution using more than one technique than for the medium and hard problems. It is possible that students are more familiar with the equations and concepts needed to solve the easy problem, which led them to answer the question within ± 1 OoM of the correct answer and are thus more likely to successfully evaluate their solution using specific techniques. On the other hand, more students evaluated their solutions through reflection using solution on the hard problem compared to the easy and medium difficulty problems. As previously

addressed, reflection using solution is the most straightforward evaluation technique for students to use, especially if they did not have as much familiarity with the problem content or struggled to conceptualize the reasonableness of their final value beyond its relative magnitude.

5.3.3 Differences in evaluation ability by concentration

While most of the students had high school level physics previous to this course (89%), the course population varied in student concentration. Most of the students were either life-science (49%) or engineering concentrators (29%) with a smaller portion of students in other disciplines (14%) like economics or history of science or had not yet declared their concentration (8%). We suspected that there is a difference in previous STEM courses depending on students' concentration, which may affect their ability to successfully evaluate their solutions. We conducted a one-way analysis of variance (ANOVA) to compare the means of the final course score among the different concentrations represented in the class, and we found that there was no statistically significant difference among the means of the groups ($F(3,96) = 1.152, p = 0.332$). We also conducted a one-way ANOVA to compare the means of the FCI normalized gain among the different concentrations, and we found there was no statistically significant difference among the means of the groups ($F(3,92) = 1.294, p = 0.281$). Additionally, we investigated whether there is an association between whether students' answers for the estimation problems were correct (within plus or minus one order of magnitude from the correct solution) and student concentration using a chi-square analysis, and we found that there was no statistically significant association ($\chi^2(3, N = 568) = 1.25, p = 0.74$). Likewise, we examined whether there was an association between whether students used valid physics concepts and equations to solve the

estimation problems and their concentration. Through a chi-square analysis, we found that there also was not a statistically significant association ($\chi^2(3, N = 568) = 2.47, p = 0.48$).

To investigate whether there was a relationship between student concentration and whether they successfully evaluated their solution we conducted a three-level multilevel (or hierarchical) analysis. We fit the following population model:

$$\begin{aligned}
 (8) \quad & \log \left(\frac{P(\text{evaluated_soln} = 1)_{ijk}}{1 - P(\text{evaluated_soln} = 1)_{ijk}} \right) \\
 &= \beta_{0j} + \beta_1 \text{concentration_LifeScience}_{ijk} + \beta_2 \text{concentration_Other}_{ijk} \\
 &+ \beta_3 \text{concentration_Undeclared}_{ijk} + \beta_4 \text{year_Sophomore}_{ij} + \beta_5 \text{year_Junior}_{ij} \\
 &+ \beta_6 \text{year_Senior}_{ij} + \beta_7 \text{previous_physics_NoPhysics}_{ij} \\
 &+ \beta_8 \text{previous_physics_HSPHysics}_{ij} + \zeta_{0j} + \varepsilon_{ijk}; \\
 &\zeta_{0j} \sim N(0, \sigma_\zeta^2); \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)
 \end{aligned}$$

where engineering is the reference category for concentration, first year is the reference category for college year, and colleges physics is the reference category for previous physics experience. At level 1 of the model is the outcome variable of whether or not a student successfully evaluates their solution. The fixed effect variable at this level is student concentration. At level 2 is the student level, where the fixed effect variable is college year and previous physics experience while the random effect variables include the variability in intercept across students, which captures the fact that some students may be more or less likely to evaluate their solution regardless of the specific assignment they are given. At level 3 of the model is the assignment level where the random effect variable includes the variability in intercept across assignments, which captures the fact that some assignments may be more or less likely to elicit student

evaluation regardless of the students. In the model above ζ_{0j} represents the random effect of assignment intercepts, and ε_{ijk} is the residual error.

The fixed effects are summarized in Table 5.3. The fixed effect of life science concentrators in reference to engineering concentrators was the only statistically significant predictor of whether students evaluate their solutions successfully or not ($\beta_1 = -0.85$, $SE = 0.35$, $p = 0.017$), indicating that life science concentrators are less likely to successfully evaluate their solutions than engineering concentrators regardless of the assignment. Specifically, life science concentrators differ by -0.85 in the log of the odds of evaluating solutions successfully from engineering students when controlling for college year and previous physics experience. We estimate that the odds for life science concentrators to evaluate their solutions is about $\exp(-0.85) = 0.427$ that of the odds for engineering students when controlling for college year and previous physics experience. This is a difference of about 9% between engineering and life science concentrators, where life science students are less likely to successfully evaluate their solutions when controlling for the aforementioned variables. We did not find that there was a statistically significant difference in whether students evaluated their solution for other or undeclared concentrations compared to engineering concentrators when controlling for college year and previous physics experience.

Table 5.4: Summary of fixed effects estimates from multi-level analysis investigating the relationship between whether students successfully evaluated their solutions and concentration controlling for college year and previous physics experience

	Estimate	SE	z-value	p-value
(Intercept)	2.421	0.767	3.158	0.002
Life science concentration	-0.848	0.354	-2.395	0.017
Other concentration	-0.713	0.466	-1.531	0.123
Undeclared concentration	-0.632	0.486	-1.301	0.193
Sophomore	-0.459	0.4622	-0.993	0.321
Junior	0.306	0.520	0.588	0.556
Senior	-0.168	0.649	-0.259	0.796
No previous physics	-0.372	0.809	-0.460	0.645
High school previous physics	-0.464	0.598	--0/776	0.438

The variance components for the random effects were estimated to be 0.411 for individual students with a standard deviation of 0.650. This indicates significant variability across students in the probability of successful solution evaluation. The variance for assignments was 0.254 with a standard deviation of 0.504. Similarly, this suggests significant variability across assignments in the probability of students successfully evaluating their solutions. We acknowledge the limitations of the model in the absence of residuals for the random effects. The lack of residuals may have impacted the reliability of the estimates and the conclusions that can be drawn from the analysis. We conducted sensitivity analysis by varying the priors and estimation algorithms, which produced consistent results.

When we focus on the relationship between student concentration and the specific evaluation techniques they use, we do not find that there is a statistically significant difference between any of the specific evaluation strategies and student concentration when controlling for

college year and previous physics experience. We do, however, find that there is a significant effect of concentration on the frequency students evaluate their solutions over the six assignments analyzed. To investigate this relationship, we fit the following population model:

$$\begin{aligned}
 (9) \quad freq_evaluated_soln_i &= \beta_0 + \beta_1 concentration_LifeScience_i + \beta_2 concentration_Other_i \\
 &+ \beta_3 concentration_Undeclared_i + \beta_4 year_Sophomore_i + \beta_5 year_Junior_i \\
 &+ \beta_6 year_Senior_i + \beta_7 previous_physics_NoPhysics_i \\
 &+ \beta_8 previous_physics_HSPHysics_i + \varepsilon_i
 \end{aligned}$$

where, as before, engineering is the reference category for concentration, first year is the reference category for college year, and colleges physics is the reference category for previous physics experience. The fitted model is shown below:

$$\begin{aligned}
 (10) \quad freq_evaluated_soln &= 4.77 - 0.79concentration_LifeScience - 0.89concentration_Other \\
 &- 0.52concentration_Undeclared - 0.44year_Sophomore \\
 &+ 0.31year_Junior - 0.39year_Senior + 0.02previous_physics_NoPhysics \\
 &+ 0.08previous_physics_HSPHysics
 \end{aligned}$$

From this linear regression model we find that life science concentrators differ by -0.79 in the frequency of successful solution evaluation from engineering concentrators when controlling for college year and previous physics experience. We estimate that engineering students on average tend to successfully evaluate closer to 5 of their 6 solutions whereas life science students on average tend to evaluate closer to 4 of their 6 solutions successfully in the class population when

controlling for college year and previous physics experience ($t(df=91) = -1.945, p=0.05$). We can therefore reject the null hypothesis that there is no difference in the average number of solutions successfully evaluated by the two groups when controlling for the above-mentioned variables. Plausible values for the association in the class population range from -1.59 to 0.02, meaning we can be confident of the mean difference in frequency of solutions successfully evaluated by the two groups. The linear regression model explains only about 0.20% of the variability of the frequency of successful evaluation, suggesting that there are many other factors that contribute to the frequency of successful evaluation.

Through a multilevel logistic regression model, we have found that there is a statistically significant difference between engineering concentrators and life science concentrators in whether they successfully evaluate their solutions when controlling for college year and previous physics experience. To further support these findings, we have found through a linear regression model that there is a statistically significant difference in the frequency of successful evaluation over the six assignments between life science and engineering students when controlling for the same variables. While there was no difference among the different concentrations in FCI normalized gain, final course score, or whether students answered the estimation problems correctly or using valid physics concepts and equations, engineering concentrators are more likely to successfully engage in evaluation than life science students. It is possible that this stems from the difference in previous courses that the different concentrations have taken. For example, engineering education often emphasizes the development of problem-solving skills, critical thinking, and the ability to apply theoretical concepts to practical situations. These skills are particularly relevant to solution evaluation. On the other hand, life science education may focus more heavily on content knowledge related to biological systems and processes, which

may not directly translate to solution evaluation skills. It is important to recognize the potential differences in student experience dependent upon their previous course work so that all students may be supported in their development of problem-solving and metacognition skills.

5.4 Conclusion

Metacognitive skills allow individuals to observe and regulate their own levels of knowledge and expertise, effectively plan and apply learning resources, and assess their own progress. Evaluation, one of the eight subcomponents of metacognition, involves appraising the products and regulatory processes of one's learning. It is thus an essential component of the physics problem-solving process as it helps to verify whether solutions obtained are reasonable and identify mistakes made during the problem-solving process. Through investigating the trends in students' metacognitive ability on estimation problems on their weekly assignments in an introductory physics course, we have discovered several notable findings. In terms of course performance, we have found that students who are able to successfully evaluate their solutions on certain assignments (specifically, 2, 6, and 12) are associated with a higher final course score, and this remains true even when controlling for student concentration, college year, previous physics experience, and FCI normalized gain. And though we have not found an association between specific evaluation techniques and higher final scores, we have found that the increased frequency of successful evaluation is a predictor of a higher final score even when controlling for the aforementioned variables. These findings suggest that students who can thoughtfully evaluate their own reasoning and problem-solving decisions are more likely to be successful in the course.

Though we did not find trends in whether students successfully evaluated their solutions over the semester aside from the second assignment to the latter ones, we did find trends in certain evaluation techniques students used for the particular estimation problem as well by the problem difficulty. Specifically, we found that there are statistically significant associations between the assignments and the evaluation techniques of reflection using problem definitions and assumptions, reflection using knowledge, reflection using solution, and mathematical answer-checking strategies. Our findings indicate that evaluation is dependent upon the task, and certain problems are better aligned with particular evaluation strategies. Similarly, evaluation is affected by the level of difficulty of the problem for students. That is, students are more likely to successfully evaluate their solutions for easier problems, and they are also more likely to use the specific evaluation techniques of reflection using problem definitions and assumptions and mathematical answer-checking strategies as well as use more than one evaluation technique for the easier problem. For the more difficult problem, however, students are more likely to evaluate their solution through reflection using solution, which is likely a result of this being the most straightforward strategy for students to use when presented with less familiar physics content.

Lastly, we found that engineering concentrators are more likely to successfully evaluate their solutions compared to life science concentrators when controlling for college year and previous physics experience. This higher likelihood of evaluation success of engineering students may be attributed to their previous course experiences that emphasize problem-solving skills and critical thinking compared to the courses life science concentrators typically take. However, we did not find a difference between engineering and life science students and the specific evaluation techniques they used. This is unsurprising as it relates to our previously addressed

findings that the evaluation technique students used was dependent upon the task itself and the level of difficulty of the problem.

Though we did observe a greater likelihood of engineering students successfully evaluating their solutions, evaluation is still a challenging component of the problem-solving process for all students, especially on new content material each week. Through our findings of how different students evaluate their solutions to different problems, we can learn how to better support students in this component of metacognition. To improve evaluation skills, students should receive explicit instructions as well as scaffolded evaluation activities at the beginning of the course. While students did receive written instructions via their course text and examples of how to evaluate solutions, they did not have the opportunity to actively engage in evaluation practice or receive feedback prior to their homework assignments. Moreover, the weekly feedback on the homework assignments needs to be both consistent and specifically targeted toward the evaluation component of the four-step problem-solving procedure. This feedback should include suggestions for how students can improve their evaluation step. With these additional practices and feedback, we hypothesize that we would see increased instances of successful evaluation throughout the semester. Likewise, we postulate that scaffolded activities and structured feedback on evaluation would diminish the gap in successful evaluation based on students' incoming preparation, as we observed between life science and engineering students. By improving their evaluation skills, students will further develop their problem-solving ability which is important beyond the classroom context.

Chapter 6

Conclusions

The aim of this dissertation is to expand our understanding of problem-solving ability in undergraduate introductory physics students. Our research comprises four main projects, each of which explores a particular aspect of measuring and assessing problem-solving ability. Here we review the conclusions drawn from these four studies and emphasize the common theme of gaining a deeper understanding of the development of problem-solving skills in introductory physics students as well as the challenges they encounter in improving their abilities.

6.1 Review

In our first study, we present the design and development of three separate problem-solving diagnostic tools. We investigate the novice-expert differences in the problem-solving process elicited from two of the surveys, which provide insight into the design and future analysis of the third problem-solving tool. One key finding from the analysis of the mechanics and E&M surveys is that experts are more detailed in their descriptions of how they would solve the problems, and they also show evidence of simplifying the problem in the plans. Novices, on the other hand, struggled to decompose the problems into more tractable subproblems, which may be a result of their inability to appropriately leverage their physics understanding. In the mechanics survey, we found that experts were less procedural in their approach compared to novices. This may be a product of how students are instructed to systematically solve problems

involving static equilibrium, while experts consider the full scenario before focusing on one specific procedure to solve the problem. Furthermore, the analysis of the mechanics survey highlights the nonlinearity of problem-solving ability, with intermediate respondents' performance sometimes aligning with expectations of being more expert-like and sometimes relating more closely with novice responses.

Interestingly, we did not find any differences in physics content respondents included in the E&M survey, which may suggest that novices are familiar enough with the content knowledge needed to solve the problem, regardless of their ability to apply their knowledge to successfully solve the problem. Overall, the problem presented in the E&M survey seemed to be challenging for novices both in devising a plan for solving the problem and in selecting the correct possible solution. How novices initially devised a plan did not lead them to successfully solve the problem, but future work should investigate how novices deal with new information or data and whether they are able to reflect and update their plan for solving the problem.

The findings and feedback from the mechanics and E&M surveys can inform the development of future problem-solving diagnostic tools and strategies to improve physics problem-solving assessment. Future surveys, including the blended problem-solving diagnostic tool, should be structured to better investigate the problem-solving process by focusing on what problem-solving decisions respondents make to solve the problem rather than the physics content knowledge in those decisions. It is also important for future surveys to give respondents opportunities to re-answer questions after being provided with additional information, thus probing how they deal with new data and their ability to reflect. Through assessing respondents with a range of physics expertise, we aim in the future to better understand the timeline and nature of experiences needed to develop problem-solving skills.

In Chapter 3 we explore the effectiveness of worked examples in facilitating the learning of problem-solving practices in physics. Working memory plays a crucial role in processing information while solving problems, and effective problem-solving methods can be fostered by considering prior knowledge and providing appropriate scaffolding to minimize cognitive load. Worked examples can therefore allow learners to direct cognitive resources toward constructing solution schemas for specific problem types. In this study, we investigated the practices involved in solving a complex, authentic physics problem and if a worked example can help students learn these problem-solving practices. We found that worked examples helped students adopt more effective problem-solving practices, including effective decomposition and data collection practices. However, this improvement in problem-solving practices did not result in students successfully solving the problem. In order to engage students in learning practices used by experts in physics, both the solution steps and the decision behind each step must be visible. Future work should focus on designing and studying different iterations of worked examples with the objective of assisting students in acquiring proficient problem-solving practices.

In Chapter 4 we present two studies investigating student epistemologies while reflecting on solutions to physics problems. In the first analysis, we examine the strategies that students use to check their answers, providing insight into the factors that may influence student epistemology of solution reflection. We found that both problem type and level of physics background have a significant impact on student strategy use. If cognitive resources were the only determining factor for strategy choice, we would not observe such differences in strategy use across different problems. However, the variation in strategy use across student levels suggests that both the availability and activation of resources may be at play. Advanced students tend to use more complex strategies, which may be due to their conditioning or access to a

broader range of strategies for problem-solving. To verify this, we examined strategy use in a different task and found no differences between student backgrounds, indicating that even advanced students can use simpler strategies in different contexts. This supports previous research suggesting that resource activation plays a crucial role in strategy choice.

In the second analysis of this study, we investigate the epistemological framing associated with answer checking to gain a more detailed understanding of this process within and across individuals. We found that more advanced physics students tend to view solution reflection as a conceptual task and associate it with specific strategies such as checking limits and identifying functional behavior or making predictions about how a system will behave. However, due to the small sample size, the correlation between strategy use and problem difficulty is biased, as novice students tend to choose easier strategies and view solution reflection as an algorithmic task involving checking force components and their calculations. It is only when these strategies fail that these more novice students turn to alternative strategies, while still viewing them as algorithmic procedures to arrive at a solution. This suggests that the resources instructors want them to use may not always be activated in the problem context.

Furthermore, we found that intermediate students exhibit a greater variation in both strategy use and epistemological framing of solution reflection. These students seem to transition from a stable but often unproductive epistemology, where solution reflection is viewed as an algorithmic procedure utilizing basic problem-solving strategies acquired in physics courses, to a more stable epistemology where solution reflection involves conceptual understanding and sense-making. Intermediate students also show some variation in strategy use based on problem difficulty.

It is often difficult to distinguish between physical and mathematical sense-making during answer checking, though we did find evidence of students shifting between algorithmic and conceptual thinking. Methodological limitations for this work include using think-aloud protocols to measure epistemology, which involves inferring how students perceive a task based on verbal reports. The solution reflection task also only examines reflection in action and does not probe reflection on action. Future work should investigate reflection on action as well as how strategy use and epistemology change when checking one's own answer. To improve reflective practices and how students perceive a solution reflection task, it is important for instructors to include targeted reflection activities integrated with physics concepts into lessons and assessments.

Finally, in Chapter 5, we explore how students engage in the metacognitive subcomponent of evaluation on their weekly homework assignments over a semester in an introductory physics course. Our findings suggest that students who can thoughtfully evaluate their solutions are more likely to be successful in the course. Specifically, we found that students who were able to successfully evaluate their solutions on certain assignments were associated with higher mean final course scores even when controlling for concentration, college year, previous physics experience, and FCI normalized gain. Additionally, we found that the increased frequency of successful evaluation is a predictor of a higher final score when controlling for the same variables. However, we did not find an association between specific evaluation techniques and higher mean final scores. It is possible that successful evaluation may require a higher level of understanding of the course material or problem-solving skills; however, it is also possible that successful and consistent evaluation indicates a higher level of work ethic and consistency of performance in other areas of the course, leading to a higher final score.

Furthermore, we found that evaluation is dependent upon the task, and certain problems are better aligned with specific evaluation strategies. Evaluation is also affected by the level of difficulty of the problem for students. For easier problems, students are more likely to successfully evaluate their solutions and do so using the specific evaluation techniques of reflection using problem definitions and assumptions as well as mathematical answer-checking strategies. For the more difficult problems, students are more likely to evaluate their solutions through reflection using solution.

Lastly, we found that engineering concentrators are more likely to successfully evaluate their solutions compared to life science concentrators when controlling for college year and previous physics experience. We did not find a difference between engineering students and life science concentrators and the specific evaluation techniques they used. Though, this may be accounted for by the previously addressed findings that the specific evaluation task students used was dependent upon the task itself. It is possible that the variation in students' evaluation performance is a result of students' previous courses for the different concentrations. For instance, engineering education tends to prioritize the cultivation of problem-solving abilities, critical thinking, and the application of theoretical concepts to real-world scenarios, which are skills that pertain to solution evaluation. Whereas life science education may place greater emphasis on knowledge acquisition about biological systems and processes, which may not be directly relevant to solution evaluation abilities.

To improve evaluation skills, students should receive explicit instructions and scaffolded evaluation activities at the beginning of the course. Weekly feedback on homework assignments should be both consistent as well as specifically targeted toward the evaluation component. With these additional practices and feedback, we anticipate an increase in the frequency of successful

solution evaluation throughout the semester and that the gap in successful evaluation based on students' incoming preparation will diminish.

6.2 Importance of teaching problem-solving skills

Learning physics can be difficult for introductory students as they often struggle to connect the abstract concepts they learn to more tangible examples. Without appropriate guidance and scaffolding support, introductory physics students may resort to more formulaic problem-solving methods that do not require a deeper understanding of the physics concepts (Heller & Reif, 1984). The development of problem-solving skills is thus a crucial aspect of learning physics, especially for guiding novice students toward expertise.

Problem solving involves the application of physics concepts, principles, and theories to real-world situations, enabling students to develop critical thinking skills as well as gain a deeper understanding of the subject. Through fostering problem-solving skills, students can learn to analyze complex problems, break them into smaller, more manageable subproblems, and use physics knowledge to devise effective solutions while reflecting and monitoring their learning throughout. Students need to engage in appropriately scaffolded problem-solving activities so that they may develop their creativity, intuition, and perseverance, which are all essential for success in physics and other STEM fields as well as in everyday life. We hope that the work presented in this dissertation has provided insight into the complexities of developing problem-solving skills in introductory physics classrooms as well as possible methods for measuring and assessing these skills so that problem-solving instruction can be improved.

References

- Adams, W. K., & Wieman, C. E. (2015). Analyzing the many skills involved in solving complex physics problems. *American Journal of Physics*, 83(5), 459-467.
<https://doi.org/10.1119/1.4913923>
- Ali, M., Talib, C. A., Ibrahim, N. H., Surif, J., & Abdullah, A. H. (2016). The Importance of Monitoring Skills in Physics Problem Solving. *European Journal of Education Studies*, 1(3), 1-10. <https://doi.org/10.5281/zenodo.51089>
- Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works*. John Wiley & Sons
- Anderson, J. R. (1993). Problem solving and learning. *American Psychologist*, 48(1), 35–44. <https://doi.org/10.1037/0003-066X.48.1.35>
- Anderson, J. R. (2000). *Learning and Memory*, Second Edition. New York: Wiley.
- Armstrong, P. (2010). Bloom's Taxonomy. Vanderbilt University Center for Teaching. Retrieved February 23, 2023 from <https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/>
- Baird, J. R. (1986). Improving learning through enhanced metacognition: a classroom study. *European Journal of Science Education*, 8(3), 263-282. <https://doi.org/10.1080/0140528860080303>
- Bing, T. J., & Redish, E. F. (2012). Epistemic complexity and the journeyman-expert transition. *Physical Review Special Topics - Physics Education Research*, 8(1), 010105. <https://doi.org/10.1103/PhysRevSTPER.8.010105>
- Brown, A. L. (1978). Knowing when, where, and how to remember: a problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology*, 7, 55-111. New York: Academic Press.
- Burkholder, E. W., Miles, J. K., Layden, T. J., Wang, K. D., Fritz, A. V., & Wieman, C. E. (2020). Template for teaching and assessment of problem solving in introductory physics. *Physical Review Physics Education Research*, 16(1). <https://doi.org/https://doi.org/10.1103/PhysRevPhysEducRes.16.010123>
- Burkholder, E., Blackmon, L., & Wieman, C. (2020). Characterizing the mathematical problem-solving strategies of transitioning novice physics students. *Physical Review Physics Education Research*, 16(2), 020134. <https://doi.org/10.1103/PhysRevPhysEducRes.16.020134>

- Burkholder, E., Salehi, S., Sackeyfio, S., Mohamed-Hinds, N., & Wieman, C. (2021). *An equitable and effective approach to introductory mechanics*.
- Cancula, M., Planinšić, G., Etkina, E. (2015). Analyzing patterns in experts' approaches to solving experimental problems. *Am. J. Phys.*, 83(4), 366. Retrieved March 2, 2023, from <https://doi.org/10.1119/1.4913528>
- Carr, R. L. B. I., Lynch D.; Strobel, Johannes. (2012). Engineering in the K-12 STEM Standards of the 50 U.S. States: An Analysis of Presence and Extent. *Journal of Engineering Education*, 101(3), 539–564. <https://doi.org/10.1002/j.2168-9830.2012.tb00061.x>
- Chari, D. N., Nguyen, H. D., Zollman, D. A., & Sayre, E. C. (2019). Student and instructor framing in upper-division physics. *American Journal of Physics*, 87(11), 875-884. <https://doi.org/10.1119/1.5120392>
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121-152. <https://doi.org/10.1207/s15516709cog0502>
- Chi, M. T. H., Glaser, R., Rees, E. (1982). Expertise in problem solving. In Sternberg, R. J. (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum.
- Chi, M.T.H. (2006). Two approaches to the study of experts' characteristics. In N. Charness, P.J. Feltovich, R.R. Hoffman, & K.A. Ericsson (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 21-30). New York, NY: Cambridge University Press.
- Clark, Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: evidence-based guidelines to manage cognitive load* (1st edition). San Francisco, CA: Jossey-Bass.
- Dewey, J. (1933). *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. Boston, MA: D.C. Heath & Co Publishers.
- diSessa, A. A. (2018). Knowledge in Pieces: An evolving framework for understanding knowing and learning. In T. Amin & O. Levrini (Eds.), *Converging perspectives on conceptual change: Mapping an emerging paradigm in the learning sciences* (pp. 9-16). London, UK: Routledge
- Eichenlaub, M., & Redish, E. F. (2019). Blending Physical Knowledge with Mathematical Form in Physics Problem Solving. In G. Pospiech, M. Michelini, & B.-S. Eylon (Eds.), *Mathematics in Physics Education* (pp. 127–151). Springer International Publishing.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251. <https://doi.org/10.1037/0033-295X.87.3.215>
- Eylon, B.-S., & Reif, F. (1984). Effects of knowledge organization on task performance. *Cognition and Instruction*, 1(1), 5–

44. https://doi.org/10.1207/s1532690xci0101_2
- Fischer, F., Chernikova, O., & Opitz, A. (2022). Learning to Diagnose with Simulations: Introduction. In F. Fischer & A. Opitz (Eds.), *Learning to Diagnose with Simulations: Examples from Teacher Education and Medical Education* (pp. 1-4). Springer International Publishing. https://doi.org/10.1007/978-3-030-89147-3_1
- Fisher, R. (1998). Thinking About Thinking: Developing Metacognition in Children. *Early Child Development and Care, 141*, 1-15. <https://doi.org/10.1080/0300443981410101>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive development inquiry. *American Psychologist, 34*(10), 906–911.
- Georghiades, P. (2004). From the general to the situated: Three decades of metacognition. *International Journal of Science Education - INT J SCI EDUC, 26*, 365-383. <https://doi.org/10.1080/0950069032000119401>
- Glaser, R., Schauble, L., Raghavan, K., & Zeitz, C. (1992). Scientific reasoning across different domains. In *Computer-based learning environments and problem solving* (pp. 345-371). Springer Berlin Heidelberg.
- Greene, J., & Roger, A. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology, 34*, 18-29. <https://doi.org/10.1016/j.cedpsych.2008.05.006>
- Gupta, A., & Elby, A. (2010). Beyond epistemological deficits: Incorporating flexible epistemological views into fine-grained cognitive dynamics.
- Hahn, K., Emigh, P., Lenz, M., & Gire, E. (2017, July 26-27) Student sense-making on homework in a sophomore mechanics course. Paper presented at Physics Education Research Conference 2017, Cincinnati, OH. Retrieved April 2, 2023 from <https://www.compadre.org/Repository/document/ServeFile.cfm?ID=14595&DocID=4772>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74. <https://doi.org/10.1119/1.18809>
- Hammer, D., & Berland, L. (2013). Confusing Claims for Data: A Critique of Common Practices for Presenting Qualitative Research on Learning. *Journal of the Learning Sciences, 23*, 37-46. <https://doi.org/10.1080/10508406.2013.802652>
- Hammer, D., & Elby, A. (2002). On the Form of a Personal Epistemology. *Personal Epistemology: The Psychology of Beliefs About Knowledge and Knowing, 169-190*.

- Hardiman, P., Dufresne, R., & Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition*, 17(5), 627-638. <https://doi.org/10.3758/BF03197085>
- Heller, J. I., & Reif, F. (1984). Prescribing effective human problem-solving processes: Problem description in physics. *Cognition and Instruction*, 1(2), 177–216. https://doi.org/10.1207/s1532690xci0102_2
- Heller, J., & Reif, F. (1984). Prescribing effective human problem-solving processes: Problem description in physics. *Cog. Instr.*, 1(2), 177-216. Retrieved March 23, 2023, from https://doi.org/10.1207/s1532690xci0102_2
- Heller, K., & Heller, P. (2000). *The competent problem solver for introductory physics*. Boston: McGraw-Hill.
- Heller, K., & Heller, P. (2010). *Cooperative Problem Solving in Physics: A User's Manual*.
- Heller, P., Heller, K., Henderson, C. R., Kuo, V., & Yerushalmi, E. (2001). Instructors' Beliefs and Values about Learning Problem Solving. Part of the PER Conference Series, July 25-26, Rochester, New York. <https://doi.org/10.1119/perc.2001.pr.005>
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141-158. <https://doi.org/10.1119/1.2343497>
- Heuvelen, A. V., & Zou, X. (2001). Multiple representations of work–energy processes. *American Journal of Physics*, 69(2), 184-194. <https://doi.org/10.1119/1.1286662>
- Honey, M., & Hilton, M. L. (2011). Learning Science through Computer Games and Simulations.
- Hull, M. M., Jansky, A., & Hopf, M. (2022). Does confidence in a wrong answer imply a misconception? *Physical Review Physics Education Research*, 18(2), 020108. <https://doi.org/10.1103/PhysRevPhysEducRes.18.020108>
- Ince, E. (2018). An overview of problem solving studies in physics education. *Journal of Education and Learning*, 7(4), 191-200. <https://doi.org/10.5539/jel.v7n4p191>
- Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: issues in definition, measurement, and instruction. *Educational Psychologist*, 22(3), 255–278. <https://doi.org/10.1080/00461520.1987.9653052>
- Jill H. Larkin & F. Reif (1979) Understanding and Teaching Problem-Solving in Physics, *European Journal of Science Education*, 1:2, 191-203, DOI: 10.1080/0140528790010208
- Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and*

- Development*, 45(1), 65-94. <https://doi.org/10.1007/BF02299613>
- Jonassen, D. H. (2014). Engineers as Problem Solvers. In A. Johri & B. M. Olds (Eds.), *Cambridge Handbook of Engineering Education Research* (pp. 103-118). Cambridge University Press. <https://doi.org/DOI: 10.1017/CBO9781139013451.009>
- Knight, R. D. (2017). *Physics for scientists and engineers: A strategic approach* (4th ed.). Boston: Pearson.
- Kohl, P. B., & Finkelstein, N. D. (2008). Patterns of multiple representation use by experts and novices during physics problem solving. *Physical Review Special Topics - Physics Education Research*, 4(1), 010111. <https://doi.org/10.1103/PhysRevSTPER.4.010111>
- Krajcik, J. (2015). Three-Dimensional Instruction: Using a New Type of Teaching in the Science Classroom. *Science and Children*, 53(3), 6-8. <http://search.proquest.com.ezp-prod1.hul.harvard.edu/scholarly-journals/three-dimensional-instruction-using-new-type/docview/1728349255/se-2>
- Kyllonen, P. C., & Cristal, R. E. (1990). Reasoning Ability Is (Little More than) Working Memory Capacity?! *Intelligence*, 14, 389-433. [http://dx.doi.org/10.1016/S0160-2896\(05\)80012-1](http://dx.doi.org/10.1016/S0160-2896(05)80012-1)
- Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208(4450), 1335–1342. <https://doi.org/10.1126/science.208.4450.1335>
- Liu, C.-C., Cheng, Y.-B., & Huang, C.-W. (2011). The effect of simulation games on the learning of computational problem solving. *Computers & Education*, 57(3), 1907-1918. <https://doi.org/https://doi.org/10.1016/j.compedu.2011.04.002>
- Maries, A., & Singh, C. (2023). Helping Students Become Proficient Problem Solvers Part I: A Brief Review. *Education Sciences*, 13(2), 156. <https://www.mdpi.com/2227-7102/13/2/156>
- Modir, B., Thompson, J. D., & Sayre, E. C. (2017). Students' epistemological framing in quantum mechanics problem solving. *Physical Review Physics Education Research*, 13(2), 020108. <https://doi.org/10.1103/PhysRevPhysEducRes.13.020108>
- Neto, A., & Valente, M. O. (1997). Problem solving in physics: towards a metacognitively developed approach. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching Oak Brook.
- Neto, A., & Valente, M. O. (1997). Problem solving in physics: towards a metacognitively developed approach. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching Oak Brook.

- Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Next Generation Science Standards: For States, By States* (2013). The National Academies Press. Retrieved February 20, 2022 from <http://www.nextgenscience.org/>
- OECD (2014), *PISA 2012 Results: Creative Problem Solving: Students' Skills in Tackling Real-Life Problems (Volume V)*, PISA, OECD publishing.
<http://dx.doi.org/10.1787/9789264208070-en>
- Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skills in statistics: a cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429-434.
<https://doi.org/10.1037/0022-0663.84.4.429>
- Pellegrino, J. W. (2014). Proficiency in Science: Assessment Challenges and Opportunities.
https://indigo.uic.edu/articles/journal_contribution/Proficiency_in_Science_Assessment_Challenges_and_Opportunities/10752938
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Into Practice*, 4, 218-225.
- Porter, A. M. (2019). *Physics PhDs Ten Years Later: Success Factors and Barriers in Career Paths*.
- Price, A. M., Kim, C. J., Burkholder, E. W., Fritz, A. V., & Wieman, C. E. (2021). A Detailed Characterization of the Expert Problem-Solving Process in Science and Engineering: Guidance for Teaching and Assessment. *CBE—Life Sciences Education*, 20(3), ar43.
<https://doi.org/10.1187/cbe.20-12-0276>
- Price, A., Salehi, S., Burkholder, E., Kim, C., Isava, V., Flynn, M., & Wieman, C. (2022). An accurate and practical method for assessing science and engineering problem-solving expertise. *International Journal of Science Education*, 44(13), 2061-2084.
<https://doi.org/10.1080/09500693.2022.2111668>
- Reif, F., & Heller, J. (1982). Knowledge structures and problem solving in physics. *Educational Psychologist*, 17, 102-127.
- Rickey, D., & Stacy, A. M. (2000). The Role of Metacognition in Learning Chemistry. *Journal of Chemical Education*, 77(7), 915. <https://doi.org/10.1021/ed077p915>
- Rozencwajg, P. (2003). Metacognitive factors in scientific problem-solving strategies. *European Journal of Psychology of Education*, 18(3), 281-294.
- Salehi, S. (2018). *Improving problem-solving through reflection*. Doctoral Dissertation, Stanford University.
- Salehi, S., & Wieman, C. (2023). *A problem-solving framework: Characterizing actions and decisions involved in solving a complex problem*.

- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. *Handbook of research on mathematics teaching and learning*, 334-370.
- Schön, D.A. (1992). *The Reflective Practitioner: How Professionals Think in Action* (1st ed.). Routledge. <https://doi.org/10.4324/9781315237473>
- Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, 26(1), 113-125. <https://doi.org/10.1023/A:1003044231033>
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351-371.
- Schraw, G., Crippen K. J., & Hartley, K. (2006). Promoting self-regulation in science education: metacognition as part of a broader perspective on learning. *Research in Science Education*, 36, 111-139.
- Schunk, D. H. (2008). *Learning theories: An educational perspective*. New York, NY: Prentice Hall.
- Schwartz, D. L., & Martin, T. (2004). Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction. *Cognition and Instruction*, 22(2), 129-184. https://doi.org/10.1207/s1532690xc2202_1
- Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. *Transfer of learning from a modern multidisciplinary perspective*, 3, 1-51.
- Shar, K., Russ, R. S., & Lavery, J. T. (2020). Student epistemological framing on paper-based assessments. *Physical Review Physics Education Research*, 16(2), 020101. <https://doi.org/10.1103/PhysRevPhysEducRes.16.020101>
- Sherin, B. L. (2001). How students understand physics equations. *Cognition and Instruction*, 19(4), 479–541. https://doi.org/10.1207/S1532690XCI1904_3
- Sikorski, T., White, G., & Landay, J. (2017, July 26-27). Uptake of solution checks by undergraduate physics students. Paper presented at Physics Education Research Conference 2017, Cincinnati, OH. Retrieved April 2, 2023 from <https://www.compadre.org/Repository/document/ServeFile.cfm?ID=14647&DocID=4824>
- Simon, H. A. (1974). How Big Is a Chunk? *Science*, 183(4124), 482-488. <https://doi.org/doi:10.1126/science.183.4124.482>

- Singh, C. (2002). When physical intuition fails. *American Journal of Physics*, 70(11), 1103-1109. <https://doi.org/10.1119/1.1512659>
- Sweller, & Cooper, G. A. (1985). The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra. *Cognition and Instruction*, 2(1), 59–89. https://doi.org/10.1207/s1532690xcio201_3
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12: 257-285. https://doi.org/10.1207/s15516709cog1202_4
- Sweller, J. (2006). The worked example effect and human cognition. *Learning and instruction*, 16(2), 165-169. <https://doi.org/10.1016/j.learninstruc.2006.02.005>
- Sweller, J. (2006). The worked example effect and human cognition. *Learning and instruction*, 16(2), 165-169. <https://doi.org/10.1016/j.learninstruc.2006.02.005>
- Taasobshirazi, G., & Farley, J. (2013). Construct validation of the physics metacognition inventory. *International Journal of Science Education*, 35(3), 447-459.
- Ton de, J., & van Joolingen, W. R. (1998). Scientific discovery learning with a computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179-201. <http://search.proquest.com.ezp-prod1.hul.harvard.edu/scholarly-journals/scientific-discovery-learning-with-computer/docview/214112548/se-2>
- Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, 36(3), 212-218. <https://doi.org/10.1016/j.cedpsych.2010.10.004>
- Wang, K. D., Salehi, S., Arseneault, M., Nair, K., & Wieman, C. (2021). *Automating the Assessment of Problem-solving Practices Using Log Data and Data Mining Techniques* Proceedings of the Eighth ACM Conference on Learning @ Scale, Virtual Event, Germany. <https://doi.org/10.1145/3430895.3460127>
- Warren, A. R. (2010). Impact of teaching students to use evaluation strategies. *Physical Review Special Topics - Physics Education Research*, 6(2), 020103. <https://doi.org/10.1103/PhysRevSTPER.6.020103>
- White Brahmia, S., Olsho, A., Smith, T. I., Boudreaux, A., Eaton, P., & Zimmerman, C. (2021). Physics Inventory of Quantitative Literacy: A tool for assessing mathematical reasoning in introductory physics. *Physical Review Physics Education Research*, 17(2), 020129. <https://doi.org/10.1103/PhysRevPhysEducRes.17.020129>
- Wieman, C. (2015). Comparative Cognitive Task Analyses of Experimental Science and Instructional Laboratory Courses. *The Physics Teacher*, 53, 349-351. <https://doi.org/10.1119/1.4928349>

- Willis, G. B. (2005). *Introduction to cognitive interviewing*. SAGE Publications, Inc., <https://doi.org/10.4135/9781412983655>
- Young, H. D., & Freedman, R. A. (2015). *University physics with modern physics* (14th ed.). Boston: Pearson.
- Zepeda, C. D., Richey, J. E., Ronevich, P., & Nokes-Malach, T. J. (2015). Direct Instruction of Metacognition Benefits Adolescent Science Learning, Transfer, and Motivation: An In Vivo Study. *Journal of Educational Psychology*, 107, 954-970.
- Zimmerman, B. J. (2006). Development and adaptation of expertise: The role of self-regulatory processes and beliefs. In N. Charness, P. J. Feltovich, R. R. Hoffman, & K. A. Ericsson (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 705–722). New York, NY: Cambridge University Press.
- Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 233–262). New York, NY: Cambridge University Press.

Appendix A

A.1 Mechanics problem-solving diagnostic tool



A ramp is leaning against (not sitting on top of) the back of a truck as shown above. A man carrying a heavy object, walks up the ramp, starting at the bottom of the ramp. The ramp is not frictionless. How far along the ramp could the man, carrying the object, travel before the ramp starts to slip?

Write out a detailed plan for how you would solve this problem. Be sure to include a list of quantities that you would need to know and assumptions you would have to make.

The following information is provided by the problem:

Length of ramp $l = 5\text{ m}$

angle between floor and ramp $= 53^\circ$

Coefficient of static friction between the ramp and the floor is the same as the coefficient of static friction between ramp and the truck $\mu_s=0.4$

Mass of ramp $m_1 = 50\text{ kg}$

Mass of man and heavy object $m_2 = 100\text{ kg}$

Given the quantities provided, which of the following assumptions would you **NEED** to make to be able to correctly solve this problem?

- ☐ the pivot point is at the center of mass of the ramp
- ☐ the pivot point is at the top of the ramp
- ☐ the pivot point is at the bottom of the ramp
- ☐ the pivot point is at some point along the ramp
- ☐ the heavy object and man are in the same position on the ramp
- ☐ the static friction forces prevent the ramp from sliding until the man gets far enough up the ramp
- ☐ the ramp will start to slide before the man reaches the truck
- ☐ all the forces are initially balanced
- ☐ friction between the ramp and the truck is negligible
- ☐ friction between the ramp and the floor is negligible
- ☐ the truck won't roll during the process
- ☐ the man's feet won't slip on the ramp
- ☐ the ramp will not bend
- ☐ the man and the heavy object are not accelerating as they travel up the ramp

List any additional assumptions (not listed above) you would NEED to make to be able to correctly solve this problem.

The following is a student solution:

Assumptions:

When ramp is not sliding, it is in static equilibrium so,

$$\Sigma F_x = 0$$

$$\Sigma F_y = 0$$

$$\Sigma \tau = 0$$

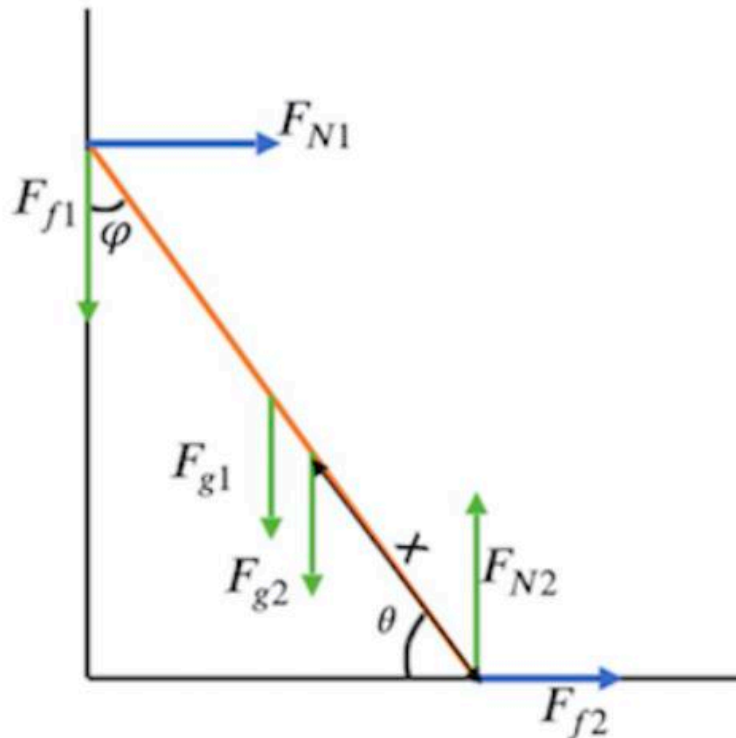
Person and bureau set as one combined mass on the ramp

Pivot point is about the bottom of the ramp

Ramp will start to slide before person on it reaches truck

Ramp will not bend

Define x = location of person on ramp just before it starts to slide



$$(1) \Sigma F_x = 0 = F_{N1} + F_{f2} = F_{N1} + \mu_s F_{N2}$$

$$(2) \Sigma F_y = 0 = F_{N2} - F_{g1} - F_{g2} - F_{f1}$$

$$0 = F_{N2} - F_{g1} - F_{g2} - \mu_s F_{N1}$$

$$(1) F_{N1} = -\mu_s F_{N2} \text{ sub (1) into (2)}$$

$$F_{N2} - \mu_s F_{N1} = F_{g1} + F_{g2}$$

$$\mu_s^2 F_{N2} + F_{N2} = (500 \text{ N} - 1000 \text{ N})$$

$$1.16 F_{N2} = -500 \text{ N}$$

$$F_{N2} = -431 \text{ N}$$

$$F_{N1} = -\mu_s F_{N2}$$

$$F_{N1} = -(0.4)(-431)$$

$$F_{N1} = 172.4 \text{ N}$$

Set pivot point about bottom of ramp:

$$(3) \Sigma \tau = 0 = \tau_{g1y} + \tau_{g2y} - \tau_{N1y}$$

$$0 = m_1 g \frac{l}{2} \sin \varphi + m_2 g x \sin \varphi - \mu_s F_{N1} l \sin \varphi$$

$$0 = (500 \text{ N}) \left(\frac{5}{2} \text{ m} \right) \sin 37^\circ + (1000 \text{ N})(x) \sin 37^\circ - (0.4)(172.4 \text{ N})(5 \text{ m}) \sin 37^\circ$$

$$0 = 752.3 \text{ Nm} + 601.8 \text{ N} * (x) - 207.5 \text{ Nm}$$

$$601.8 \text{ N} * x = 544.8 \text{ Nm}$$

$$x = 1.10 \text{ m}$$

This means that beyond 1.10 m to the right of the pivot the ramp will begin to slide.

Group the following criteria according to the importance for evaluating this students' problem solving process (focus on things that are procedurally important as opposed to things important to getting the correct answer). Be sure to rank the importance within each box as well.

Items	Most important for evaluating problem-solving process
Free body diagram includes all forces in the correct direction	
Forces in the x-direction add to zero	
Forces in the y-direction add to zero	
Torques add to zero	
Solution has a free body diagram	
Signs of the forces are correct	
Frictional forces have been calculated	
No algebra mistakes have been made	
The conditions for static equilibrium have been determined	
Assumptions have been clearly stated	
Important variables are clearly defined	
Solution is evaluated for reasonableness	

Are there any additional criteria (not listed in the previous question) that are important for evaluating the solution? If so - list them here and indicate the level of importance (most, medium, least).

The student solution may have an error (or multiple errors). Clearly identify any error(s). If there are multiple errors - list them in order of importance and explain your choice of ranking.

If there is an error(s) - provide a detailed description of the modification(s) that must be made to correct the solution.

List any physical assumptions that the student makes (in his/ her solution) but fails to list.

Check the forces in the x-direction and the forces in the y-direction. Are they balanced for each solution? Why or why not?

Check the torques. Are they balanced? Why or why not?

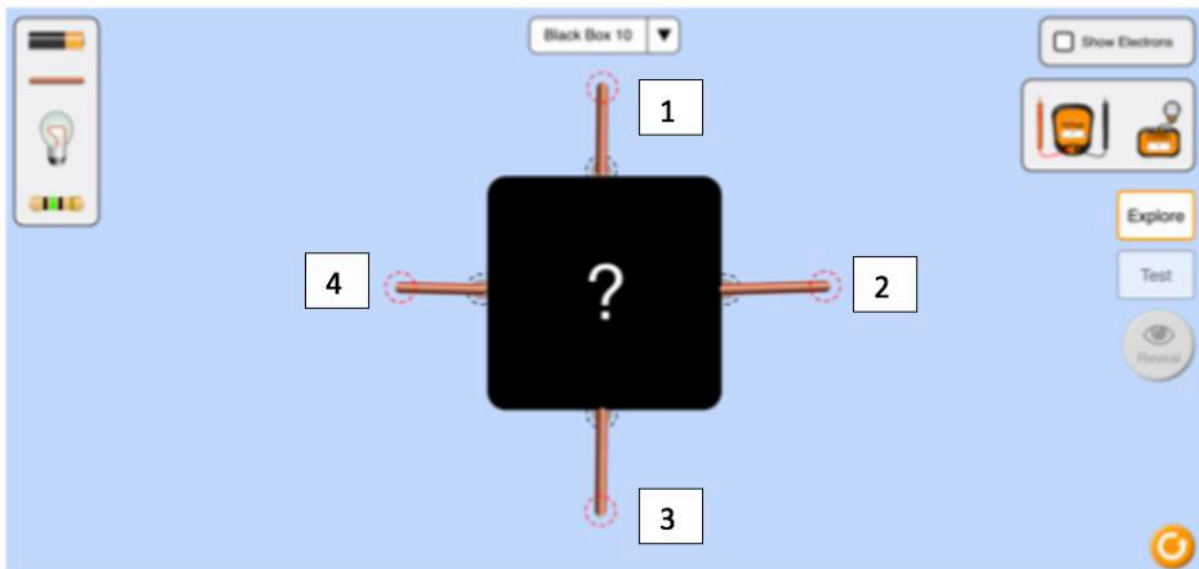
Is the student's interpretation of the answer correct? Explain why or why not.

What is the physical interpretation of the answer? (i.e. what does it mean?)

Discuss the reasonableness the final answer. Does it make sense? Why or why not.

What conditions need to be present for something not to slide?

A.2 E&M problem-solving diagnostic tool



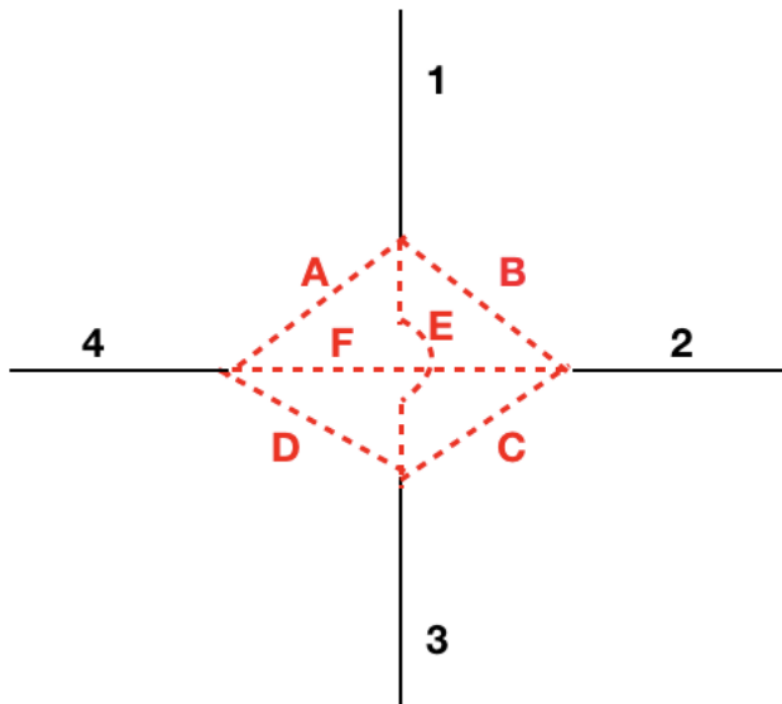
You will be presented with a circuit puzzle to solve. The questions that follow are designed to probe your thinking about the puzzle as well as your problem solving approach and process.

First, you will be asked to think about a general approach for solving the puzzle. Then, you will be presented progressively with more information to help you rethink or validate your strategy to solve the puzzle. We are interested in your approach to solving the puzzle rather than your facility with the relevant physics concepts. Please be as detailed as possible when describing your thoughts about process.

The black box above hides a mystery electrical configuration.

Your job is to figure out a possible electrical configuration that is hidden behind the black box by collecting data from the four wires protruding from the box. The mystery circuit could be comprised of the following components: wires, batteries, resistors. The circuit does not necessarily have all 3 types of components.

Each branch (depicted below as A, B, C, D, E, F) could be a wire, a battery, a resistor, or nothing. In other words, you could place at most 1 circuit element directly between any 2 wires. It is OK to connect a battery to the wires without any resistance.



You will be given the following items to use to figure out the mystery circuit:

- wires (as many as you need)
 - ammeter
 - voltmeter
 - resistors (you can use as many resistors as you need)
 - 9V battery
 - light bulb
-

Write out an approach for how you would figure out the mystery circuit configuration and the values (resistances, voltages) for any of the components.

Be sure to thoroughly describe your thought process. What would you do and why?

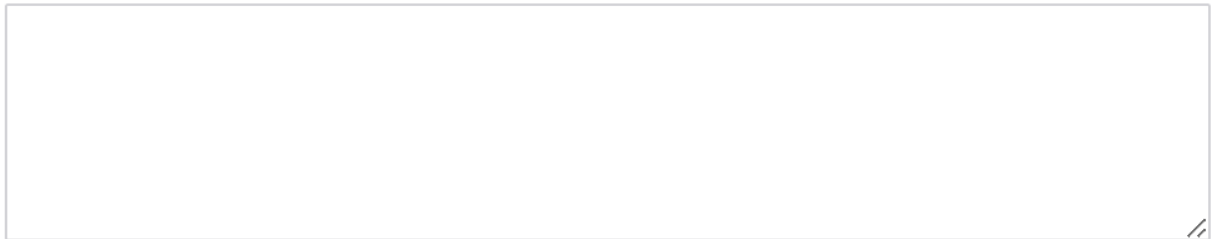
Explain why you think the circuit element configuration you chose in the previous question would be MOST beneficial for collecting data to determine the mystery circuit.

Focusing in on just the data part of planning, what data would you collect and how would you collect it?



Focusing on the data collection part of your plan, which electrical component(s) you would need to collect the data? Explain how you would connect the component(s) to the mystery circuit.

If you feel you have already answered this, indicate that below.



You place a voltmeter between pairs of protruding wires to measure the voltage. You use the ammeter to measure the current flowing through each of the protruding wires. The data you collect is presented in Table 1 (below).

The diagram below shows how you connect the voltmeter to measure the voltage between wires 2 and 3 and how you connect the ammeter to measure the current in wire 1.

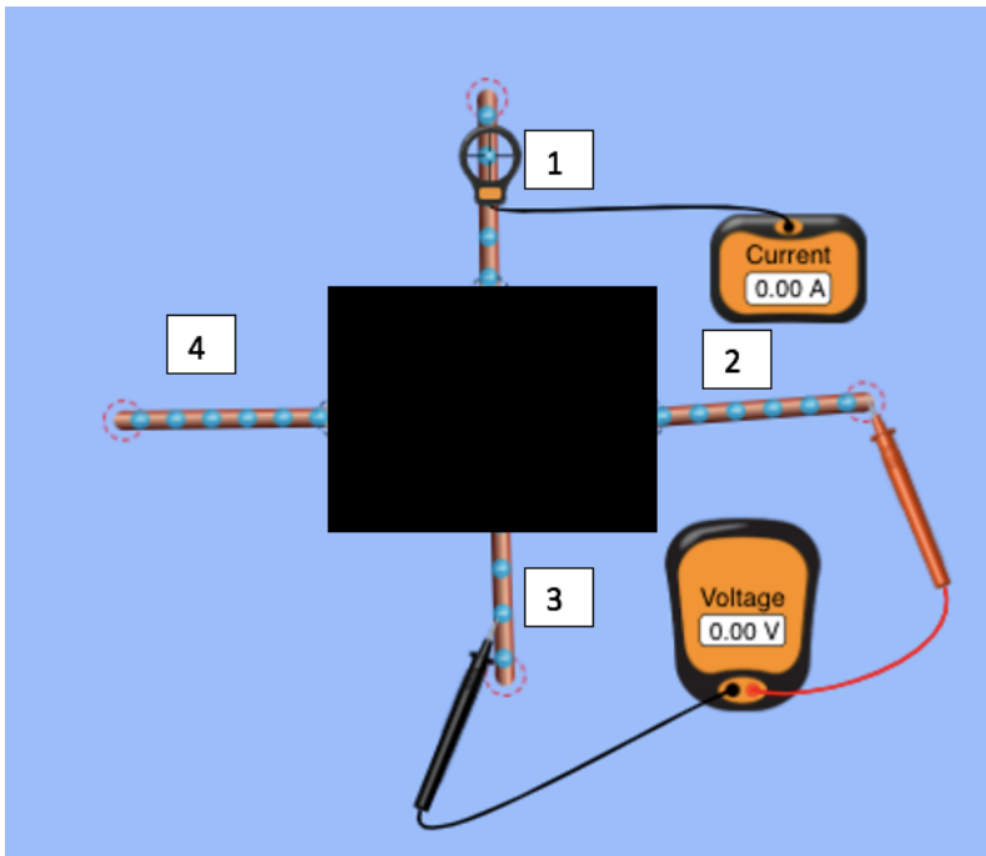


Table 1

Wires	Voltage (V)
1 & 2	0
2 & 3	0
3 & 4	0
4 & 1	0
1 & 3	0
2 & 4	0

Wire	Current (A)
1	0
2	0
3	0
4	0

Based on these data what can you conclude? Be specific.

Given the data presented in the previous question, write out a detailed plan (list of steps) for how you would figure out the mystery circuit configuration and the values (resistances, voltages) for any of the components.

If you feel like you have already answered this question - write "already answered"

You connect a 9 V battery between protruding wires 1 and 2. You place a voltmeter between pairs of protruding wires to measure the voltage. You use the ammeter to measure the current flowing through each of the protruding wires. The data you collect is presented in Table 2 (below). The diagram below shows how you connect the voltmeter to measure the voltage between wires 1 and 2 and how you connect the ammeter to measure the current in wire 2.

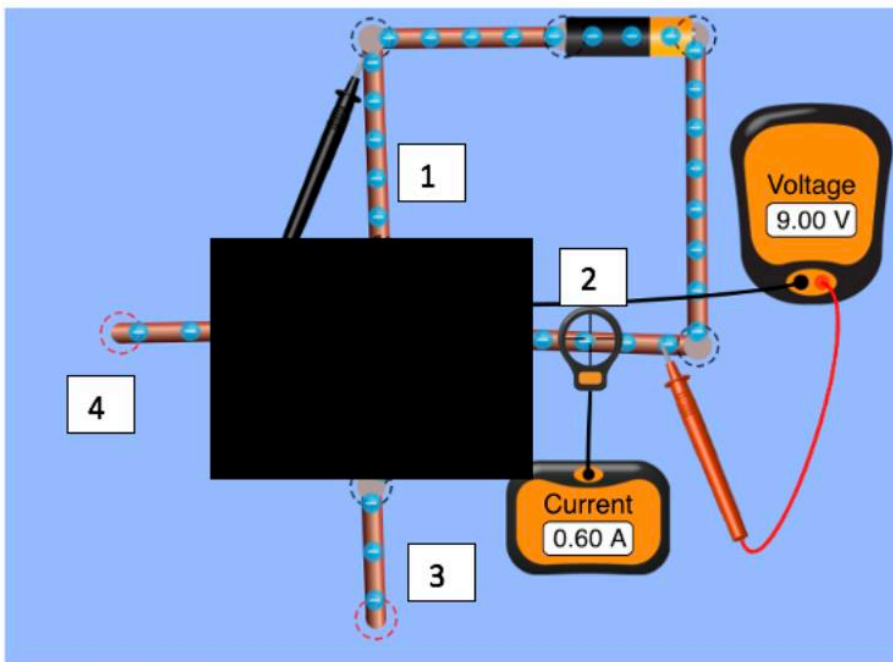


Table 2

Wires	Voltage (V)
1 & 2	9
2 & 3	0
3 & 4	0
4 & 1	0
1 & 3	9
2 & 4	0

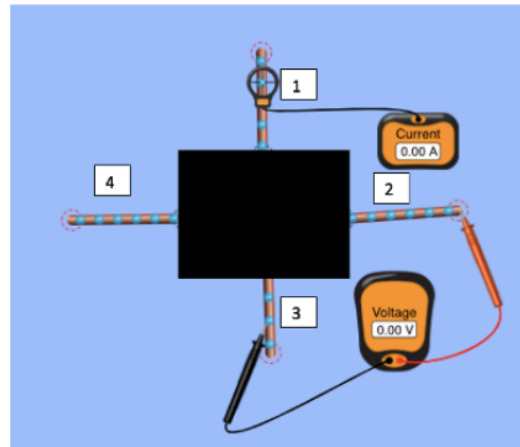
Wire	Current (A)
1	0.6
2	0.6
3	0
4	0

Data already collected:

Table 1

Wires	Voltage (V)
1 & 2	0
2 & 3	0
3 & 4	0
4 & 1	0
1 & 3	0
2 & 4	0

Wire	Current (A)
1	0
2	0
3	0
4	0



Based on these new data what can you conclude? Be specific.

Given this new data presented, write out a detailed plan (list of steps) for how you would figure out the mystery circuit configuration and the values (resistances, voltages) for any of the components.

If you feel like you have already answered this question - write "already answered"

You connect a 9 V battery between protruding wires 1 and 4. You place a voltmeter between pairs of protruding wires to measure the voltage. You use the ammeter to measure the current flowing through each of the protruding wires. The data you collect is presented in Table 3 (below). The diagram below shows how you connect the voltmeter to measure the voltage between wires 1 and 4 and how you connect the ammeter to measure the current in wire 1.

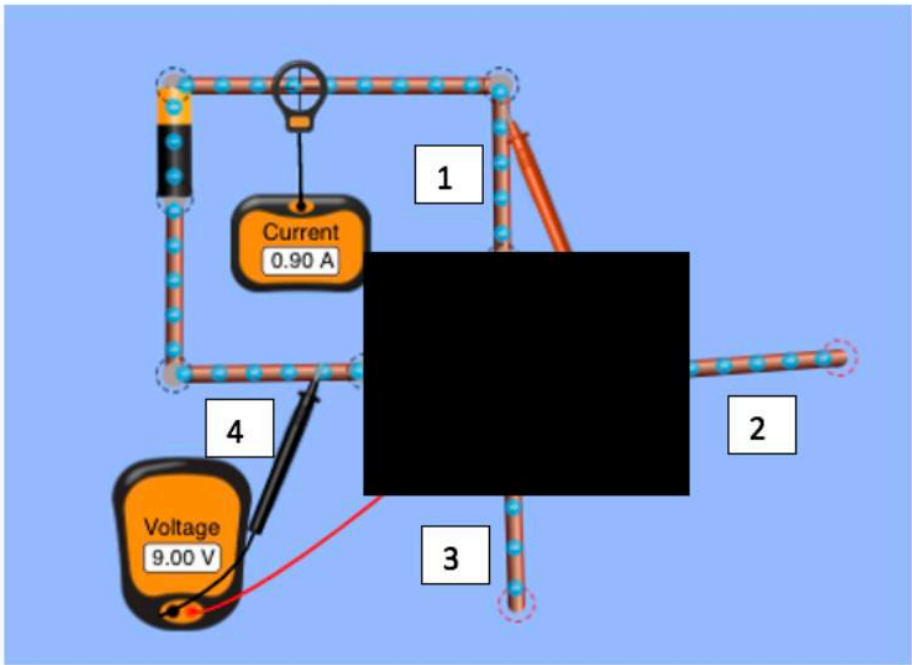


Table 3

Wires	Voltage (V)
1 & 2	0
2 & 3	0
3 & 4	9
4 & 1	9
1 & 3	0
2 & 4	9

Wire	Current (A)
1	0.9
2	0
3	0
4	0.9

Data already collected:

Table 1

Wires	Voltage (V)
1 & 2	0
2 & 3	0
3 & 4	0
4 & 1	0
1 & 3	0
2 & 4	0

Wire	Current (A)
1	0
2	0
3	0
4	0

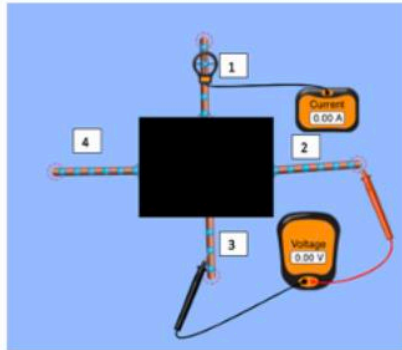
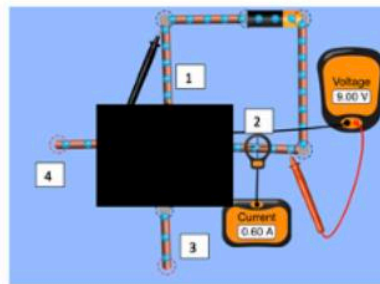


Table 2

Wires	Voltage (V)
1 & 2	9
2 & 3	0
3 & 4	0
4 & 1	0
1 & 3	9
2 & 4	0

Wire	Current (A)
1	0.6
2	0.6
3	0
4	0



Based on these new data what can you conclude? Be specific.

Given this new data presented, write out a detailed plan (list of steps) for how you would figure out the mystery circuit configuration and the values (resistances, voltages) for any of the components.

If you feel like you have already answered this question - write "already answered"

You connect a 9 V battery between protruding wires 3 and 4. You place a voltmeter between pairs of protruding wires to measure the voltage. You use the ammeter to measure the current flowing through each of the protruding wires. The data you collect is presented in Table 4 (below). The diagram below shows how you connect the voltmeter to measure the voltage between wires 3 and 4 and how you connect the ammeter to measure the current in wire 3.

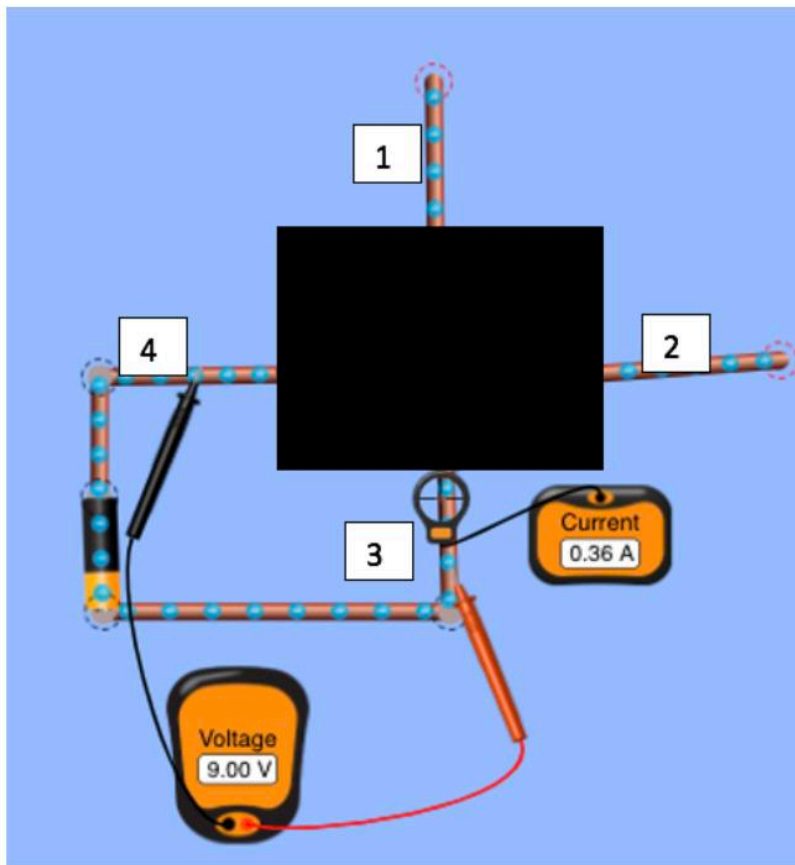


Table 4

Wires	Voltage (V)
1 & 2	5.4
2 & 3	0
3 & 4	9
4 & 1	3.6
1 & 3	5.4
2 & 4	9

Wire	Current (A)
1	0
2	0
3	0.36
4	0.36

Data already collected:

Table 1

Wires	Voltage (V)
1 & 2	0
2 & 3	0
3 & 4	0
4 & 1	0
1 & 3	0
2 & 4	0

Wire	Current (A)
1	0
2	0
3	0
4	0

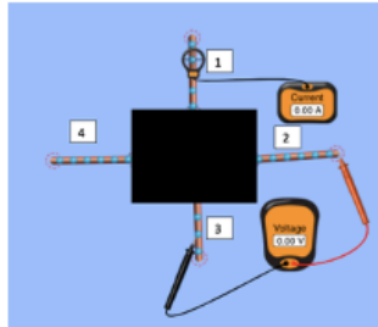


Table 2

Wires	Voltage (V)
1 & 2	9
2 & 3	0
3 & 4	0
4 & 1	0
1 & 3	9
2 & 4	0

Wire	Current (A)
1	0.5
2	0.5
3	0
4	0

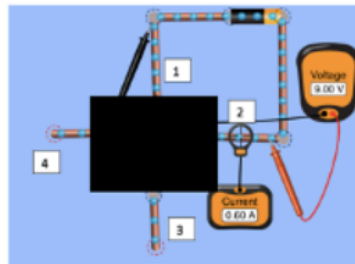
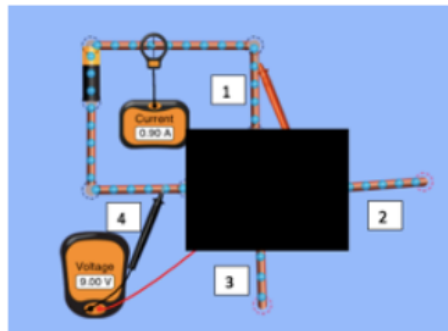


Table 3

Wires	Voltage (V)
1 & 2	0
2 & 3	0
3 & 4	9
4 & 1	9
1 & 3	0
2 & 4	9

Wire	Current (A)
1	0.9
2	0
3	0
4	0.9



Based on these new data what can you conclude? Be specific.

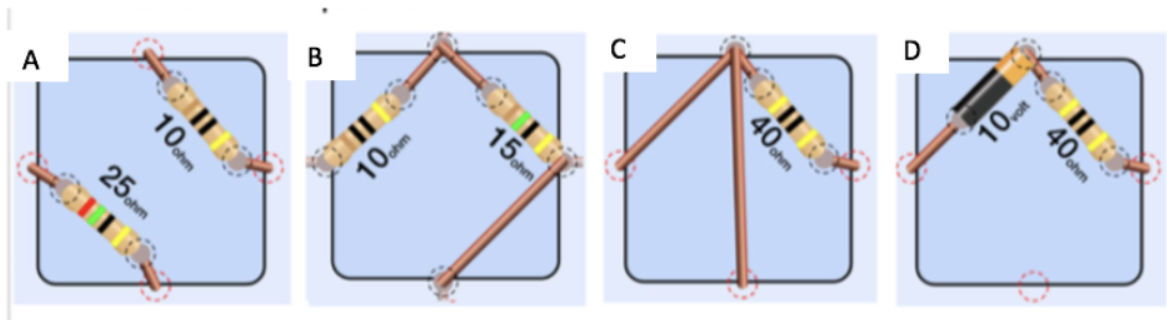
Given this new data presented, write out a detailed plan (list of steps) for how you would figure out the mystery circuit configuration and the values (resistances, voltages) for any of the components.

If you feel like you have already answered this question - write "already answered"

Examine the four circuits shown below. Which of them (if any) could be the mystery circuit, based on the data collected?

Explain, in detail, how you would determine this.

A summary of all the data collected is provided below.

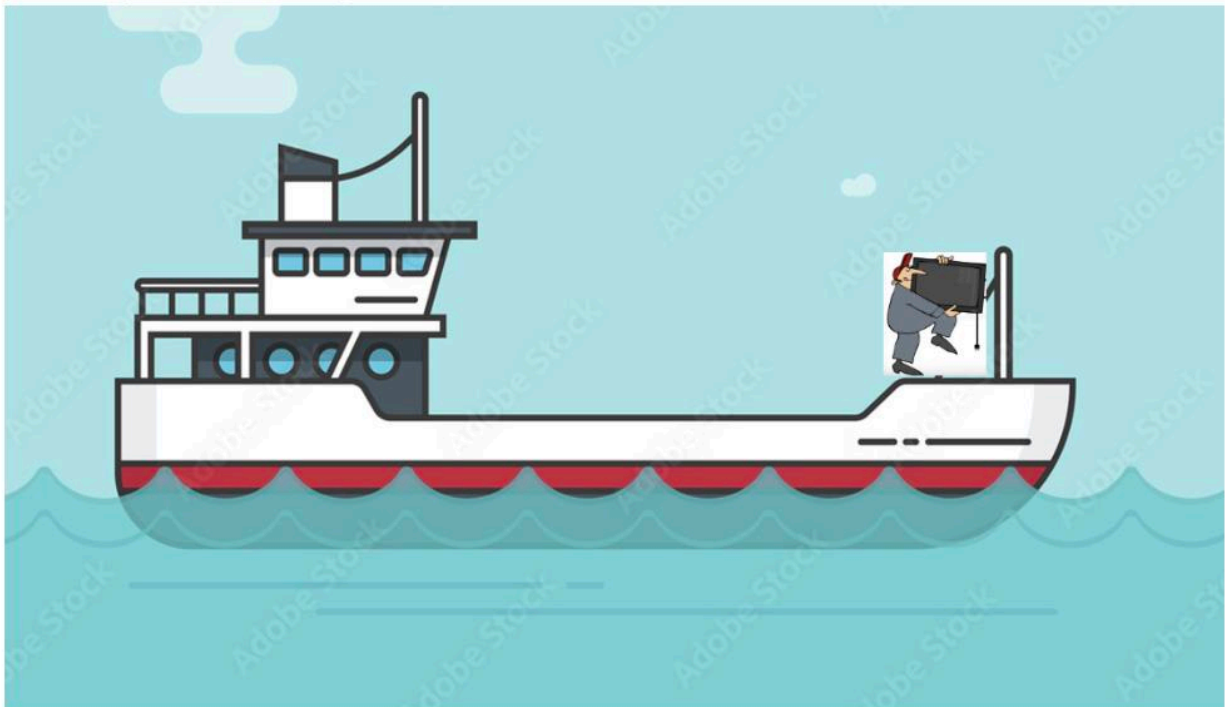


Are there any other possible solutions that could be the mystery circuit? How do you know?

If you were given a new mystery circuit, what would be your plan or strategy for figuring it out? Summarize your approach to solving a similar mystery circuit.

A.3 Blended problem-solving diagnostic tool

A ship building company has just finished building a new steel ship and they are outfitting the interior. They need to mount flat screen monitors in many of the rooms. The traditional way to do this is to drill holes in the steel walls and bolt the monitors on. This takes a lot of labor drilling holes in the steel walls, and it makes it very difficult to move the monitors to different locations as needs change. Someone in the company suggested the idea of bolting magnets to the back of the monitors and then holding them on the walls magnetically. They come to you as a physics consultant to figure out if this method would be practical. They also provide you with 4 magnets.



Imagine you are working for a physics consulting firm and you are recommending the most secure way to hang the monitor on the wall of the ship (assuming you only have 4 magnets). Write a detailed plan of your recommendation. Include as many details as you think are necessary for someone else to be able to implement it. Your response should be 1-2 paragraphs long.

What information would you need to determine if your method of hanging the monitor in this way would work? (if you feel like you have already answered this in the first question - enter "already stated".)

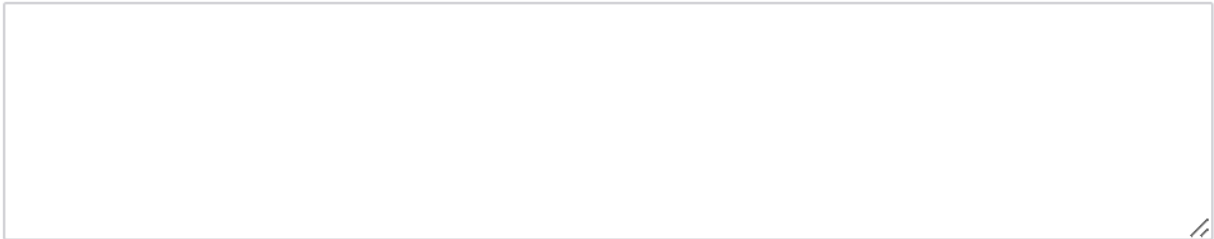
What assumptions or approximations would you need to make to determine if your method of hanging the monitor in this way would work? (if you feel like you have already answered this in the first question - enter "already stated".)

Draw any representation you think would be useful for solving this problem. Be sure to include important quantities you would need to consider.

Upload the file here.

Drop files or click here to upload

Which conditions need to be satisfied for the monitor to stay in place?



Which physics concepts would you use in solving?



Write an equation which would represent the conditions for the monitor to stay in place. Upload a screenshot of that equation below.

Drop files or click here to upload

If you would like to revise your initial diagram, please upload it here. If not, please continue to the next question.

Drop files or click here to upload

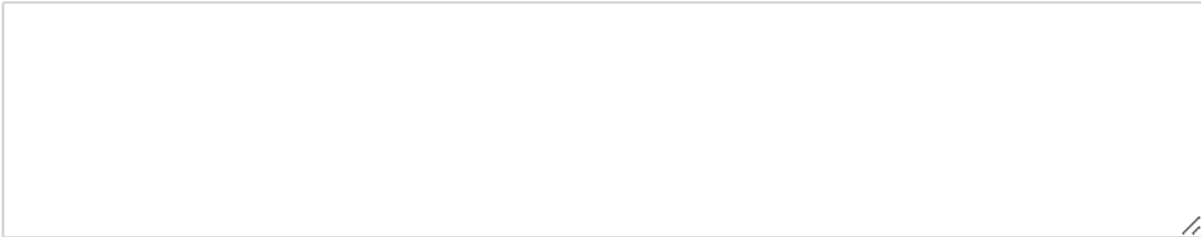
If your boss now said you should do experiments to collect the information needed to make sure this method would work, what measurements would you make? List your top three measurements you would make and rank them from most critical first, second most critical second, etc.

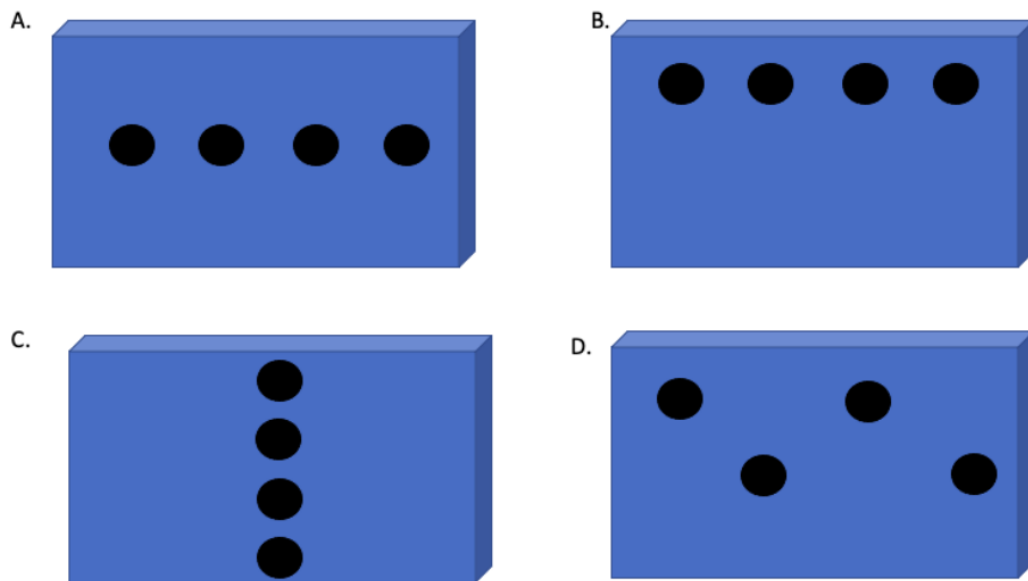
Imagine you test your method first in your office. What would be different about mounting the monitor in your office compared to on the ship?



Below are the assumptions/ approximations you made earlier to determine if your method of hanging the monitor in this way would work:

If you would like to revise these assumptions, do so below. If not, please continue to the next question.





Above are four configurations for the placement of the magnets on the back of the monitor. Which magnet configuration do you think would work the best for successfully mounting the monitor? Rank the configurations in order of best (1) to worst (4).

- A
- B
- C
- D

Explain your reasoning for the order that you ranked the configurations in the question above. If you think there is a better configuration not depicted in the question above, please describe it here and explain why it is better.



Appendix B

B.1 Complete transcripts for selection of students

None:

Okay great so, since they're both labeled as theta they're equal to one another right?
Great. Alright so this will make this a triangle. Without solving the problem yourself determine whether Jessica's answer, Sarah's answer, or neither is correct.
Okay, well, ooh...so, we have one component that is going downwards. Okay, we have tension force going up at an angle. Correct. Now, if we're using this angle now the only question we have to break it into its x and y components, right?
So, we only really care about the vertical component because that's what's going to counteract force of weight pulling downwards.
So F_y is equal to F exerted by gravity plus...let's see opposite over hypotenuse. So it is $\sin^2 \theta$. Sine theta is what you get. Um, there the coefficient here is gonna be F_1 or...what happened to...greater than or equal to zero.
without solving the problem yourself determine if Jessica's answer, Sarah's answer, or neither is correct.
Neither is correct, why the hell would they divide? I don't know if that's right though. You need the...As written neither of their equations gives...No, no, no, no it does.
Yeah, yeah, yeah, yeah because FT_1 is equal to, if I remember correctly, I don't know why they have a W . It was $\cos \theta$ of the angle 1, $\cos \theta$ of the angle 2, divided by FT , right?
We're seeing that the rope is the same, so $FT_1 = FT_2$. So 2, the angle has to be the same. If the angle would be the same then they both are together.
Why would you divide by...it doesn't seem right that you're dividing, that you're getting. I think I'm gonna disappoint you on this one, I'm a little stuck.
I don't think we should be dividing the weight by an angle. That just doesn't seem right to me, maybe it's because I haven't seen it before. It just doesn't seem right.
We want to know the tension force and this...so why would we set this relate...Oh. Oh, oh, oh. F_{rope} , $F_{r2} \sin \theta$ oh I get what they did. T
They put a W . Oh, oh, oh, oh we're getting somewhere. $F_{r2} \sin \theta - W$ is equal to zero. That's an expression for static equilibrium.
They're both right. They're just using...no. No, no. Jessica's not right. I don't know why she's using cosine. Because if you're using cosine, it would be adjacent which is this guy over hypotenuse, and we don't want the x-component.
Jessica figured out the x-component, we don't want that. Sarah on the other hand figured out the y-component and that's what we want. I think Sarah's correct

This student is from the group with no prior physics experience, and they are checking the answer to the chandelier problem. The student starts out by writing down force balances and drawing a force diagram, essentially following a memorized procedure until they remind themselves of what the task at hand is. They then start to parse the expression a bit more carefully (why is it divided by the angle), but again they are connecting to their memories of what the answers of problems like this are supposed to look like. They then cycle back and forth between these three behaviors, eventually arriving at the correct solution. Interestingly though, the logic for justifying the solution is still quite algorithmic: it has to be sine theta because that is the y-component of the force. At no point during this episode do we see evidence of conceptual manipulation of the system or expressions. The strategy shifts are frequent, but the overall framing of the task remains relatively consistent.

Intro:

Two students, Sarah and Jessica are working together on studying for their physics midterm. ON the practice exam they come across this question. You have to change a lightbulb in your basement. To reach the lightbulb you lean a lightweight aluminum ladder of length l against the wall such that it makes an angle α with the floor. The floor and the wall are both made of concrete which has a coefficient of friction μ with the aluminum ladder. What is the maximum height off the ground h to which you can climb before the ladder starts to slide down and away from the wall. They decide to solve the problem independently, they compare their answers.

Their answers are: the height equals coefficient of friction times l over one over coefficient of friction squared sine squared α over cosine α plus μ sine α . H is equal to l over 1 plus the coefficient of friction squared cosine squared α over sine α plus μ cosine α . Without solving the problem yourself determine if Jessica's answer, Sarah's answer, or neither is correct. Without solving the problem. Hmm. I guess the further, the higher up you are l would increase.

Okay so look at the sine squared over cosine. The height would matter. Would the height matter more or the distance away? Height would matter more.

At the same angle, the greater height would result in μ against the wall being higher.

There would also be translational. Okay, so, so there's both at the bottom and at the top.

Is Jessica or Sarah correct? I think Jessica is wrong, so.

As l increases, so as l increase it would increase an angle I think of sine squared alpha. As l increases h increases by $\sin^2 \alpha$ not $\cos^2 \alpha$ because the higher you are up the more force you apply at that spot.

Unlike the student with no prior physics experience, this student starts off by thinking conceptually about the expression, rather than by drawing force diagrams and writing down equations. They think about how quantities should vary with respect to one another, as well as what variables should be in the final expression. They are very much thinking about the physical model that underlies this system but appear to be having trouble converting that into an appropriate mathematical expression. They ultimately rely on the shape of the sine vs. cosine graphs, but do not take into account the entirety of the expression, they only focus on one small piece of it, ultimately leading them to an incorrect conclusion. Interestingly, this student does shift between algorithmic and conceptual framing of the task, but does not really shift their strategy at any point.

Major:

Alright two students are working together. A lot of studying going on. Okay you need to change a lightbulb, you lean a...coefficient of friction μ . Oh, okay makes an angle α which, oh it has a coefficient of friction on both sides. Okay, what's the maximum height you can climb. So you're, you're climbing up the ladder okay. This, uh, determine who is right.

μ times okay so we know cosine of $\pi/2$ is zero. We know sine of $\pi/2$ is one. So we know that, and then we know that h of $\pi/2$ is, should like, okay wait. The limit as α goes to uh $\pi/2$ of h is infinity.

You should be able to go up forever because it doesn't, it won't slip.

So then we know that let's do Sarah's thing. Okay, so we can actually just see which one's right, we know that Sarah's right because it's over cosine.

This physics major clearly demonstrates that they are thinking about limiting cases of the expression and making predictions about how the system behaves. However, they start this

approach in a very algorithmic way, plugging $\pi/2$ and 0 into the expression and calculating the limits before making any physical predictions. After they get the results of their calculations, they do then bring in physical reasoning to make a correct conclusion about the solution in front of them. This represents a stable strategy with some frame shifting again, indicating that these frame shifts may be necessary for these types of tasks.

Grad:

Two students, Sarah and Jessica, are working together on studying for their physics midterm. On the practice exam, they come across this question:

Consider a chandelier of weight W that is attached to the ceiling by two ropes of equal length, as shown below. One end of each rope is attached to the center of the chandelier, and the other end is attached to the ceiling. And θ . How strong does the rope have to be as a function of the weight of the chandelier?

They decided to solve the problem independently.

Okay, so obviously the strength of the rope will be proportional to the weight .

So based upon the numerator, I would say that both of them have possibly be answering incorrect.

Now, as far as that is concerned, if you are at a smaller θ That means a smaller component of the tension is balancing out in the vertical direction with the mass, not with the weight of the block

And since you have a smaller component of tension. Balancing out of the weight of the block, that means your tension overall has to be much larger to balance out the weight of the block.

So what I'm trying to say is that are small θ , the tension would have to be much greater in both ropes and therefore we require much much more strength.

Um, and so what I would expect is I would expect this f rope to increase with decreasing θ . So I would expect F to increase or decrease in θ . Yes.

Denominator. Expect F rope to increase. With decreasing θ . And the only one that does that is Sarah's answer because $\cos \theta$. No is that right?

Sarah's answer is correct. Answer is correct Because $\sin \theta$ increases with θ at least between zero and 90 degrees. Whereas $\cos \theta$ decreases with θ in the same range.

This graduate student again relies on making predictions to explain how the system will behave.

They are able to immediately identify that the answer should be proportional to the weight of the object because that is the only force in the problem. They then think about how cosine and sine

vary with angle and how that is related to the variation in tension that would physically occur. Unlike the physics major, at no point do they start to plug in different values, they use entirely variational reasoning. There is no shift out of the conceptual frame, nor is there a shift of strategy.

Appendix C

C.1 Solutions to the estimation problems

Assignment 2:

Problem 3. What is the combined distance walked by all the humans in the history of the world? Please state your assumptions clearly and describe how you arrive at your estimation.

1 Getting Started

This question boils down to estimating the distance walked by an average human being and then estimating how many humans have roamed the planet since the dawn of humanity.

2 Devise Plan

To obtain the distance walked by a single person in a lifetime, I can begin by estimating the distance walked by a single person in a given day and then multiply by the number of days in a year and by the average number of years

in a lifetime. Then all I need is to do is multiply by the number of humans that have lived in the history of the world. This estimation is perhaps the trickiest, but I know the population of Earth is about 7 billion and I also know that the population has grown rapidly over the last decades.

3 Execute Plan

If d is distance walked by a single person in a day, then the distance D walked in a year is $D = 365d$. If the average number of years in a life time is T , and N is the number of humans in the history of the world, then distance walked by a single person in a lifetime is DT , and the answer I am looking for is NDT .

From tracking steps, I know that I walk around 10,000 steps/day, which at 0.5 m/step yields $d = 5$ km. The lifespan in developed countries is around 80 yrs, but for the entire world and over the history of humanity, I'll put that at a lower value: $T = 60$ yrs. So, the distance walked by an average human of their lifespan is $(5 \text{ km/day})(365 \text{ days/yr})(60 \text{ yrs}) \approx 10^5$ km.

Now the total number of humans that have ever lived on earth. I know the current population is around 7 billion (order 10^{10}), and so I know that the total number of humans that have ever lived should be more than 10^{10} . Also, because the population has increased steeply in the last several hundred years, 10^{12} is probably unreasonably high. Using this alone, I can state with reasonable confidence that the total number of humans to have ever lived is of the order 10^{11} .

Multiplying the distance walked in a person's lifespan by the number of people N , I get for the total distance walked by all humans who ever lived

$$(10^5 \text{ km})(10^{11}) = 10^{16} \text{ km}.$$

4 Evaluate Answer

Wow! I had no idea an average human would walk 100,000 km in a lifespan—about two-and-a-half times the circumference of the globe! Is that number reasonable? I know that a car can easily go that distance over its lifespan. But cars average at 10x higher speeds than humans. On the other hand, humans live about 10 times longer than an average car. So, maybe the result is not that crazy. The only assumption that goes into the calculating that distance is the 5 km/day. I actually checked my phone and saw that I do indeed average about 7,000 steps or 3 miles/day, which is not far off. But even if I had taken 2 km/day, I would still have ended up with over 40,000 km in a lifespan, which gives the same order of magnitude (10^5 km). The other assumption in my calculation is the number of people who ever lived (something no one knows with any accuracy) and I feel pretty confident that 10^{12} would be too high and 10^{10} would be too low. That makes me feel reasonably confident about my answer.

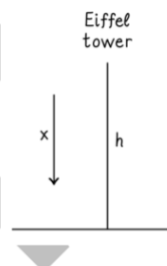
Problem solving tip: Yikes! This all sounds a bit wishy-washy! Is that science? You may think this type of exercise is off the deep end and that there are too many assumptions, oversimplifications, and uncertainties. While you wouldn't want to bet your fortune on the answer (even though no one really knows the answer), the power of developing estimation skills is that it helps you develop a general feel for a problem. You can put this skill to use in many ways: check on a solution you obtained; quickly see if something someone tells you makes any sense at all; get a feel for the solution before you seriously embark on working it out in detail, etc. Often, you will find that the answer you obtain is within an order of magnitude of the actual answer.

Assignment 3:

Problem 3. A penny is dropped from the top observation deck of the Eiffel Tower in Paris, France. Estimate its speed upon hitting the pavement. Please state your assumptions clearly and describe how you arrive at your estimation.

1 Getting Started

To get started, I make a sketch to visualize the problem. I represent the Eiffel tower in the simplest possible way — a vertical line. As the penny will fall straight down, I'll choose the x axis to point downward. The penny will accelerate downward under the influence of gravity. As the penny falls in air, it will experience air resistance during its fall, but as I haven't learned (yet) how to deal with air resistance, I will ignore this air resistance for now and revisit my simplification at the end of the problem. By assuming the penny only experiences gravitational force, its motion is reduced to constant acceleration.



2 Devise Plan

As I don't know the height h of the Eiffel Tower from the top of my head, I will need to come up with a way to estimate it. Once I have h , I can use Eq. 3.7 again to calculate the penny's speed upon hitting the pavement.

3 Execute Plan

I could Google the height of Eiffel Tower, but that's not as much fun (and satisfying) as trying to estimate it. It should be of the same order of magnitude as a skyscraper in, say Manhattan. I know from visiting New York, that the taller skyscrapers have around 100 stories and the height of a typical floor is about 3 m. Therefore

$$\Delta x \approx 100 \times 3 \text{ m} = 300 \text{ m}$$

I'll use the constant-acceleration Eqs. 3.4 and 3.7 from the textbook,

$$\Delta x = v_{x,i}\Delta t + \frac{1}{2}a_x(\Delta t)^2$$

$$v_{x,f} = v_{x,i} + a_x\Delta t$$

Because the penny is dropped and not thrown down, I can safely assume that $v_{x,i} = 0$. And as I'm ignoring air resistance, $a_x = +g$. I can then obtain the time travelled from the first equation and substitute the expression in the second equation to obtain the final speed upon hitting the ground:

$$v_{x,f} = g\Delta t = g \sqrt{\frac{2\Delta x}{g}} = \sqrt{2g\Delta x} \approx \sqrt{2(9.8 \text{ m/s}^2)(300 \text{ m})} \approx 10^2 \text{ m/s}.$$

Thus, my estimation gives around 100 m/s for the penny's speed as it hits the ground.

4 Evaluate Answer

Note that I rounded my answer to the nearest power of 10. My calculator gives 77 m/s (with a lot more digits actually), but as this is a (very rough) estimation, all I can really do is provide an order of magnitude.

That's a *really* high speed — about twice the speed of a fast baseball pitch (almost 300 km/h or 200 mph)! What if my estimated height of the Eiffel tower is not correct? Say it's 150 m high. Or 600 m high. As you can easily verify, that would change the final speed by just a factor of $\sqrt{2} \approx 1.4$. So, even if my estimate is off by a factor of 2, I'm within an order of magnitude. Ignoring air resistance, as I did, is more problematic — I know that if I stick my hand outside a car going 55 mph on a highway (a lot less than the speed I obtained), my hand experiences a large force. I don't really have another choice, as I don't yet know anything about air resistance.

Given the (unreasonable) assumption to ignore air resistance, however, the speed I obtained is not off-scale — a pitcher can reach half that speed in a simple arm throw (but probably not with a penny).

At least my calculation tells me that the penny cannot go *faster* than 100 m/s!

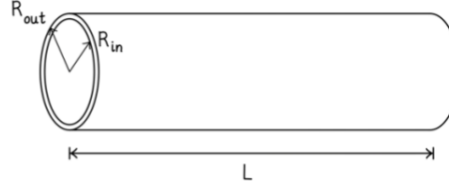
Problem solving tip: This problem may leave you with an uncomfortable feeling. The assumption that air resistance can be ignored is simply not reasonable. However, note the value of the last sentence: At least I know that the penny cannot go faster than the speed I obtained. So at least I have a reasonable upper bound for the penny's speed. If someone were to tell you, "If you drop a penny from the top of the Eiffel tower, it will reach the speed at which a jetliner flies," knowing that jetliners fly at around 500 mph, you can tell that person with full confidence that they are wrong. Now *that* is cool!

Assignment 6:

Problem 3. The Keystone pipeline is among the longest oil pipeline systems in the world. It carries oil 3,500 kilometers (2,200 mi) from Alberta in Canada all the way to Texas in the US. To reduce pollution, an engineer suggests that the entire pipeline should be emptied and raised 1 m above the ground. Estimate the amount of energy this would require.

1 Getting Started

To get started on this problem, I make a sketch of the pipeline, which essentially is a cylindrical tube of length L with inner diameter R_{in} and outer diameter R_{out} . The pipeline needs to be raised by 1 m.



2 Devise Plan

Raising the pipeline increases the gravitational potential energy of the Earth-pipeline system and this energy must be supplied to lift it. I can use Eq. 7.19, $\Delta U^G = mg\Delta x$, to calculate this energy. I know the values for g and Δx , so I will need to estimate the inertia m of the entire pipeline. I have seen pictures of pipelines and I'm eyeballing their diameter to be around 1 m. They need to be relatively robust, so I'll assume the walls are made from steel and are about $\frac{1}{2}$ " thick, which is around 0.01 m. This should permit me to calculate the volume of steel. If I then estimate the density of steel, I can obtain an estimate of the inertia of the entire pipeline and calculate the energy needed.

3 Execute Plan

For a given length L of pipeline, I can calculate the volume of steel by subtracting the inner empty volume of the pipe from the total (outer) volume:

$$V_{steel} = V_{pipe} - V_{empty} = \pi R_{out}^2 L - \pi R_{in}^2 L = L\pi(R_{out}^2 - R_{in}^2),$$

and so the volume of steel per unit length is

$$\frac{V}{L} = \pi(R_{out}^2 - R_{in}^2) = \pi[(0.50 \text{ m})^2 - (0.49 \text{ m})^2] = 0.03 \text{ m}^2.$$

I know that the density of water is about $1 \times 10^3 \text{ kg/m}^3$, and the density of steel is significantly higher. I'll take it to be 10 times that of water. Because density is defined as the inertia per unit volume ($\rho = m/V$), I have

$$\frac{m}{L} = \frac{V}{L} \rho = (0.03 \text{ m}^2)(1 \times 10^4 \text{ kg/m}^3) = 300 \text{ kg/m},$$

which means that 1 meter of pipeline has an inertia of 300 kg. The pipeline is 3,500 km long, so the total inertia of the pipeline is

$$m = (300 \text{ kg/m})(3.5 \times 10^6 \text{ m}) \approx 10^9 \text{ kg}.$$

Using this inertia, I can now determine the amount of energy required to raise the pipeline by 1 m:

$$\Delta U = mg\Delta h = (10^9 \text{ kg})(10 \text{ m/s}^2)(1 \text{ m}) = 10^{10} \text{ J}.$$

4 Evaluate Answer

Searching online, I find that pipeline diameters vary from 17 to 40 inches (or 0.43 m to 1.01 m) and their wall thicknesses range from 9.3 mm to 13 mm. My estimate of the diameter is at the upper end and that of the thickness at the lower end, so my estimate of the volume of steel per meter of pipeline is definitely in the right ballpark. Looking up the density of steel, I find a value $8.05 \times 10^3 \text{ kg/m}^3$, close to my estimate. Therefore, my answer of 10^{10} J to raise the pipeline can't be far off.

With some further research online, I find that the Keystone pipeline has a capacity of 1.5 million cubic meters of oil per day. A good number to remember is that 1 liter of fuel contains on the order of 30 MJ of chemical energy, so in a year the pipeline transports about 10^{18} J (another good number to remember is that number of seconds in a year is $\pi \times 10^7$ to within a fraction of a percent). Thus, the amount of energy required to raise the pipeline 1 m is miniscule compared to the amount of energy in oil that it transports in a year. Cool!

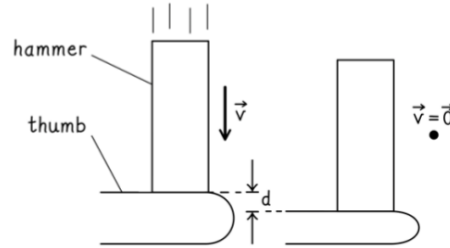
Problem solving tip: For many problems you can use estimation procedures like the one above to get a feel for the answer in the Evaluate Answer step and see if the answer your obtained makes sense. As this last problem already is an estimation problem, it generally doesn't make much sense to use estimation to "evaluate the estimation." So, unless you can come up with a totally different estimation approach, it's OK to look up some facts to gain confidence in your answers. Note finally, how a repertory of "handy numbers to remember" is useful — the density of water, energy content of fuel, seconds per year, etc. If you train yourself to carry out these estimations regularly — when reading the news, thinking about facts you hear mentioned, etc. — you'll develop such a list automatically. Additionally, you'll start feeling much more confident of your ability to make sense of information in unfamiliar domains. A powerful skill that will permit you to tread new ground more easily.

Assignment 7:

Problem 3. While helping your young niece build a birdhouse you demonstrate how to use a hammer. Unfortunately, you miss the nail and bring the hammer down on your thumb which is holding the nail. Estimate the impulse the hammer delivers to your thumb.

1 Getting Started

The hit of the hammer is a collision between the hammer and (ouch!) my thumb, and as in any collision, momentum is transferred between the hammer and the thumb. I'll assume that the collision is totally inelastic. The impulse delivered by the hammer to my thumb is due to the contact force F_{ht}^c that is exerted by the hammer on my thumb. That force forms an interaction pair with the force F_{th}^c exerted by my thumb on the hammer, which slows the hammer down to a stop as it strikes my thumb. I make a sketch representing the impact of the hammer from the instant it first touches my thumb until it comes to rest after "compressing" my thumb over a distance d .



2 Devise Plan

The impulse delivered by a force is given by Eq. 8.25 and requires knowing both the magnitude of the force exerted on my thumb and the time interval over which this force is exerted. I'll assume the hammer undergoes a constant acceleration as it slows down over the distance d and use kinematics to determine the hammer's acceleration a . To do that, I'll need to estimate the velocity the hammer has just before it strikes my thumb. I can then use Eq. 3.4 to determine the product $a\Delta t$ of the acceleration a and the time interval Δt over which the hammer comes to a stop.

Because the vector force sum of the forces exerted by my thumb on the hammer is given by the product of the hammer's acceleration and its inertia m , I can multiply my estimate of $a\Delta t$ with m to obtain the impulse.

3 Execute Plan

If the hammer has speed v just before hitting my thumb and I treat the collision as being totally inelastic (that is, the hammer comes to rest against my thumb), I can use Eq. 3.13 to determine the hammer's acceleration:

$$v_{x,f}^2 = v_{x,i}^2 + 2a_x\Delta x,$$

or

$$a_x = \frac{v_{x,f}^2 - v_{x,i}^2}{2\Delta x}.$$

Given that $v_i = v$, $v_f = 0$, and $\Delta x = d$, I can write for the magnitude of the acceleration

$$a = \frac{v^2}{2d}.$$

Eyeballing the motion of a hammer, I estimate it to move about a meter in half a second or about 2 m/s. As it hits my thumb it is stopped over a distance that is much shorter than the thickness of my thumb — I'll take it to be 10 mm. Substituting these values, I get

$$a = \frac{(2 \text{ m/s})^2}{2(0.01 \text{ m})} = 200 \text{ m/s}^2.$$

If I take the inertia of the hammer to be 1 kg, the vector sum of the forces exerted on the hammer is

$$\sum F_h = ma = (1 \text{ kg})(200 \text{ m/s}^2) = 200 \text{ N}.$$

The vector sum of the forces exerted on the hammer consist of the force exerted by my thumb on it plus the force of gravity. I'll assume that during the interaction with my thumb the force F_{th}^c exerted by my thumb on the hammer is the only force, and so

$$\sum F_h \approx F_{th}^c = F_{ht}^c = 200 \text{ N}.$$

The time it takes for the hammer to slow down is given by $v_{x,f} = v_{x,i} + a\Delta t$, and so

$$\Delta t = \frac{v_{x,f} - v_{x,i}}{a_x} = \frac{0 - (-2 \text{ m/s})}{200 \text{ m/s}^2} = 0.01 \text{ s}.$$

Now I can determine the magnitude of the impulse delivered by the hammer to my thumb:

$$J = F_{ht}^c \Delta t = (200 \text{ N})(0.01 \text{ s}) = 2 \text{ N} \cdot \text{s}.$$

4 Evaluate Answer

Hitting my thumb with a force of 200 N is equivalent to placing a 20-kg weight on it — like one leg of a pretty heavy 4-legged 80-kg table. From experience I know this would be pretty painful, just like the blow of a hammer would! Googling the weight of a hammer, I see hammers are between 1 and 2 kg, so my estimate is within the right range.

The impulse — the momentum transferred to my thumb — is not that large, but that is because the duration of the impact is so short. The energy transferred is equal to the kinetic energy of the hammer, which is

$$K = \frac{1}{2}mv^2 = \frac{1}{2}(1 \text{ kg})(2 \text{ m/s})^2 = 2 \text{ J}.$$

That is not a very large amount of energy — equivalent to dropping a 1-kg block from a height of 20 cm —, but this energy would be delivered to a rather small area of my thumb.

Another way to calculate the impulse delivered to my thumb is to determine the momentum of the hammer, as all of it is transferred to me:

$$p = mv = (1 \text{ kg})(2 \text{ m/s}) = 2 \text{ kg m/s},$$

which is the same answer, as 1 kg m/s is equal to 1 N s.

Finally, I note that the acceleration of the hammer, 200 m/s² or 20 *g*, is very large. This is to be expected as the hammer comes to a stop over a very short distance (10 mm) and a very short time interval (0.01 s). That also means that for the purpose of this estimation problem, I can safely ignore the effects of gravity.

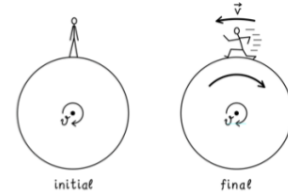
Problem solving tip: Note that for this problem too, I found a much simpler way of calculating the impulse during the last step, Evaluating Answer: just calculate the initial momentum of the hammer. Maybe you thought of that approach right away (good for you!). I didn't. Problem solving is not a straight-forward, clean-cut, one-shot activity. What's even more important: for any real, authentic problem there usually isn't a single valid solution.

Assignment 10:

Problem 3. Estimate the change in angular velocity of the earth if the entire human population simultaneously faced West and took one step forward.

1 Getting Started

I begin by representing with a simple sketch. For simplicity I'll consider just one human initially at rest on the surface of Earth, itself also at rest. The person then takes a step forward, pushing Earth backward, causing it to rotate in a clockwise direction in my sketch. Choosing my system to include both the human and the Earth, I know that there are no external forces that are exerted on the system that affect rotation, and so conservation of momentum tells me that the angular momentum of the system is constant. I will choose my positive direction of rotation to be clockwise (the direction in which I expect Earth to rotate)



2 Devise Plan

In the initial state the angular momentum of both human and Earth are zero. After stepping forward, the human gains counterclockwise (negative) angular momentum and Earth must gain a compensating clockwise (positive) angular momentum. I will begin by determining the final angular momentum of the human after taking a step forward using Eq. 11.35, $L_{\theta} = rmv_t$. Once I have determined the angular momentum change for a single person, I can multiply this change by the human population. This change angular momentum causes a corresponding opposite change in the Earth angular momentum. The Earth angular momentum is given by Eq. 11.34, $L_{\theta} = I\omega_{\theta}$. Treating the Earth as a rigid sphere, I can use the expression $I = \frac{2}{5}MR^2$ from Table 11.3 to determine the Earth rotational inertia.

3 Execute Plan

I will denote an average person's inertia by m_p and the velocity acquired by stepping forward by v . What value of r do I use in Eq. 11.35? I could set $r \approx R_E$, with R_E the radius of the Earth, but not everyone is at the equator. Most humans are closer to the axis of rotation than R_E . If I assume the bulk of Earth's population to be at 45° latitude the distance to the axis of rotation is $R_E/\sqrt{2}$, so

$$L_{p\theta,f} = -m_p v R_E / \sqrt{2},$$

where I have added a minus sign to indicate that this angular momentum is counterclockwise. Denoting the human population by N , I must multiply the above expression by N to obtain the final angular momentum of all humans ON Earth stepping West.

Given that the initial angular momentum of the system is zero, I know that

$$L_{p\theta,f} + L_{E\theta,f} = -Nm_p v R_E / \sqrt{2} + \frac{2}{5} M_E R_E^2 \omega_{E\theta,f} = 0.$$

Solving for the Earth final rotational velocity $\omega_{E\theta,f}$, I get

$$\omega_{E\theta,f} = \frac{5}{2} \frac{Nm_p v R_E}{M_E R_E^2 \sqrt{2}} = \frac{5}{2} \frac{Nm_p v}{M_E R_E \sqrt{2}}.$$

I'll take the inertia of an average human to be $m_p \approx 50$ kg, and I know that the current population is around $N \approx 8$ billion people. I'm going to assume that the velocity of an average person's step is $v \approx 1$ m/s, and I remember from earlier in the course that Earth's radius is about $R_E = 6,400$ km. Next I need to estimate Earth's inertia. Assuming Earth to be a sphere, knowing that the radius is about 6,400 km, I can estimate the volume of Earth to be

$$V_{\text{earth}} \approx \frac{4\pi r_{\text{earth}}^3}{3} \approx 10^{21} \text{ m}^3.$$

I know that the density of rock is significantly larger than that of water — rocks sink — so I'll assume it is about 3 times the density of water. However, the inner core of the earth might have a higher density — maybe 10 times that of water. So overall, I estimate the average density of the earth to be about 6 times that of water, which gives me

$$\rho_E \approx 1 \text{ g/cm}^3 \times 6 = 6 \text{ g/cm}^3 = 6,000 \text{ kg/m}^3.$$

From the above estimations of volume and density, I obtain the inertia of the earth

$$M_E = \rho_E V_E \approx (10^{21} \text{ m}^3)(6 \times 10^3 \text{ kg/m}^3) \approx 10^{25} \text{ kg}.$$

Using these values, I obtain

$$\omega_{E\theta,f} = \frac{5}{2} \frac{Nm_p v}{M_E R_E \sqrt{2}} = \frac{5}{2} \frac{(8 \times 10^9)(50 \text{ kg})(1 \text{ m/s})}{(10^{25} \text{ kg})(6.4 \times 10^6 \text{ m})\sqrt{2}} \approx 10^{-20} \text{ rad/s}.$$

4 Evaluate Answer

My estimation shows that the change in Earth's angular velocity is a mere 10^{-20} rad/s when every human faces West and takes a step forward. Online I find that Earth's angular velocity is about 7.3×10^{-5} rad/s, so the 10^{-20} rad/s change would not be noticeable. This reminds me of the problem earlier in the semester where I calculated the change in the (linear) momentum of Earth as the entire population gathers in one spot and jumps up. In that case too, the effect was too small to be noticeable. In both cases, the change is so small because of the 13 orders of magnitude difference between Earth's inertia (10^{25} kg) and the human population's inertia $(8 \times 10^9)(50 \text{ kg}) \approx 10^{12}$ kg.

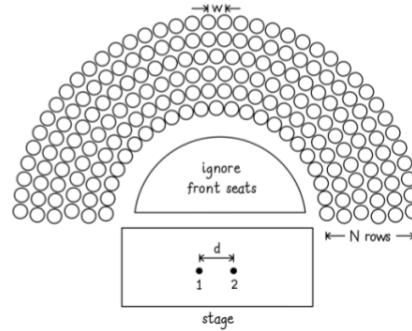
Problem solving tip: Note that we reuse some values we've used or estimated before, like the Earth radius. By doing these estimations regularly, you'll end up remembering a lot of these "good-to-remember" values and build up an arsenal of them that will make things go faster.

Assignment 12:

Problem 4. Estimate the number of seats in Sanders Theater where you would hear a reduced sound level if two clarinet players on center stage standing directly next to one another played the exact same tone at 440 Hz.

1 Getting Started

I begin by making a sketch of the situation. Having been to Sanders, I know that the seats are arranged in a semicircular arrangement of seats. For simplicity, I will ignore the seats right in front of the stage and those on the balcony. I'll denote the number of semicircular rows by N and the separation between adjacent seats by w (the width of a seat). I will treat the clarinet players each as a point source, separated by a distance d . From Chapter 17 I know that the sound waves emitted by two point sources can interfere with each other, resulting in destructive interference that reduces the sound level at certain locations. As shown in Figure 17.13 the locations of destructive interference lie on nodal lines that radiate out from the point between the two sources. My task is to estimate how many seats like on a nodal line.



2 Devise Plan

On p. 442 in Chapter 17 it says that the number of nodal lines on either side of a straight line connecting two point sources is the greatest integer smaller than or equal to $2d/\lambda$. Knowing the frequency, I can determine the wavelength and using an estimate for the distance between the players, I can determine how many nodal lines intersect the semicircular arrangement of seats. Next, I can model the geometry of Sanders theater to estimate the number of seats affected by the nodal lines.

3 Execute Plan

I begin by calculating the wavelength, given the frequency provided (440 Hz), and the speed of sound given in the textbook (roughly 340 m/s):

$$\lambda = \frac{c}{f} = \frac{340 \text{ m/s}}{440 \text{ Hz}} = 0.77 \text{ m}.$$

As the two clarinet players are standing right next to each other, I estimate the distance between them to be 1 m. I can then use the principle on p. 442 in Chapter 17 to determine the number of nodal lines

$$n \leq \frac{2d}{\lambda} = \frac{2(1 \text{ m})}{0.77 \text{ m}} = 2.6$$

Because the number of nodal lines must be an integer number, I conclude that there are 2 nodal lines on either side of the line connecting the clarinet players. As the audience is entirely in front of them, I only need to consider one side of the line, and so there are two nodal lines traversing the audience (see, for example Figure 17.17).

I know Sanders Theater can accommodate CS50, which enrolls about 750 students. The theater therefore likely has around 1000 seats, some of which are balcony seats. Ignoring the balcony seats, the bottom level has around 750 seats. As can be seen in my sketch, the rows of seats become longer (and contain more seats) as we move further away from the stage. (Note that for simplicity, I also ignored the region in the middle of the theater near the front, where seats are usually arranged in straight rows instead of semicircles. As there are not many seats in this area, this should not affect my estimation by much.)

To calculate the number seats affected by the two nodal lines, I first need to estimate the number of rows N in the seating area. I estimate that the average distance between adjacent seats in a row is 0.80 m (taking into account aisles and empty spaces) and that the first row contains about $N_1 = 20$ seats. This means the first row has a perimeter of

$$0.80 \text{ m/seat} \times 20 \text{ seats} = \frac{1}{2}\pi D_1,$$

where D_1 is the diameter of the first semicircular row. Solving for D_1 , I get

$$D_1 = 10 \text{ m}.$$

Estimating that the distance between adjacent rows is also 0.80 m, the diameter of the second row is $D_2 = 10 \text{ m} + 2 \times 0.8 \text{ m} = 12 \text{ m}$ and its perimeter is

$$0.80 \text{ m/seat} \times N_2 = \frac{1}{2}\pi D_2.$$

Substituting $D_2 = 12 \text{ m}$ and solving for N_2 yields

$$N_2 = 24 \text{ seats}.$$

So each subsequent row has 4 seats more than the previous row, forming an arithmetic series:

$$N_n = N_1 + 4(n - 1) \quad n = 1, 2, 3, \dots$$

As the number of rows is N , the last row (the row for which $n = N$) has $N_N = N_1 + 4(N - 1)$ and the sum of the arithmetic series is

$$\sum_{n=1}^N N_n = \frac{N}{2} [2N_1 + 4(N - 1)] = \frac{N}{2} [40 + 4(N - 1)] = 2N^2 + 18N.$$

Using my estimate for the total number of seats, I get the following quadratic equation for the number of rows N :

$$2N^2 + 18N = 750.$$

Solving this equation for N yields

$$N \approx 15.$$

The last thing I need to calculate is the probability that the nodal line encounters an audience member's head (as opposed to empty space between seats). I estimate that the width of a person's head is about 0.2 m. As the average distance between adjacent seats in a row is 0.80 m, there is a $(0.2 \text{ m})/(0.8 \text{ m}) = 25\%$ chance of the nodal line encountering a person's head. Thus, I estimate that there are

$$(\text{number of nodal lines}) \times (\text{number of rows}) \times (\text{percentage chance}) = 2 \times 15 \times 0.25 \approx 8 \text{ seats}$$

on the main floor of the Sanders Theater where the sound level from the two clarinets is reduced.

4 Evaluate Answer

Looking at the actual seating plan of Sanders Theater (<https://ofa.fas.harvard.edu/boxoffice/sanders-theatre-seating-charts>), my estimate of 15 semicircle rows in the theater is not far off from the actual number of rows (10 rows). The difference is likely due to the considerable number of seats in the front middle section, which I ignored for simplicity. However, this overestimation is partially offset by my *underestimate* of the number of seats in the first row (there are 32 seats instead of 20).

Now, it seems strange that if two clarinets are playing, there would be areas where you would hear nothing. Having been to many concerts, I have never experienced two “mute” players! However, in reality things are more complicated than the simple picture I assumed: 1) we have two ears so one ear might hear less sound, but the other wouldn’t; 2) the clarinets don’t emit pure sine waves at 440 Hz and there are many higher frequencies which form the unique timbre of the instrument’s sound (see Chapter 17, p. 438); 3) the clarinets are unlikely to emit exactly the same sound intensity; if they emit different intensities, for the interference cannot be completely destructive; 4) and, finally, 5) the walls reflect sound back causing the nodal lines to further blur out.

Problem solving tip: OK, I could probably have eyeballed the number of rows in Sanders theatre and I would probably have chosen a number somewhere between 10 and 20, saving myself a lot of trouble and obtained the same result. It’s still great to see how all the things we have learned work out!