Spring 5-10-2019

Using Question Variations to Access Alternate Student Thinking About the Same Physical Situations

Jeffrey Mathis Hawkins
University of Maine, jeffrey.hawkins@maine.edu

Follow this and additional works at: https://digitalcommons.library.umaine.edu/etd
Part of the Physics Commons, and the Science and Mathematics Education Commons

Recommended Citation
Hawkins, Jeffrey Mathis, "Using Question Variations to Access Alternate Student Thinking About the Same Physical Situations" (2019). Electronic Theses and Dissertations. 3037.
https://digitalcommons.library.umaine.edu/etd/3037

This Open-Access Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.
USING QUESTION VARIATIONS TO ACCESS ALTERNATE
STUDENT THINKING ABOUT THE SAME PHYSICAL
SITUATIONS

By
Jeffrey Mathis Hawkins
B.S. Physics, University of Washington, Seattle, 2007
B.S. Astronomy, University of Washington, Seattle, 2007

A DISSERTATION
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
(in Physics)

The Graduate School
The University of Maine
May 2019

Advisory Committee:
John R. Thompson, Professor, Co-Advisor
Michael C. Wittmann, Professor, Co-Advisor
Brian W. Frank, Associate Professor
Natasha M. Speer, Associate Professor
Jonathan T. Shemwell, Associate Professor
Assessing students’ understanding of physics is a critical element for improving physics education. This dissertation presents and demonstrates a method to better assess student understanding of physics by varying the types of questions asked of students. The method presented herein, which is based on the Specific Difficulties and Resources frameworks, involves modifying the task students are asked to complete when analyzing the physics of a particular physical situation. Creating question variations that ask students to address a particular correct or incorrect outcome of the physical situation, eliminate an incorrect outcome, or justify a correct outcome provided to them can provide more information about students’ physics ideas. This dissertation applies this methodology across multiple content areas to see how often these question variations elicit novel ideas as well as investigate patterns in students’ responses to these different variations across multiple content areas.

These variations provide a more detailed view of students’ ideas by demonstrating that students use some ideas exclusively in response to variations or that they use ideas differently. In some cases, providing students with the correct
outcome and asking them to justify it demonstrates that students can sometimes express correct reasoning to support the correct response more often than when it is not provided. Other times students treat the provided correct outcome as anomalous data and disagree with it, or accept it and use it as the basis for their reasoning. Furthermore, students are able to develop unique reasoning to justify why an outcome is not what occurs; and asking students to justify or eliminate the response “zero” often leads to a difference in the ideas they express. Finally, the results identify response options that a majority of a class has ideas to both justify and eliminate.

The additional information gathered from asking question variations has the potential to impact research by providing new tools to investigate students’ ideas, and to improve instruction by informing instructors about students’ ideas.
DEDICATION

This dissertation is dedicated to my mother Paula Mathis.
Many people deserve a tremendous amount of thanks for assisting me in the completion of this dissertation.

First, I would like to thank my co-advisors, Dr. Michael Wittmann and Dr. John Thompson, who established the Physics Education Research Laboratory (PERL) at the University of Maine (UMaine) and provided me with the environment and resources I needed to complete this work, become a member of the physics education research community, and grow as a person. I also appreciated their willingness to work together and advise me on the research presented in this dissertation, and I value the many lessons I have learned from them.

Second, I would like to thank Dr. Brian Frank, who went above and beyond in his duties as a post-doctoral researcher to assist me in planning, implementing, and presenting of this research, even continuing to work with me after leaving his role at UMaine.

I also want to thank many other people at the University of Maine, including my other committee members, Dr. Natasha M. Speer and Dr. Jonathan T. Shemwell, who both provided valuable resources and input while expanding my knowledge of discipline-based education research beyond the field of PER; all members of PERL, who worked together to create a great research community and made me a much more well versed and mindful researcher than I would have been able to become without them; and the faculty and staff in the Physics Department at the University of Maine, especially Dr. David Clark, Dr. George Bernhardt, and Frank Dudish, who continuously worked with me in implementing research projects in their courses.

I would also like to thank all members of the PER community, who were a more welcoming and supportive research community than I could have hoped for. I would particularly like to thank the all members of the Physics Education Group at the
University of Washington, who developed my interest in PER when I was an undergraduate student and researcher at the University of Washington—especially Dr. Peter Shaffer, for his patience and encouragement while advising me.

In addition, I would like to thank the National Science Foundation for funding this research through grants #DRL-0633951 and #DUE-0941191.

Finally, I want to thank my family and friends who have been instrumental in the completion of this project. It would not have been possible without their emotional support and encouragement. I would especially like to thank my mother for the countless ways she continued to support me throughout this entire process.
TABLE OF CONTENTS

DEDICATION ........................................................................................................ iii

ACKNOWLEDGEMENTS ...................................................................................... iv

LIST OF TABLES .................................................................................................. xvi

LIST OF FIGURES ............................................................................................... xxi

Chapter

1. INTRODUCTION ............................................................................................ 1
   1.1 Motivation for Investigating Students’ Ideas and Thinking in Physics .......... 1
   1.2 Improving Methods of Investigation .......................................................... 2
   1.3 Dissertation Overview ............................................................................. 3

2. MOTIVATIONAL LITERATURE REVIEW .................................................. 7
   2.1 The Research and Development Cycle .................................................... 7
   2.2 The Specific Difficulties Framework for Investigating Student Understanding ........................................ 10
   2.3 Global Difficulties .................................................................................. 11
      2.3.1 Difficulties Across Content Area .................................................... 12
      2.3.2 Difficulties Across Physical Situation .......................................... 13
   2.4 Previous Research on Question Variations ............................................ 13
      2.4.1 Question Representation ............................................................... 14
      2.4.2 Question Format ............................................................................ 16
5. STUDY DESIGN, ADMINISTRATION, TASK DESIGN, AND ANALYSIS METHODS ........................................................... 53

5.1 The Across-Content Study ........................................................................ 53

5.2 Administration Method ........................................................................... 54

5.3 Variation Design .................................................................................... 58
  5.3.1 Consider, Given Correct, and Given Information .............................. 59
  5.3.2 Eliminate ......................................................................................... 60

5.4 Pretest Design ....................................................................................... 62
  5.4.1 Questions Without Variations ......................................................... 62
  5.4.2 Number of Variations Tested per Question ....................................... 63
  5.4.3 Pretest Schedule ............................................................................... 65

5.5 Data Analysis ....................................................................................... 66
  5.5.1 Reasoning as Evidence of Ideas ....................................................... 66
  5.5.2 Coding Data .................................................................................... 67
  5.5.3 Differences in a Particular Reasoning or Response Between Question Versions .............................................................. 69
6. ACROSS-CONTENT STUDY EXAMPLE 1 - WORK-BY-WALL

   QUESTION ........................................................................ 73

6.1 Previous Research on Students' Understanding of Work................. 73

   6.1.1 Students' Difficulties with the Definition of Work.............. 74

   6.1.1.1 Inappropriate Use of an External Coordinate System .................. 74

   6.1.1.2 Failing to Use the Appropriate Displacement .................. 75

   6.1.1.3 Distinguishing Between a Force and Work .................. 76

6.1.2 Difficulties with Zero Displacement Forces and Systems............ 76

6.2 Designing Variations of a Research Task.................................. 78

   6.2.1 Previous Responses to the Work-By-Wall Question.............. 78

   6.2.2 Creating Variations ................................................. 79

   6.2.3 Design of the Whole Pretest ....................................... 81

6.3 Administration ..................................................................... 83

6.4 Discussion of the Coding ................................................... 83

   6.4.1 Content-Specific Reasoning Codes.................................. 84

   6.4.1.1 Coordinate System Dependent .......................... 84

   6.4.1.2 Pushes Back .............................................. 85

   6.4.1.3 Always Positive........................................... 86

   6.4.1.4 Force Means Work ....................................... 86

   6.4.1.5 Displacement Means Work .............................. 87

   6.4.1.6 Force and Displacement.................................. 88

   6.4.1.7 No Displacement (Correct Reasoning).................. 89

   6.4.1.8 Spring Does the Work ................................... 90
6.4.2 General Coding Categories - No Reasoning, Unclear, and Other ......................................................... 91
6.4.3 Coded Responses for All Versions ................................. 92
6.5 Analysis and Discussion .................................................... 95
   6.5.1 No Displacement Reasoning on Given Correct ............... 95
   6.5.2 Zero - Most Eliminated and Most Chosen Response .......... 98
   6.5.3 Force Means Work and Displacement Means Work Reasoning ..................................................... 102
6.6 Conclusion ................................................................... 103

7. ACROSS-CONTENT STUDY EXAMPLE 2 - CHANGE IN KINETIC ENERGY QUESTION ............................................................ 105
   7.1 Student Understanding of the Work-Energy Theorem ............... 105
   7.2 Research Task Design ....................................................... 108
      7.2.1 Design of the Variations ............................................ 108
      7.2.2 Design of the Pretest ............................................... 111
   7.3 Administration .............................................................. 113
   7.4 Discussion of the Coding ................................................... 114
      7.4.1 Content-Specific Reasoning Codes ................................. 114
         7.4.1.1 Compensation ............................................ 114
         7.4.1.2 More Velocity ............................................. 115
         7.4.1.3 More Mass ................................................ 116
         7.4.1.4 Equal Applied Force ..................................... 117
         7.4.1.5 Conservation .............................................. 118
         7.4.1.6 Unequal Elements ........................................ 119
7.4.1.7 More Mass or Unequal Elements ......................... 120
7.4.1.8 Need to Know .............................................. 120
7.4.2 General Coding Categories - No Reasoning, Unclear, and Other ................................................................. 121
7.4.3 Coded Responses for All Versions ............................. 121
7.5 Analysis and Discussion ............................................. 124
  7.5.1 More Common Incorrect Reasoning on Given Correct .... 124
    7.5.1.1 Equal Applied Force Reasoning on Given Correct ..... 124
    7.5.1.2 Compensation Reasoning on Given Correct .......... 126
    7.5.1.3 Accessibility of Reasoning for Given Correct ....... 127
    7.5.1.4 Incorrect Reasoning on Given Correct Conclusion .. 128
  7.5.2 Novel Unequal Elements Reasoning on Consider Equal To ... 129
  7.5.3 Interaction with Acceleration Question .................... 134
7.6 Conclusion.................................................................. 135

8. STUDENT RESPONSES TO GIVEN CORRECT VARIATIONS .... 139
  8.1 Introduction ................................................................ 139
  8.2 Research on Anomalous Data .................................... 140
  8.3 Anomalous Data in the Current Study ........................... 142
    8.3.1 Why This is an Appropriate Framework ................. 142
    8.3.2 Theoretical Framework Compatibility.................... 143
8.4 Chinn and Brewer’s Theoretical Framework ........................................... 145
  8.4.1 Responses to Anomalous Data .................................................. 145
  8.4.2 Conditions Leading to Particular Responses ........................... 146
    8.4.2.1 Entrenchment of Ideas ............................................. 147
    8.4.2.2 Availability of an Alternative Set of Ideas .................... 148
    8.4.2.3 Credibility of the Data ........................................... 149
    8.4.2.4 Ambiguity of the Data ........................................... 149
  8.5 Students Disagreeing with Given Correct ....................................... 149
    8.5.1 Ignore and Reject Response Categories ............................... 150
    8.5.2 Ignore and Reject Responses ......................................... 151
      8.5.2.1 Pretest 4: Car-Over-Hill Question ............................ 152
      8.5.2.2 Pretest 10: Change in Energy and Momentum Question ........ 153
      8.5.2.3 Pretest 12: Rotating Disk/Wheel Question .................. 154
    8.5.3 Correlation Between Difficulty and Disagreement .................... 156
    8.5.4 Discussion of Disagreement ........................................... 158
      8.5.4.1 Existence of Disagreement .................................... 158
      8.5.4.2 Correlation with Question Difficulty .......................... 161
  8.6 Students Reinterpreting Given Correct ........................................... 162
    8.6.1 Reinterpret Response Category ........................................ 163
    8.6.2 Reinterpret Responses ................................................ 163
    8.6.3 Discussion of Reinterpreting ......................................... 166
8.7 Students Excluding, Holding in Abeyance, or Claiming Uncertainty about Given Correct ......................... 168

8.7.1 Exclude, Abeyance, and Uncertainty Response Categories ...... 168

8.7.2 Exclude, Abeyance, and Uncertainty Responses .............. 170

8.7.3 Discussion of Exclude, Abeyance, and Uncertainty .......... 171

8.8 Students Accepting Given Correct ........................................... 172

8.8.1 Change and Peripheral Change Response Categories ........... 172

8.8.2 Change and Peripheral Change Responses ...................... 173

8.8.3 Discussion of Change and Peripheral Change .................. 174

8.9 Discussion ........................................................................ 175

8.9.1 Students Treat Given Correct as Anomalous Data .......... 175

8.9.2 Hierarchy of Response Types ...................................... 176

8.10 Conclusion ...................................................................... 178

9. GENERAL RESULTS FROM ALL CONTENT AREAS ................. 180

9.1 Introduction ..................................................................... 180

9.2 Methods for Comparing Variations ..................................... 181

9.3 Differences in Response Prevalence ................................... 182

9.3.1 Response Similarities from Eliminate ......................... 183

9.3.2 Response Differences from Consider .......................... 186

9.3.3 Differences in Response Prevalence Conclusion ............ 187

9.4 Unique Reasoning in Variations ........................................... 187

9.4.1 Unique Reasoning from Eliminate ............................... 188

9.4.2 Unique Reasoning from Consider ................................. 193
9.4.3 Unique Reasoning from Given Correct .................................................. 193
9.4.4 Unique Reasoning Conclusion ................................................................. 196
9.5 Differences in Reasoning Prevalence ........................................................... 197
  9.5.1 Informative Differences - Eliminate ..................................................... 198
  9.5.2 Expected Difference Example - Eliminate ........................................... 201
  9.5.3 Absence of Differences - Eliminate ..................................................... 202
  9.5.4 Informative Differences - Consider ..................................................... 203
  9.5.5 Expected Differences Example - Consider .......................................... 205
  9.5.6 Absence of Differences - Consider ..................................................... 206
  9.5.7 Informative Differences - Given Correct ............................................ 206
  9.5.8 Expected Differences - Given Correct ................................................ 208
  9.5.9 Absence of Differences - Given Correct ............................................ 208
  9.5.10 Differences in Reasoning Conclusion .............................................. 208
9.6 Zero as a Response Option ........................................................................ 211
9.7 Compensation-like Reasoning ................................................................. 215
9.8 Reasoning Based on the Provided Outcome .............................................. 217
9.9 Unique Reasoning to Justify Why Something is Incorrect ...................... 219
9.10 Conclusion ................................................................................................. 222
  9.10.1 Summary of Results from Eliminate .................................................. 223
  9.10.2 Summary of Results from Consider .................................................. 224
  9.10.3 Summary of Results from Given Correct ......................................... 225
  9.10.4 Summary of Results Spanning Variations ......................................... 226
10. CONCLUSION............................................................................................... 228

10.1 Findings on Alternate Ideas........................................................................ 228

10.1.1 Unique Reasoning .............................................................................. 230

10.1.2 Informative Differences ...................................................................... 231

10.2 Findings About Question Variations .......................................................... 234

10.3 Variation Utility, Applications, and Suggestions for Use .............................. 237

10.3.1 Consider Utility and Applications .......................................................... 237

10.3.2 Eliminate Utility and Applications .......................................................... 239

10.3.3 Suggestions for Using Eliminate ............................................................ 240

10.3.4 Given Correct Utility and Applications ................................................... 241

10.3.4.1 Suggestions for Using Given Correct ................................................. 244

10.4 Future Research ......................................................................................... 244

10.5 Summary ..................................................................................................... 250

REFERENCES ..................................................................................................... 252

APPENDIX A – ELECTRIC CIRCUITS HANDOUTS ......................................... 257

APPENDIX B – ACROSS-CONTENT STUDY PRETEST QUESTIONS ............ 262

APPENDIX C – RANDOM GROUPINGS IN CATALYST ................................. 300

APPENDIX D – GROUP SIMILARITY IN THE ACROSS-CONTENT

STUDY .............................................................................................................. 302

APPENDIX E – DEMOGRAPHICS OF PHYSICS COURSE .............................. 305

BIOGRAPHY OF THE AUTHOR ...................................................................... 307
LIST OF TABLES

Table 3.1  Examples of each question variation................................. 39
Table 3.2  Possible responses that can be targeted with each question
version. ................................................................. 40
Table 3.3  Number of choices offered for each version......................... 40
Table 5.1  Pretest designation, content area, and types of variation
administered for each pretest administration. ........................... 55
Table 5.2  Examples of consider, given correct, and given information
variations of a single question...................................... 59
Table 5.3  Examples of consider and eliminate variations that could be
made of a single question and the possible responses. ................. 61
Table 5.4  The statistical power of Fisher’s Exact Test. ......................... 64
Table 5.5  A contingency table for comparing students’ use of correct
reasoning on the original and given correct versions. ................. 69
Table 5.6  A table comparing increase responses on different question
variations. ............................................................. 70
Table 6.1  Results of Lindsey’s administration of the Work -by-Wall
question at the University of Washington (N = 339) [60]............. 79
Table 6.2  Results of Lindsey’s administration of the Work -by-Wall
question at the University of Washington (N = 339) [60]............. 79
Table 6.3  Layout of Pretest 9, which had three groups and two questions
with multiple versions. ............................................. 82
Table 6.4  Reasoning for all three versions of the question, with common color coding. ............................................................ 94

Table 6.5  Number and percentage of students using no displacement, or any other type of reasoning, in response to the original and given correct versions. ............................................................ 96

Table 6.6  Number and percentage of students selecting or eliminating each response on the original and eliminate versions respectively. ............................................................ 99

Table 6.7  Number and percentage of students giving the response zero or a different response to the original and eliminate versions. ...... 100

Table 6.8  Number and percentage of students giving the response negative or a different response to the original and eliminate versions. ............................................................. 101

Table 7.1  Table from Lawson and McDermott showing the percentage of students giving each particular response and reasoning before an intervention from the interviewer [23].............................. 107

Table 7.2  Layout of Pretest 9, which had three groups and two questions with multiple versions. ...................................................... 112

Table 7.3  Number and percentage of students selecting each response. ........... 112

Table 7.4  Number and percentage of students selecting correct or incorrect responses. ............................................................. 113

Table 7.5  Reasoning for all three versions of the question with common color coding. ............................................................ 123
Table 7.6  Number and percentage of students using equal applied force reasoning, or any other type of reasoning, in response to the original and given correct variations. ................................. 125

Table 7.7  Number and percentage of students using compensation reasoning, or any other type of reasoning, in response to the original and given correct variations. ................................. 126

Table 7.8  Number and percentage of students using compensation reasoning, or any other type of reasoning, when justifying the correct response equal to on the original and given correct variations. ............................................................... 128

Table 7.9  Reasoning for students providing the incorrect responses greater than, less than, or NOT equal to on the original and consider equal to versions. ............................................................... 130

Table 7.10 Key differences between the types of reasoning used to justify incorrect responses on the original and consider equal to versions. ................................................................. 131

Table 7.11  Number and percentage of students using unequal elements reasoning, or any other type of reasoning, in response to the original and consider equal to variations. ................................. 131

Table 7.12  Number and percentage of students using more velocity reasoning, or any other type of reasoning, in response to the original and consider equal to variations. ................................. 131

Table 8.1  The percentage of students who disagreed with the information presented as part of the given correct variation for all given correct variations. ................................................................. 152
Table 8.2  The percentage of students who choose an incorrect response on the *original* version and disagree with the provided outcome on the *given correct* variation, and the rate of disagreement, for each *given correct* variation. ....................... 158

Table 9.1  A summary of the results of all variations from the across-content study. ................................................... 183

Table 9.2  Pretest designation, question name, and type of variation for each question administration in the across-content study. .......... 184

Table 9.3  Number of students selecting, or eliminating, each response on the *original* and *eliminate* versions for all conflicted responses in the across-content study. ................................. 185

Table 9.4  Number and percentage of students giving the response *greater than*, on the *original* and *consider greater than* versions of the Oval Track question. ........................................... 187

Table 9.5  Number and percentage of students using *velocity must change* reasoning in response to the *original* and *eliminate* versions. .... 189

Table 9.6  Number and percentage of students using *non-compensation* reasoning in response to the *original* and *given correct* versions of the Change in Momentum question from Pretest 10. ............. 194

Table 9.7  Number and percentage of students using *otherwise* reasoning in response to the *original* and *eliminate* versions of the Car-Over-Hill question from Pretest 4................................. 199

Table 9.8  Number and percentage of students using *otherwise* reasoning in response to the *original* and *eliminate* versions of the Gymnast question from Pretest 6................................. 199
Table 9.9  Number and percentage of students using *velocity is zero* reasoning in response to the *original* and *consider zero* versions of the Simple Harmonic Oscillator question from Pretest 13................................................................. 204

Table 9.10  Number and percentage of students using *equilibrium* reasoning in response to the *original* and *consider zero* versions of the Simple Harmonic Oscillator question from Pretest 13................................................................. 205

Table 9.11  Features and results of *consider* and *given correct* variations with *zero* or *equal to* as a response option. ......................... 212

Table 9.12  A summary of the results of all variations from the across-content study. ................................................................. 223

Table D.1  Number of students selecting each response or some other response for each group of students on a base question. ............ 303
LIST OF FIGURES

Figure 2.1 The “five blocks” problem used to investigate student understanding of buoyancy. ................................. 8

Figure 2.2 A sketch of the apparatus used in interviews by Lawson and McDermott. ..................................................... 12

Figure 2.3 Questions Stetzer et al. used to investigate student difficulties with understanding the conservation of current [26]. .......... 14

Figure 2.4 The free-response and multiple-choice multiple-response formats of the question used by Wittmann et al. ................. 18

Figure 2.5 The speed-cuing version (A) and distance-cuing version (B) of the question administered by Frank et al. ................. 20

Figure 2.6 The different distributions of incorrect reasoning in response to each of the two versions of the questions used by Frank et al. ................................................................. 21

Figure 2.7 The two versions of the task administered by Sabella and Redish. .................................................................. 22

Figure 2.8 The results of Heckler’s task. ............................................ 23

Figure 4.1 The original version of the question [3]. ......................... 43

Figure 4.2 The full text of each question version and the associated distribution of students’ responses (or reasoning, in the case of the given correct variation). ........................................ 44

Figure 5.1 How the pretest uses the randomization task to break students into random groups. ........................................ 57
Figure 6.1 A question used by Lindsey to assess students’ understanding of work and systems [3, 60]................................. 77

Figure 6.2 The *given correct* variation of the question. ......................... 81

Figure 6.3 The *eliminate* variation of the question. .............................. 82

Figure 7.1 A diagram of the demonstration used by Lawson and McDermott [23]. ........................................................ 106

Figure 7.2 The *original* version of the Change in Kinetic Energy question [3]. ................................................................ 109

Figure 7.3 The *given correct* variation of the Change in Kinetic Energy question................................................................. 110

Figure 7.4 The *consider equal to* version of the Change in Kinetic Energy question............................................................. 111

Figure 8.1 The *given correct* variation of the Car-Over-Hill question administered as part of Pretest 4. ................................. 153

Figure 8.2 The *given correct* variation of the Change in Momentum question from Pretest 10. ............................................ 154

Figure 8.3 The *given correct* variation of the Rotating Disk/Wheel question from Pretest 12. .................................................. 155

Figure 8.4 A plot of the percentage of students who disagree with the provided outcome on the *given correct* variation versus the percentage of students who answer the *original* version of the question incorrectly. .................................................. 157
Figure 8.5  A plot of the rate of disagreement on the *given correct*
variation versus the percentage of students who answer the
*original* version of the question incorrectly............................ 159

Figure 8.6  The *original* version of the Gymnast question [54].............. 164

Figure 8.7  The *given correct* version of the Gymnast question. ............ 164

Figure A.1  The *original* version of the Electric Circuits question [3]. ....... 258

Figure A.2  The *consider increase* variation of the Electric Circuits
question..................................................................................... 259

Figure A.3  The *eliminate* variation of the Electric Circuits question. ....... 260

Figure A.4  The *given correct* variation of the Electric Circuits question. ...... 261

Figure B.1  The *original* version of the Oval Track question from
Pretest 2 [3]. .................................................................. 262

Figure B.2  The *consider greater than* variation of the Oval Track question
from Pretest 2................................................................. 263

Figure B.3  The *consider less than* variation of the Oval Track question
from Pretest 2........................................................................ 263

Figure B.4  The *consider equal to* variation of the Oval Track question
from Pretest 2........................................................................ 264

Figure B.5  The *original* version of the Pendulum question from
Pretest 3 [89]. ................................................................. 265

Figure B.6  The *eliminate* variation of the Pendulum question from
Pretest 3................................................................. 266
Figure B.7  The consider upward variation of the Pendulum question from Pretest 3................................................................. 267
Figure B.8  The given correct variation of the Pendulum question from Pretest 3................................................................. 267
Figure B.9  The original version of the Car-Over-Hill question from Pretest 4 [54]. ........................................................... 268
Figure B.10 The eliminate variation of the Car-Over-Hill question from Pretest 4................................................................. 269
Figure B.11 The consider greater than variation of the Car-Over-Hill question from Pretest 4................................................................. 269
Figure B.12 The given correct variation of the Car-Over-Hill question from Pretest 4................................................................. 270
Figure B.13 The original version of the Newton’s 3rd Law question from Pretest 5................................................................. 271
Figure B.14 The eliminate variation of the Newton’s 3rd Law question from Pretest 5................................................................. 272
Figure B.15 The consider greater than variation of the Newton’s 3rd Law question from Pretest 5................................................................. 273
Figure B.16 The original version of the Gymnast question from Pretest 6 [54]. ........................................................... 274
Figure B.17  The eliminate variation of the Gymnast question from Pretest 6................................................................. 275
Figure B.18  The given correct variation of the Gymnast question from Pretest 6................................................................. 276
Figure B.19  The original version of the 1-D Conservation of Momentum A question from Pretest 7a [3]................................................. 277

Figure B.20  The eliminate variation of the 1-D Conservation of Momentum A question from Pretest 7a............................................. 278

Figure B.21  The consider greater than variation of the 1-D Conservation of Momentum A question from Pretest 7a. ............................... 279

Figure B.22  The original version of the 1-D Conservation of Momentum B question from Pretest 7b [3]............................................ 280

Figure B.23  The consider less than variation of the 1-D Conservation of Momentum B question from Pretest 7b............................... 281

Figure B.24  The original version of the 2-D Collision question from Pretest 8 [3]. ............................................................ 282

Figure B.25  The given correct variation of the 2-D Collision question from Pretest 8................................................................. 282

Figure B.26  The original version of the Change in Kinetic Energy question from Pretest 9 [3]. ............................................................. 283

Figure B.27  The consider equal to variation of the Change in Kinetic Energy question from Pretest 9................................................... 284

Figure B.28  The given correct variation of the Change in Kinetic Energy question from Pretest 9....................................................... 285

Figure B.29  The original version of the Work-by-Wall question from Pretest 9b [60]................................................................. 286

Figure B.30  The eliminate variation of the Work-by-Wall question from Pretest 9b. ................................................................. 287
Figure B.31 The *given correct* variation of the Work-by-Wall question
from Pretest 9b. ........................................................ 288

Figure B.32 The *original* version of the Change in Momentum question
from Pretest 10 [3]. ..................................................... 289

Figure B.33 The *consider equal to* variation of the Change in Momentum question from Pretest 10. ............................................... 290

Figure B.34 The *given correct* variation of the Change in Momentum question from Pretest 10. .............................................. 291

Figure B.35 The *original* version of the Pivoting Rod question from Pretest 11 [3]. ........................................................... 292

Figure B.36 The *consider equal to* variation of the Pivoting Rod question from Pretest 11.......................................................... 293

Figure B.37 The *given correct* variation of the Pivoting Rod question from Pretest 11................................................................ 293

Figure B.38 The *original* version of the Rotating Disk/Wheel question from Pretest 12 [63]..................................................... 294

Figure B.39 The *consider greater than* variation of the Rotating Disk/Wheel question from Pretest 12. ................................. 295

Figure B.40 The *given correct* variation of the Rotating Disk/Wheel question from Pretest 12. ................................................................. 296

Figure B.41 The *original* version of the Simple Harmonic Oscillator question from Pretest 13 [3]......................................................... 297

Figure B.42 The *consider zero* variation of the Simple Harmonic Oscillator question from Pretest 13. ................................................................. 298
Figure B.43  The *given correct* variation of the Simple Harmonic Oscillator question from Pretest 13. .............................................. 299

Figure C.1  The randomization question used to assign the students to random groups. ................................................................. 301
CHAPTER 1
INTRODUCTION

1.1 Motivation for Investigating Students’ Ideas and Thinking in Physics

Two goals of science education are to train a technically skilled workforce and create a scientifically literate society. Currently in the United States, these goals are not being met, and improvements to science education are needed to help achieve these goals [1]. One way to improve science education is through discipline-based education research, such as physics education research (PER) [2]. This dissertation focuses on improving physics and science education by enhancing research methods and investigating students’ ideas and thinking in an introductory physics course.

Physics education research enhances science education by improving how we teach students physics. To achieve this goal, many PER-related research studies have focused on improving students’ understanding of physics by studying the ideas students bring into the classroom and then using the results to develop improved curricular materials or instructional methods. Knowing what ideas students have prior to instruction, and how to address them, is a critical part of improving science education. Therefore, many PER-related studies focus on investigating students’ understanding of particular physics content areas and improving instruction to help students develop a correct understanding.

Many PER-based curricular or instructional reforms have been based on studies investigating what students know about particular physics content areas and how students think about physics. Examples of studies and curricular reforms investigating students’ ideas are those that improve curriculum by addressing students’ difficulties in particular content areas [3, 4, 5, 6, 7, 8] whereas examples investigating students’ thinking are those that focus on developing students’
meta-cognitive abilities [9, 10]. Other examples utilize information about students’ thinking and ideas together by organizing the order of curricular materials [11, 12] or improving teacher training by utilizing information about students’ ideas and thinking to improve teachers’ pedagogical content knowledge [13].

1.2 Improving Methods of Investigation

To further physics education, the work presented herein was completed with the goal of improving our ability to learn more about students’ ideas in all physics content areas while simultaneously investigating students’ thinking.

Given that an understanding of students’ ideas and thinking is so commonly used for improving physics courses, we feel that additional work to improve how we research students’ ideas and thinking in physics will further enhance physics instruction. Therefore, we present a method to conduct a more extensive investigation of students’ ideas in any content area, thereby providing insights into how students think about physics and develop responses to physics questions.

One common method of PER investigates students’ ideas by asking them what they think happens in carefully chosen physical situations and asking them to explain the reasoning they used to come to that conclusion. These investigations also commonly utilize many questions about different physical situations in order to uncover the ideas students commonly apply and enable researchers to thoroughly understand those ideas. Although such investigations are able to determine the ideas students commonly use to answer physics questions in particular content areas, they do not provide an in-depth picture of the range of ideas a student may choose to use to assess any particular physical situation. Previous research has shown that students may use a variety of ideas to respond to a physics question depending on how the question is asked [14, 15, 16]. Therefore, there is value in investigating the range of ideas students may apply to a single physical situation so
that we can learn more about many of students’ ideas rather than only those commonly applied or expressed in response to a particular question’s phrasing.

The method presented herein involves changing the task students are to complete when analyzing the physics of a particular physical situation. Here, “physical situation” means the objects and surroundings students are asked to consider when answering a question, and as well as their attributes. Instead of just asking students what happens in a particular situation, an array of questions are used to learn more about students’ ideas. By asking students to address a particular correct or incorrect outcome of the physical situation, eliminate an incorrect outcome, or justify a correct outcome provided to them, we can learn more about the physics ideas students have. Applying this methodology across multiple content areas highlights how prevalent alternate ideas are and enables us to investigate whether patterns exist in how students respond to these different question variations across multiple content areas.

Expanding the understanding of students’ ideas and thinking is expected to lead to improved instruction. Learning more about the question variations used may also allow instructors to apply them as tools in the classroom or researchers to use them as tools to improve the elicitation of students’ ideas about physics content.

1.3 Dissertation Overview

In this dissertation, our goal is to demonstrate the effectiveness of our question-variation-based research methodology by applying it to learn about students’ alternate ideas and how students respond to question variations.

Through asking question variations applied consistently across a wide range of physics content areas we can perform a broader study of the alternate ideas students possess, and how common they are; demonstrating the effectiveness of our methodology. Additionally, we can look at students responses to particular question
variations in many content areas to learn how students respond differently to these variations.

The set of research questions about alternate ideas that we hope to address are:

- How prevalent are alternate ideas across many physics content areas?
- What alternate ideas do we observe in each content area?

The set of research questions about our question variations that we hope to address are:

- Do these question variations elicit alternate ideas? Is our methodology able to consistently identify alternate ideas, or do they need to be targeted with content-specific question modifications?
- How do students engage with different variations of qualitative multiple-choice physics questions?
- Are the types of alternate ideas we find influenced by the variations we ask?
- Do certain content areas or question features more commonly elicit alternate ideas in combination with these question variations?

In order to provide answers to these questions this dissertation will describe the relevant literature, present the development of question variations, provide examples of applications of the variations and the types of reasoning seen in response, and discuss patterns observed in responses to the question variations.

Chapter 2 presents prior research and frameworks that show how curriculum is developed based on students’ ideas and how those ideas can vary. We discuss a common PER cycle of research and curriculum development along with the

\(^1\)We will present a few more specific research questions related to our variations after the presentation of the exact question variations we choose to use in this dissertation, presented in Section 3.3.
framework it uses. A discussion of prior research on question representation and format is followed by information on studies demonstrating that students’ ideas can vary and depend on the details of the questions to which they respond. Finally, we present a framework that accounts for the variability of students’ ideas and discuss the terms and framework we will use in this dissertation.

Chapter 3 presents the methodology developed based on the literature reviewed in Chapter 2, providing a more extensive investigation of the range of ideas students may use when responding to questions about a single physical situation. The methodology developed consists of four question variations that constrain the responses students are allowed to address.

Chapter 4 discussed the pilot study used to test our methodology, including the adaptation of the methodology to this particular implementation, and the results of the study. Although the results of the pilot study are somewhat limited by its smaller sample size, this chapter serves as an example of the types of studies that make up the rest of this dissertation. Despite the small sample size, the evidence suggests that we can learn more about students’ ideas by asking multiple questions about the same physical situation.

Chapter 5 details the design, administration, and analysis methods of a much larger study designed to test our methodology in a wide range of content areas. This study, called the across-content study, administered question variations to the same classroom of students on a weekly basis via an online pretest system, thereby enabling us to investigate the prevalence of alternate ideas in a wide variety of content areas. We also discuss methods for creating variations of existing questions based on our methodology.

Chapters 6-9 present results from the across-content study. Chapters 6 and 7 focus on two administrations of question variations from the across-content study that utilize our methodology to investigate students’
understanding of work. We review the literature on students’ understanding of the specific content area discussed in each chapter, the application of our methodology to that specific content area, examples of the analysis methods used, and the results of each investigation. Significant differences emerge in how students answer the different question variations, and these differences tell us more about students’ ideas in both content areas.

Chapter 8 includes a comparison of the types of responses we received when asking students to justify a provided correct outcome of a physical situation to Chinn and Brewer’s [17] framework, which categorizes students’ responses to anomalous data. We make this comparison because providing students with a correct outcome to a physical situation is one way of providing students with anomalous data; in our study, the students’ reactions to the provided correct outcome align well with Chinn and Brewer’s framework, with our students rejecting the provided information or reinterpreting the physical situation.

Chapter 9 provides the results of applying our methodology to many different content areas. The results are presented in less detail than in Chapters 6 and 7, but show a broader picture of how the methodology performs when applied to a wide range of content areas. We discuss the information each type of question used in our methodology allows us to find and how often those findings were observed. Each question variation was able to access alternate ideas, but with different levels of success; several results spanned variations.

Finally, Chapter 10 provides a discussion of the overall success of the methodology, potential uses of question variations, and suggestions for applying the methodology before proposing future research projects.
CHAPTER 2
MOTIVATIONAL LITERATURE REVIEW

As discussed in the introduction, understanding students’ ideas and how they think about physics is an important step in being able to improve physics instruction. This chapter presents existing research and frameworks that guide the work presented herein. We first discuss a common cycle of research and curriculum development in physics education, including the framework it uses and types of results it observes. We then examine previous research utilizing question variations in PER, discuss how these studies demonstrate the variability of students’ ideas, and present a framework that accounts for the observed variability. Finally, we explore the level of compatibility between the two frameworks discussed in this chapter and describe how the research in this dissertation will employ both frameworks to utilize the variability of students’ ideas to improve the research and curriculum development cycle.

2.1 The Research and Development Cycle

Many research-based curricula and teaching practices use an understanding of students’ ideas to improve instruction; although they may use different theoretical frameworks or have differing pedagogical beliefs, they all use some form of a cycle that starts with investigating students’ understanding of physics followed by making changes to instructional practices based on the results of their investigations [3, 9, 11, 18, 19, 20, 21].

In this dissertation, we focus on investigating students’ understanding by using question variations to more thoroughly study students’ ideas and learn how students’ responses are affected by questions variations. In this way, we hope to provide tools and results for other researchers to use when working to improve instruction through research.
In this section, we walk through one cycle of research and development to demonstrate how research into students’ ideas can enhance their understanding. The discussion uses Heron’s [22] example from her paper describing the research and curriculum development cycle used by the Physics Education Group at the University of Washington.

As Heron explains, the initial step of their process is to identify an area or situation in which students are not able to provide satisfactory explanations of the physics. In this case, the group uses the students’ inability to explain the buoyancy of a “Cartesian diver” as an example, starting the research from there.

After identifying a target topic or situation, the group investigates students’ ideas in this content area to determine what difficulties they encounter when trying to analyze this situation. To this end, they developed a research task (question) about buoyancy that was similar to the situation with the Cartesian diver and asked students to develop and explain their responses to this task (see Figure 2.1).

In administering this task, the researchers were able to identify several specific difficulties students had, including a difficulty determining the magnitude of the buoyant force. Students incorrectly related the magnitude of the buoyant force to the mass or position of the block instead of its volume. Although multiple
difficulties were identified in this example, we follow only the impact of the buoyant force difficulty.

Once this difficulty was identified, it was further probed by designing additional research tasks to more thoroughly understand the ideas students were using to answer these questions.

Once the difficulties were well understood the curriculum development process began. A tutorial—namely, a worksheet to be completed while working in a small group—was developed to address the difficulties students showed when answering the research tasks. To address the difficulty in determining the correct buoyant force, the tutorial was designed to guide students “from the assumption of linear variation of pressure with depth to the conclusion that fluid exerts a net upward force on a submerged object that does not depend on the object’s depth or mass” [22, p. 356]. It did this by having students work through several free-body diagrams of rising, sinking, and stationary blocks of different masses submerged in liquid, paying special attention to the forces acting on each surface of the blocks. This activity was specifically designed to address students’ efforts in incorrectly relating the magnitude of the buoyant force to mass or position, which was observed in the research.

The effect of the tutorial was then measured and found to have led to improved student understanding. The researchers noted that many students continued to relate the magnitude of the buoyant force to the mass of the object even after completing the tutorial; thus, they investigated this difficulty further and implemented changes to the curriculum through repetitions of the same cycle of research and curriculum development that led to the creation of the tutorial.

Through this process to understand students’ ideas and modify instruction to help students develop a correct understanding of the content, physics education can be improved. Although this is one example of one specific cycle of research and
development, many other research-based physics curricula use different techniques and frameworks to achieve the same purpose. Thus, this example should not be viewed as the only way to enact this cycle [3, 9, 11, 18, 19, 20, 21].

We adopted the general view that knowing more about students’ ideas better enables us to modify physics instruction to help students develop a correct understanding. We took this view because of the wide range of physics education research studies that have led to improvements in instruction and the wide range of instructional changes that have led to improved student understanding [2]. This research cycle is one example of using students’ ideas to improve instruction; the research methodology presented in this dissertation could be useful in any similar cycle of research and instructional improvement.

2.2 The Specific Difficulties Framework for Investigating Student Understanding

Heron’s research and curriculum development cycle discussed in the previous section uses a framework for investigating and improving student understanding of physics referred to as the specific difficulties framework. In this dissertation we use many research techniques and modified research materials from studies using the specific difficulties framework. This section presents the relevant aspects of the framework and some of the results found using this framework. Because this dissertation focuses on learning about students’ ideas, not curriculum development, we focus only on the research portion of the specific difficulties framework.

The specific difficulties framework has been used to investigate students’ ideas in many physics content areas [8, 23, 24, 25], with the goal being to learn about students’ ideas and reasoning in a way that can lead to improved instruction and student learning [4, 5, 6].
The specific difficulties framework is a cognitive framework that assumes an individual student has ideas related to a physical situation or content area as well as reasoning abilities that reside in the individual and are influenced by their experiences and the instruction received. In answering physics questions, students often provide responses inconsistent with accepted physics understanding. It is therefore assumed, that in these cases, students have some difficulty with invoking or applying a correct understanding of the physics. This framework is not particularly concerned with understanding what causes students to have these difficulties, but instead focuses on understanding and addressing them. Because these difficulties can change with instruction or experience, appear in some students and not others, and can be given as a response to only a particular set of questions, these difficulties are referred to as “specific difficulties.” Here, the term “specific” signifies that the difficulties may be localized to a specific student, content area, or question.

Although this framework allows for students’ ideas or reasoning to be localized, many specific difficulties are expressed in a global manner. Because these difficulties, apply in multiple contexts, are used by many students, and are resilient to instruction, the framework is often used to investigate them [5].

2.3 Global Difficulties

The specific difficulties framework has been used to investigate students’ ideas in many content areas, and these studies almost always focus on investigating specific difficulties that students express globally. The most common way students’ ideas are seen to be global is when students express the same idea in response to many different content areas or many different physical situations within a content area. Therefore, this section discusses investigations that used a specific difficulties framework to investigate students’ understanding of several physics content areas.
and found that students exhibited the same specific difficulties in multiple content areas and in response to multiple physical situations.

### 2.3.1 Difficulties Across Content Area

One example of a specific difficulty seen in multiple content areas comes from Lawson and McDermott’s [23] investigation of students’ understanding of the work-energy and impulse-momentum theorems.

In their investigation, Lawson and McDermott showed individual students a demonstration where two pucks of unequal mass were pushed for the same distance, with the same force, on a frictionless surface. Figure 2.2 presents a diagram of the apparatus used in this demonstration. After seeing the demonstration, students were asked questions about the physical situation they had observed, such as comparing the final kinetic energy and momenta of the two pucks after they had traveled the distance for which the force was exerted. Students’ responses were categorized into groups of similar responses, from which the authors identified various specific difficulties.

When asked to compare the kinetic energy and the momentum, students often expressed the same specific difficulty. The researchers referred to this specific difficulty as “compensation reasoning.” When using compensation reasoning, students would explain that the greater mass of one would be compensated for by
its slower velocity (or vice versa); therefore, the two would have equal kinetic energies or momenta. This reasoning is insufficient to determine if the energies and momenta are equal and, thus, is considered incorrect.

The results of this study show that some specific difficulties can be expressed across content areas such as work and momentum.

2.3.2 Difficulties Across Physical Situation

In addition to specific difficulties being expressed across content areas, other studies find the same, or similar, specific difficulties being expressed in response to questions about various physical situations. For example, Stetzer et al. [26] found similar difficulties across several questions about electric circuits. They investigated students’ understanding of the conservation of current in an electric circuit and showed that difficulties with this idea persisted across multiple questions and physical situations as well as across introductory physics students, undergraduate physics majors, and graduate students. To perform this investigation, Stetzer et al. administered several different questions about electric circuits (see Figure 2.3) to students with various levels of expertise in physics. These findings indicate that the difficulty in applying the correct understanding of the conservation of current is a global difficulty that occur in student responses to questions about many physical situations and among introductory, upper-level, and graduate physics students.

2.4 Previous Research on Question Variations

The research in this dissertation uses question variations to investigate students’ understanding of physics content; thus, this section reviews prior research in physics that studied the effects of changing question representations and formats. These studies show that altering the questions researchers use to assess students’
Figure 2.3. Questions Stetzer et al. used to investigate student difficulties with understanding the conservation of current [26]. Many students had difficulty demonstrating a correct understanding of conservation of current in response to each of these questions.

understanding may change the results while asking multiple question variations can provide a broader picture of students’ understanding.

2.4.1 Question Representation

Prior research has extensive studied students’ understanding of representations and the benefits of teaching students using multiple representations. Here, we instead focus on studies that compare the solution methods students use when responding to different representations.¹

In mathematics education, Koedinger and Nathan [28] compared students’ solution methods between word-based representations and formula-based representations. They found that the solution method students used was linked to

¹For a review of research on representations in general, see De Cock [27].
the representation they were given. In addition, Lem et al. [29] studied student responses to varied representations in statistics and found that, when students were shown a histogram, box plot, or descriptive statistics, they had varying levels of success. The researchers concluded that this result was due to students’ varying aptitudes in using each representation.

In physics education, Meltzer [30] studied the solution methods of introductory physics students when presented with different representations. This study varied the representation among verbal, mathematic, graphical, and diagrammatic representations. The results showed that students sometimes solved a problem more successfully when presented with a particular representation, and is backed up by a recent study by Susac et al.[31] which also showed students have varying levels of success with different representations. Meltzer also noted that some students had difficulty with vector representations. Hawkins et al.’s [32, 33] and Van Deventer’s [34] follow-up studies comparing students’ abilities to perform graphical vector addition and subtraction found that students were often able to complete these vector operations more successfully in a mathematics context.

Similar to Meltzer’s study, Kohl and Finkelstein [35] compared students’ performance on physics tasks using the same set of varied representations. They again found that representation can impact how successful students are at completing the tasks. More specifically, students were often more successful when shown a mathematical representation. This result is consistent with the work by Hawkins et al. and Van Deventer, and Kohl and Finkelstein proposed that this result is due to students’ familiarity with “plug-and-chug” solution methods, which tend to rely heavily on mathematical formalism.

Additional research comparing math and physics representations by Thompson and others commonly shows that students have difficulty successfully transferring their mathematics knowledge to a physics context [36, 37, 38, 39, 40].
In summary, students have varying levels of success with differing
representations, and varying the representation they are shown in a question can
influence their ability to answer correctly or affect their approach to solving the
problem.

Therefore, if a question is asked in only one representation, it may be unclear
how well students are able to understand and work with the same content in
different representations. This is a clear demonstration that features of the
questions students are being asked can affect their responses, even if the questions
are highly similar in other regards.

2.4.2 Question Format

Another common type of question variation in physics is a comparison of
multiple-choice and free-response question formats.

This topic has been widely studied in science education. In a review of studies
making these comparisons, Rodriguez [41] concluded that, as long as both versions
of the question have similar stems and the questions are carefully designed to
“measure the same aspects of the specified content and cognitive domains” the
scores from the two measures will be very similar.

Similar studies within physics education research confirm this broader trend
from science education, but these studies also caution that there are still instances
where the question format can influence students’ responses [42, 43].

The two studies discussed here, by Steinberg and Sabella [43] and Wittmann
et al. [42], both focused on an introductory physics class at the University of
Maryland, and involved asking the same students both question formats and then
comparing the responses.

In the study designed to compare multiple-choice questions from the Force
Concept Inventory (FCI) [44] to free-response equivalents, Steinberg and
Sabella [43] found student performance to be correlated between the two questions, but there were still instances where the question format affected students’ responses.

The study consisted of administering the FCI to the entire course, and then administering several free-response versions of FCI questions on a final exam the students took a week later.

Regarding items where students’ responses varied across question formats, the researchers concluded that students may be inconsistent in their thinking about physics due to having multiple mental models, and that the different question formats could elicit different responses. One possible reason for the different responses is that the multiple-choice response options trigger different mental models.

In another study investigating students’ understanding of waves, Wittmann et al. [42] found that students provided different responses to free-response and multiple-choice multiple-response versions of a question. In this study, students were asked about the factors affecting the speed of a pulse propagating down a string after being created by the motion of a hand (see Figure 2.4). The researchers found that 60% of students responded to the free-response version by only stating that the hand motion affected the speed of the pulse, but responded to the multiple-choice multiple-response version by stating that both the hand motion and the medium affect the speed of the pulse.

To explain this result Wittmann et al. made the same claims as Steinberg and Sabella: students have multiple mental models, and the free-response and multiple-choice multiple-response questions trigger different or multiple models. They also noted that one type of question alone would not have given such a complete picture of students’ understanding, thus, using multiple question formats may be essential when students have many—possibly conflicting—ideas about a situation.
Figure 2.4. The free-response and multiple-choice multiple-response formats of the question used by Wittmann et al.. They found that roughly 60% of students did not answer these questions consistently [42].

The notion that students’ ideas can be more fully assessed by using multiple question formats to investigate their ideas is a fundamental aspect of this dissertation, and the basis for our methodology of using question variations to more broadly investigate students’ ideas across all content areas.

The next section examines studies that followed up on investigating how students’ responses are influenced by changes to questions and continued to show that students’ ideas are often variable and influenced by small changes to research tasks.

2.5 Variability of Students’ Ideas

Although the majority of studies using the specific difficulties framework to investigate students’ ideas have focused on difficulties that appear globally, some recent studies have shown that students’ ideas can be localized, even within a single content area.
Building on studies investigating question format, the studies presented here all made minor modifications to the tasks presented to students and found that those small changes led to different responses from students. Here we present a summary of these recent results and describe how this dissertation will use variability as a tool for researching students’ ideas and thinking; we then present a theoretical framework that accounts for variability in students’ ideas and discussion of how it explains these results.

2.5.1 Variability Due to Visual Cuing

Frank et al. investigated how students’ responses to a kinematics question could be influenced by visual cues [14]. In their experiment, they asked students questions that were similar, but emphasized different information presented in the problem via visual cues. The kinematics question they modified is about three experiments in which identical balls are rolled off of a table at three different velocities. Each student in a large lecture class received one of two versions of this question emphasizing either the distance the balls travel or the velocity with which they leave the table (see Figure 2.5).

As shown in Figure 2.6, Frank et al. found that students responding to the two different versions answered correctly with equal prevalence, but provided different distributions of incorrect responses. Students responding incorrectly to the speed-cuing version concluded that, the farther the ball travels, the more time it will take to hit the ground, as often as they responded that the farther the ball travels, the less time it will take to hit the ground. However, students responding incorrectly to the distance-cuing version of the task were more likely to respond that the farther the ball travels, the longer it will take to hit the ground. Thus, students responding to the speed-cuing version of the question were relatively more likely to say the balls take less time to hit the ground when they are going faster, and students receiving
The distance cuing version are relatively more likely to say the balls take more time to hit the ground when they are traveling farther. In other words, students’ incorrect responses can be influenced by the version of the question they received.

This result shows students’ responses to questions about the same physical situation can be influenced by emphasizing different features of the physical situation, even when there is no actual difference in the physical situation the students are assessing.

2.5.2 Variability Due to the Content of Preceding Questions

Sabella and Redish investigated how asking students questions about forces would influence their ability to use reasoning about work and energy on subsequent questions. They administered two different versions of a survey (i.e., the “long version” and “short version”) to students in an introductory physics course. Both versions were identical, except the long version had two preceding questions that asked students about force and acceleration (see Figure 2.7). To determine how the additional questions about force and acceleration influenced students’ ability to use
Figure 2.6. The different distributions of incorrect reasoning in response to each of the two versions of the questions used by Frank et al.. The questions are shown in Figure 2.5. Roughly the same percentage of students gave incorrect responses on the two versions [14].

work and energy ideas, the researchers looked for evidence of students using work- or energy-related reasoning on the last question of either version (which is the same on both versions).

Sabella and Redish found that, when students were asked the additional questions about forces, it inhibited their ability to access useful reasoning about energy on subsequent questions [15]. Students who did not receive additional preceding questions about force were better able to use work and energy ideas appropriate to the problem they were solving and then use those ideas to achieve a correct solution.

This provides a clear example of how students’ responses to physics questions can be influenced by the content they are asked to consider in preceding questions, even when the question to which they are responding has not been modified in any way.
Figure 2.7. The two versions of the task administered by Sabella and Redish. The long version contains all four parts of the problem. The short version contains only the two parts in bold. An example solution is also shown in red [15].

### 2.5.3 Variability Due to Suggested Solution Method

Heckler [16] demonstrated that requiring students to draw a free-body diagram (FBD) when solving force and motion questions could cause students to answer incorrectly more often when not required to do so. In the first experiment of the study, groups of students from the same courses received a question about force and motion requiring them to find a numerical value for either the minimum mass for a
Figure 2.8. Heckler’s results. These show that students prompted to draw a free-body diagram are less likely to answer these two questions correctly than those not prompted to draw a free-body diagram [16].

box to stay stationary on a surface with friction (Problem 1) or the magnitude of a force (Problem 2). These students were placed into one of two conditions: (1) asked to draw a FBD before solving the problem or (2) no such prompt received. Students asked to draw a FBD prior to solving the problem were less likely to answer the question correctly (see Figure 2.8).

Similar to Sabella and Redish’s results, this example of how the ideas students use to solve physics problems can vary based on small changes to the task they should complete, even when the physical situation considered is exactly the same.

2.5.3.1 Variability Due to Accessibility

In another study, Heckler and Bogdan [45] showed that the accessibility of explanatory factors students are presented with when analyzing a physical situation influence what other explanatory factors they bring up.
In cognitive science literature, ideas that are accessible are those that are invoked by the subject, or “what comes to mind;” and ideas that are available are those that are not invoked by the subject, but which the subject views as relevant when the ideas is brought up [46].

As an example Quinn and Markovits [47] present the physical situation of “someone’s pupils are dilated” to which almost all subjects in their study offered the explanatory factor “light” demonstrating this explanatory factor is highly accessible. The explanatory factor “drugs” was offered by one-fifth of the subjects, indicating that it is less accessible. However; when asked to verify the statement "If someone takes a psychototropic drug, then the pupils of the eyes will dilate" more than one-fifth of subjects verified it as correct, showing that the explanatory factor “drugs” is available to more of the population than to which it is accessible.

In the study by Heckler and Bogdan, they showed that the accessibility and availability of explanatory factors could influence students’ responses to conceptual physics questions by demonstrating that offering highly accessible explanations leads to less frequent consideration of alternate explanations.

To demonstrate this, they first measured the accessibility of explanatory factors for six different physical situations by asking a free-response question. For example, they described two pendulums with different periods and asked students to list the reasons the pendula would have different periods, finding that the two most accessible explanatory factors were length (67%) and mass (48%). They considered all explanatory factors even if incorrect.

Then they determined the availability of the most accessible explanatory factors by directly asking students if explanatory factors were relevant or not. For example asking if the pendulum with the longer period has a length that is longer, shorter, or doesn’t matter. If students responded longer or shorter length was considered an
accessible explanatory factor for them, if the responded *doesn’t matter* length was considered to not be an accessible factor.

Finally, they proposed and tested what they refer to as an “accessibility rule” which states that presenting students with a highly accessible explanatory factor will result is students bringing up alternate explanatory factors less often than when presenting students with explanatory factors that are less accessible.

They tested this by describing the physical scenario and asking students if a single explanatory factor was a valid conclusion — such as by asking if the statement “Pendulum A has more mass than Pendulum B” is a valid statement, or if students indicated that another factor was relevant.

They found that their accessibility rule held in most cases, and that offering students a explanation that was highly accessible lead to students offering up alternative explanations less often than when less accessible factors were offered.

This research demonstrates that the ideas students express can be influenced by the accessibility of explanatory factors presented to them. Specifically, if a highly accessible explanatory factor is presented to them, they are less likely to seek out other explanatory factors. This again reinforces that students’ ideas can be variable, and what is *accessible* to students can change based on the question students are asked.

### 2.5.4 Results of Variability Studies

All of the studies discussed in this section, and the work on question representation and format, show that students’ responses to physics questions can be influenced by features of the questions themselves. Students express different ideas in response to these question variations, despite all of them asking about the same physical situation. These results have important consequences for how we
assess and teach students, and more research is needed to understand students’ ideas and how they are influenced by question features.

The different reasoning seen in all of these examples, as well as Wittmann et al.’s and Steinberg and Sabella’s work, emerges as a difference in the prevalence of a particular kind of response between students responding to two different versions of a research task. Sometimes these are the same students, but often they are groups of students assumed to be highly similar, but who only respond to one version of the research task each. Both methods are used to conclude that the difference in prevalence of particular types of reasoning shows that some individual students will respond to questions differently due to the variations in the questions being asked. The research presented here uses a similar methodology and identifies similar types of evidence.

What we find interesting about these results is that these different ideas can be elicited by questions asking about exactly the same physical situation, and same physics content. Without changing the physical situation or the aspects students are being asked about, we would like students to give consistently correct responses even with varied question formats or preceding tasks. However, these studies show that students may not have strong internally consistent ideas for analyzing these physical situations and may have multiple ideas that they can apply to the exact same physical situation.

Because some students may have multiple ideas, one question format is likely insufficient for understanding all the ideas students relate to that physical situation. As previously discussed, knowing what ideas students use to assess and make sense of a physical situation is important for improving instruction; thus, multiple question formats should be able to help us learn about all of students’ ideas, and enable the development of better instructional materials and practices. Therefore, in this dissertation, we develop a technique to easily create varied question formats
as asking about the same physical situation in order to investigate students’ ideas more broadly.

### 2.5.5 Resources Framework

Many studies on the variability of students’ ideas use the *resources* framework to guide their research [48]. Hammer [48] developed this framework based largely on the work of the *Knowledge-in-Pieces* framework developed by diSessa [48, 49]. The *resources* framework focuses on explaining students’ mental process in developing reasoning about physical situations. This explanatory framework can account for the results of studies that demonstrate the variability of students’ ideas.

The *resources* framework models students’ ideas as consisting of fine-grained pieces of knowledge that are assembled into structures as needed. These structures of connected knowledge pieces form students’ reasoning about a physical situation. Hammer calls these pieces of knowledge “resources,” because they are the resources students use to generate their reasoning about a physical situation. For example, when considering the physical situation of a book resting on a table, some commonly applied resources are *blocking*, which can be used to reason that the book does not fall through the table because the table is in the way, or *springiness*, which can be used to reason that the table acts like a spring and exerts an upward force on the book.

Because resources are fine-grained pieces of knowledge, they are not inherently correct or incorrect, but can lead to correct or incorrect reasoning depending on their application. For example, when attempting to explain why the earth is warmer in the summer, students often respond that the earth is closer to the sun in the summer. This can be viewed as students applying the *closer is stronger* resource to arrive at a response. In this application, the response they develop is not correct; however, this same resource can be correctly used to explain why music is louder.
when closer to a speaker, or why a light from a bulb is more intense at closer distances.

This dissertation uses Hammer’s term “resources” to describe the pieces of knowledge that make up students’ reasoning, but expands it to include non-intuitive elements of knowledge, consistent with Frank et al. [14, 48]. For example, Frank et al. refer to going faster implies taking less time, going farther takes more time, and going faster implies getting farther as resources students are using to develop reasoning [14].

One key aspect of the resources framework is that students’ thinking is generated on demand, which accounts for variability or a lack of self-consistency in students’ ideas. Although certain pieces of knowledge may assemble in a similar manner in many different situations, they are not required to; each assembly may be unique to the situation. As Frank et al. put it, “although they account for inconsistency, such frameworks do not rule out stability: common and stable patterns of thinking may emerge out of particular combinations of knowledge pieces that become reliably activated across many contexts” [14, p. 020102-3].

This dissertation refers to assemblies of knowledge pieces or resources as “knowledge structures” [15].

We use the resources framework because it can explain many of the results discussed in Section 2.5 related to the variability of students’ ideas. Frank et al. discussed their work as being consistent with the resources framework in that the different strong visual cues presented in the questions they asked could likely be causing the activation of different resources, thereby causing the creation of different knowledge structures that could lead to different responses from students [14].

Both Sabella and Redish, and Heckler discuss how the different responses observed when students respond to additional questions about force or draw free-body diagrams are consistent with the resources framework. Sabella and Redish
suggested that prompting students to think about forces may activate resources closely associated with forces more often. They refer to this condition as being in a “local coherence” of knowledge related to forces [15]. They then suggested that being in a local coherence about forces makes it more difficult for students to access useful resources outside of that particular coherence [15, 16]. In Sabella and Redish’s study, this led to students not using work and energy ideas. In Heckler’s study, it led to students using formal solution methods that started with the drawing of a FBD, rather than other informal solution methods that more successfully allowed them to make sense of the situation.

The resources framework is clearly useful for making sense of the variability in students’ ideas, and explaining why we sometimes see different responses when we ask students different questions. The next section discusses how we use elements of both the resources and specific difficulties frameworks to discuss students’ responses to question variations.

2.6 Parallelizing Frameworks

The research discussed in this dissertation was clearly influenced by both the resources and specific difficulties frameworks. This section discusses the consistency between these frameworks and the language we use in this dissertation.

In Scherr’s paper, she discusses the consistency of theoretical frameworks that take two different perspectives on students’ thinking. She calls one group the “misconceptions model,” which would contain the specific difficulties framework; the other group she calls the “pieces model,” which would contain the resources framework. Scherr argued that these frameworks are not inconsistent and can both explain the data we see in physics education studies, but that the views change the way research is performed and instructional materials are developed [50]. Furthermore, both Heron and Hammer argue that the specific difficulties and
resources frameworks are not in conflict with one and other; in fact, the resources framework may be a useful way to explain the mechanisms that cause students to have particular specific difficulties [5, 48].

Because the study in this dissertation is informed by both methods, it is important to look at the results from both perspectives. To this end, we need to—to the extent possible—use a combined framework that allows a consistent analysis from both perspectives. We follow Scherr’s lead and use both frameworks providing the impetus for this research together.

When discussing students’ understanding independent of framework, or from a combined perspective, Scherr used the language “ideas” as an umbrella term for misconceptions/difficulties” and “knowledge pieces.” In combining these frameworks in this way, we end up with a term where “ideas” can be stable and appear across multiple content areas, which is more akin to specific difficulties, but not disallowed by resources, or an idea may be malleable and change easily, which is more resources like, but not disallowed by specific difficulties [50]. We use this term, which symbolizes a combination of frameworks, in this dissertation, where the design, results, and analysis are compatible with both frameworks.

Using a term that combines frameworks is necessary because the research is designed from a merged perspective. The results are, in parts, strongly aligned with both perspectives, and we would be remiss to not consistently acknowledge that. We want to be consistently authentic to both frameworks that influenced the current study’s design and analysis. This work should also be accessible and usable to researchers from both perspectives as well as other perspectives. Finally, in this dissertation ideas are:

- Something that could be a knowledge structure, a specific difficulty, or another compatible element from any other theoretical framework. In general, we think the phrase “ideas” could be used as a framework agnostic term (i.e., a
general term that spans elements from multiple frameworks, not just the two we specify);

• What students use to develop reasoning in response to a question;

• Influenced by the question to which students are responding, either by the knowledge structures that the students generate or the specific difficulties they may express in response to that particular question version;

• Often stable, and similar across a large portion of the student population, but not required to be; and

• Variable and influenced by small changes to questions posed to students.

Students may use one or multiple ideas to develop a response. Therefore, we most often use the plural “ideas” to avoid attributing a “grain-size” to them.

Another key term we use here is “alternate ideas.” In the work presented in this dissertation, we create variations of existing research tasks; “alternate ideas” refer to ideas used in response to a variation only or in a different manner or with a different prevalence than on the unmodified version of the task.

When students express different types of reasoning, we take that as a direct indication that they are using different ideas to develop those types of reasoning. These different ideas can be seen as experiencing different specific difficulties, or generating different knowledge structures.

One reason we are able to make this combination of frameworks is that this study is conducted from a question-centric perspective. Because we only ask students for written responses to physics questions, and only see the final reasoning, we are unable to see changes in a students’ knowledge structure as their reasoning develops. Therefore, with only this level of detail, an underlying difficulty specific to that question and an underlying knowledge structure generated in response to that
question are indistinguishable in our data. In a more in-depth study with different
data, these issues may not be indistinguishable, and using the term “ideas” as we do
herein may not be appropriate at that point.

2.7 Investigating Alternate Ideas in Many Content Areas

As we stated in Chapter 1 alongside our research questions, the goal of the work
presented in this dissertation is to demonstrate the effectiveness of a methodology
we developed which enables us to learn more about students’ ideas, and how
students respond to question variations.

By combining knowledge about the variability of students’ ideas with common
methods of investigating students’ ideas, this dissertation demonstrates a method
for performing a deeper investigation of students’ ideas within many content areas
by utilizing question variations. Many studies have explored the difficulties that
students have in many content areas; other studies have explored the variability of
students’ ideas in a few specific content areas, with studies that are designed specific
to those content areas. We explore the variability of students’ ideas in a wide range
of content areas with a methodology that is not content area specific. This research
differs from previous specific difficulties studies because it focuses on using question
variations to access ideas that may not show up in response to only one version of a
question. Moreover, this research builds on previous research on the variability of
students’ ideas in physics by exploring the effect of the same type of variability in
many different content areas.

This is an important area of research because it has the potential to help develop
a better understanding of ideas through both the development and implementation
of tools to access alternate ideas. It has been shown, and explicitly mentioned in
Wittmann et al. [42], that using a single question format may not fully elicit the
ideas students have whereas using multiple question formats can provide more
information about students’ ideas, thereby leading to improved instruction. We are also able to learn more about how students develop responses to various question formats, independent of their ideas in a specific content area.

In seeking to explore the variability of students’ ideas over many content areas in a consistent manner, we are required to use a method that causes students to vary their ideas via a content-independent mechanism. For this reason, we developed a method to explore students’ ideas using variations to the question statement of a common type of PER question. Instead of asking students only to select a correct response and explain why that response is correct, we ask students to consider whether particular responses are correct, to select an incorrect response, and to justify a correct response we provide for them. Students are then asked to explain the reasoning they used to arrive at a response or to provide reasoning that justifies a provided response. This information gives us a content-independent set of variations we can apply to existing PER research tasks in multiple content areas.

Looking at these question variations across multiple content areas allows us to see what alternate ideas students have about each of these physical situations. We can determine how common these alternate ideas are across an entire class of students, and across an entire semester of physics content. We feel that it has been established that resources can be useful for explaining why students express variable ideas in a content area, but we look instead to see what types of ideas we find and what question variations work well for eliciting them in a wide range of content areas.

Additionally, by asking the same types of question variations repeatedly, we are able to investigate how students interact with these question variations, and are able to identify patterns in the types of responses they give to each variation.

The next chapter presents our methodology for systematically investigating alternate ideas across a wide range of content areas. We discuss the different
question variations we ask and why we think these particular questions elicit alternate ideas.
CHAPTER 3
METHODOLOGY FOR INVESTIGATING ALTERNATE IDEAS

The methodology we developed for investigating alternate ideas was based on previous studies that investigated students’ ideas and the variability in those ideas. The goal of this methodology is to enable a more extensive investigation of students’ ideas by developing variations of a question about a single physical situation that can be applied independent of content area; thus allowing us to learn more about students’ alternate ideas and how they respond to question variations. We developed four question variations that meet these criteria and act as tools to elicit a broader range of students’ ideas than any single version of the question.

This chapter explains how our question variations are derived from constraining the response options available, the four variations we developed with this constraint, and why we believe they are capable of accessing alternate ideas.

3.1 Imposing Constraints to Access Alternate Ideas

Our general approach to creating questions to access alternate ideas was to constrain the set of possible response options with which students are presented. We used two types of response constraints. The first constraint was to limit the number of responses from which students were allowed to choose. The second constraint was having students justify why a particular outcome is incorrect, rather than justifying why one is correct. As demonstrated in in Section 3.3, asking students to explain why something is incorrect is a natural consequence of presenting students with only one response option.

These constraints were expected to elicit alternate ideas but were also selected for a few pragmatic reasons as well. With these constraints, we could develop isomorphic versions of many types of questions, with the only real difference between them being the phrasing of the question being asked. This approach meant
we did not have to worry about the effects of changing other parts of the question, such as altering the physical situation or inadvertently providing cues. As these tasks are similar to existing tasks, students should also easily understand these variations and be able to engage with them successfully.

The next section explains why we think constraining response options will elicit alternate ideas. The following section presents the question variations we developed and explains how constraining response options leads us to them.

3.2 Constraining Response Options to Access Alternate Ideas

Our main principal behind the development of question variations is that, by constraining response options, we change the focus of students’ attention and hopefully cause them to express different specific difficulties or activate different resources and develop different knowledge structures due to this shifted focus. For example, a student who would select greater than on an unmodified version of a task may respond differently when responding to a task that only asks them only to consider the response less than because the student is prompted to think about less than, but not prompted to think about greater than.

The main reason we believe that this shift in attention will have a noticeable impact on students’ responses is due to Wittmann et al.’s [42] work discussed in Section 2.4.2. They showed that providing students with different response options may cause them to activate alternate resources or consider resources from another local coherence related to that particular response; constraining their response options is expected to work in a similar way. Additionally, the visual cues used in Frank et al.’s study, discussed in Section 2.5.1 [14], are a similar level of alteration as the modifications in our variations and also work by shifting students’ focus.

Frank et al. used the resources framework to explain that the resulting changes are likely due to students activating different resources when presented with
different visual cues. In our study, students could have a similar reaction by activating different resources related to the response they are being focused on or when being asked to explain why a response is incorrect. In addition, removing response options may prevent resources related to those response options from being activated and change the knowledge structures that develop.

Students are expected to be even more likely to express alternate ideas if the response they are provided or focused on is inconsistent with the knowledge structure, local coherence, or specific difficulty they would have used to answer an unmodified version of the question, as some of the elements they use to formulate their response to the unmodified version may no longer be applicable.

Different variations could focus student attention differently, and the magnitude of the shift in focus depends on what response the student would have chosen on the original version.

When a variation removes a response option that a student would have selected as correct on an unmodified version, those variations will force the student to consider a different response, which could dramatically shift the knowledge structures the student generates and perhaps prevent the student from using resources or knowledge structures they would use otherwise.

Additionally, we ask students to eliminate responses, which may cause students to use alternate ideas when responding. Asking students to eliminate a response could influence their reasoning because explaining why something does not happen often requires different reasoning than explaining why another thing does happen.

The choice to have students eliminate a response was not derived from prior research or the resources framework, but as a consequence of offering only a single response option. As such, no strong prior evidence suggests that this change will elicit alternate ideas. One of the key questions we hope to answer through this
research is whether students use the same ideas when justifying and eliminating responses.

3.3 Description of Question Versions

All of the variations discussed in this section are modifications made to an existing question. We refer to the existing version of the question as the *original* version. The *original* version describes a physical situation and students are asked about the outcome of that situation. Students are offered a set of possible outcomes to select as a response, such as *greater than*, *less than*, or *equal to*. They are then asked to select the correct outcome of the physical situation and explain why that outcome is correct. This is a common type of conceptual question used in physics education research and research-based instructional materials [3, 18, 24, 40].

The set of four question variations we developed are the *eliminate*, *given information*, *consider*, and *given correct* variations. The *eliminate* variation provides students with the same set of possible outcomes as the *original* version, but asks students to eliminate an outcome they think is incorrect and explain their reasoning for eliminating it. The *consider* variation asks students to consider one particular outcome from the options presented in the *original* version, state whether they think that particular outcome is correct or incorrect, and to explain their reasoning. The *given correct* variation provides students with the correct outcome and asks them to justify why it is correct. Finally, the *given information* variation tells the students that a particular outcome does not occur and asks them to justify why it does not. Table 3.1 provides examples of each question version. All versions ask students to explain their reasoning.

Because *consider* and *given information* variations can be created for each particular outcome of the physical situation, their names are amended to reflect the particular outcome on which they focus. For example, the pendulum question in
Table 3.1. Examples of each question variation.

<table>
<thead>
<tr>
<th>Version</th>
<th>Example of Question Phrasing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>At the bottom of its swing, is the acceleration of a pendulum <em>upwards, right, or zero</em>? Explain.</td>
</tr>
<tr>
<td>Eliminate</td>
<td>For the following question, please select a response that is incorrect and explain why it is incorrect: At the bottom of its swing, is the acceleration of a pendulum <em>upwards, right, or zero</em>? Explain.</td>
</tr>
<tr>
<td>Consider</td>
<td>At the bottom of its swing, is the acceleration of a pendulum upwards? Explain.</td>
</tr>
<tr>
<td>Given Correct</td>
<td>At the bottom of its swing, the acceleration of a pendulum is upwards. Explain.</td>
</tr>
<tr>
<td>Given Information</td>
<td>At the bottom of its swing, the acceleration of a pendulum is not zero. Explain.</td>
</tr>
</tbody>
</table>

Table 3.1, *consider* versions could be made with each one asking about a particular outcome (*upwards, right, or zero*). The *consider right* variation would be the one that asks “At the bottom of its swing, is the acceleration of a pendulum to the right? Explain.” The *given information right* variation would say “At the bottom of its swing, the acceleration of a pendulum is not to the right. Explain.” Consider variations can be made for both correct and incorrect outcomes while *given information* variations can only be made for incorrect outcomes.

The *original* and *eliminate* versions do not target a particular outcome in their phrasing, and the *given correct* variation can only target the correct outcome, so there are not multiple versions of these questions. The possible target responses for each question version are shown in Table 3.2.

As discussed in the previous section, this set of variations was defined by constraining response options. This set of variations is a complete set, as it spans the space of question variations that limit the number of responses students are allowed to choose and asks them to explain why something is correct or incorrect.
Table 3.2. Possible responses that can be targeted with each question version.

<table>
<thead>
<tr>
<th></th>
<th>Correct Response</th>
<th>Incorrect Response A</th>
<th>Incorrect Response B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eliminate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consider</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Given Correct</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Given Information</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 3.3. Number of choices offered for each version. In the consider variation students have the option to justify why the outcome is correct or incorrect, so it spans both columns.

<table>
<thead>
<tr>
<th>Number of Possible Responses</th>
<th>Justify Why an Outcome is Correct</th>
<th>Justify Why an Outcome is Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 or More</td>
<td>Original</td>
<td>Eliminate</td>
</tr>
<tr>
<td>2</td>
<td>Consider</td>
<td>Consider</td>
</tr>
<tr>
<td>1</td>
<td>Given Correct</td>
<td>Given Information</td>
</tr>
</tbody>
</table>

The spanning of this space is shown in Table 3.3 By spanning this space, the goal is for all of our question variations to function in distinctly different ways leading to different results.

In this dissertation, we implement our question variations in multiple content areas and compare the responses to the original version to each variation, identifying alternate ideas, and investigating responses to these particular variations.

3.4 Conclusion

The methodology developed uses four question variations to investigate students’ alternate ideas about a single physical situation. These four question variations were defined by constraining the number of response options presented to students and asking students to justify why responses are correct or incorrect. These
variations are expected to elicit alternate ideas by focusing students’ attention on particular outcomes or making them explain why a response is incorrect.

In addition to the research questions presented in Section 1.3, we investigate some additional research questions specific to these variations intended to help us learn more about how students respond to different types of questions:

- How do students respond to a question that is not asking them to provide a correct response? Do they typically try to give the right response anyway? To what extent can students explain why an incorrect response is incorrect?
- How do students respond to a question that is providing them with a correct response? Do they often have or develop reasoning to support it?
- Do all the variations lead to similar types of responses, or does the amount the responses are constrained matter?

The next chapter discusses a pilot study used to test our methodology and determine if these question variations are capable of eliciting alternate ideas. Chapter 5 then presents a larger study that aims to answer all of the research questions presented.
Following the development of the methodology for accessing alternate ideas with question variations, we developed a pilot study to test the methodology. This section briefly describes the implementation of the methodology used for the pilot study and the results. The goals of the pilot study were to investigate whether students knew how to engage with the question variations and if students’ responses indicated that they were using alternate ideas when responding to the variations.

The pilot study demonstrated that most students were capable of understanding and responding appropriately to the variations, and they appeared to answer the questions in a way that gives us more knowledge about their ideas than the original version alone does.

This chapter presents the methodology developed for implementing our question variations, discusses how we developed our question variations in this content area, and presents an analysis of the results.

4.1 Pilot Study Methodology

During the pilot study, we distributed one question version to each student in a single-large-calculus-based introductory physics class at the University of Maine (UMaine). Consistent with the general methodology described in Chapter 3, our four question versions were all based on the same physical situation, and each student answered only one question. The questions were distributed evenly during a lecture, with each question going to approximately one quarter of the class.

We made variations of an electric circuits pretest question from *Tutorials in Introductory Physics* [3]. The original version asks for a correct response and an explanation (see Figure 4.1). We selected this pretest question because it was from an appropriate content area that was about to be introduced in the introductory
The circuit at right contains an ideal battery, three identical light bulbs, and a switch. Initially the switch is open.

After the switch closes:

Does the brightness of bulb A increase, decrease, or remain the same? Explain.

Figure 4.1. The *original* version of the question [3].

physics courses at UMaine, and there was published research on students’ ideas about this content that could guide the development of our variations and the coding of students’ reasoning [8].

This question asks students to explain what happens to an indicator bulb (Bulb A) when a switch on a branch is closed. The correct response is that, once the switch is closed, the total resistance of the circuit decreases; therefore, the total current flow though the battery increases. As Bulb A has the same current flow as the battery, Bulb A becomes brighter.

In deciding which question variations to use, we looked at the results of previous research, which showed that most students have many specific difficulties related to electric circuits [8]. McDermott and Shaffer found that “most students in an introductory physics course do not develop the type of conventional understanding that enable them to apply the basic electrical concepts” [8, p. 1002]. The specific difficulties they documented range from a failure to apply the concept of a complete circuit, with 50% of students failing to draw a complete circuit diagram for one question, to several difficulties related to voltage and resistance.

Because of this prior research, we expected students to answer this question incorrectly often; therefore, our designed variations focused students on the correct response whenever possible. We expected that directing students’ attention to a

\[1\]This specific research task was not addressed explicitly in McDermott and Shaffer’s paper [8].
The circuit at right contains an ideal battery, three identical light bulbs, and a switch. Initially the switch is open. After the switch closes: Does the brightness of bulb A increase, decrease, or remain the same? Explain.

Imagine you are taking an exam with the question shown in the box below. You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate?

The circuit at right contains an ideal battery, three identical light bulbs, and a switch. Initially the switch is open. After the switch closes, does the brightness of bulb A increase? Explain.

We first used the original version. Asking the original version provided a reference point to which variations could be compared.
The second question version we used was an *eliminate* variation, which asks students to select an incorrect response and explain their reasoning. Asking the *eliminate* variation indicates which responses students think are incorrect and what ideas students use to explain why they are incorrect.

The third question version, *consider increase*, asks students if *increase* is a correct response and for them to explain their reasoning. Asking the *consider increase* variation lets us know what students think of the correct response even if they would not have chosen it as the correct one. It also indicates which ideas students use to explain why the correct response is incorrect.

The fourth question version was the *given correct* variation, which tells students that *increase* is the correct response and asks students to justify why it is correct. The *given correct* variation indicates whether students have correct ideas for explaining why the correct response is correct, even if they may have initially thought of it as incorrect.

We chose not to use the *given information* variation because we did not want to divide our students amongst even more questions and believed this variation was similar enough to the *given correct* or *eliminate* variations that we did not need to test it. In addition, a *given information* variation would have told students a particular response was incorrect, which did not fit our targeted objective of learning what students thought about the correct response.

Because the adaptation of question variations to a specific research task is applicable to all of the results in the dissertation, the development of these questions, especially the *eliminate* variation, is discussed in further detail in Section 5.3.
4.2 Analysis

Because of the low number of students in this study, our analysis is somewhat different from the analysis used in the rest of this dissertation. These question variations were designed to elicit alternate ideas, and we investigated ideas through student reasoning. Therefore, the focus is on the reasoning students use to learn more about the ideas they are using.

However, most pilot study questions did not receive enough student responses to allow for conclusions to be drawn about changes in reasoning by comparing their prevalence. Differences in response distributions are an indication that different ideas are being used, thereby allowing for conclusions to be drawn from the data; yet identifying these differences does not provide as much information about the ideas students are using as an analysis of student reasoning would. However, we will comment on some aspects of student reasoning as appropriate, even if we can not always make meaningful quantitative comparisons based on student reasoning.

One exception is that all students were provided with a response on the given correct variation, so we must look at student reasoning on this variation. As all students provided reasoning for the same response, it condenses the variety of student reasoning in a way that does not occur with the other questions and allows us to draw meaningful conclusions despite the low number of responses received. The categories of student reasoning we use here were developed based on those discussed by McDermott and Shaffer [8].

The analysis of the pilot study results highlighted two ways that responses to this set of question versions can be discussed. Results can be broken up by version to determine the distribution of responses to each question or by possible response to see what the set of questions indicates about each particular response. We first briefly look at the questions versions individually and then at the possible responses individually.
4.2.1 Results by Question Version

The results of all four question versions are displayed in Figure 4.2, and each version tells us something about what ideas students are bringing into the course. Most students did not respond correctly to the original version, with remains the same being the most commonly chosen response. On the eliminate variation, students often chose increase and remains the same as incorrect, but not decrease. Nearly all students answering the consider increase question indicated increase as incorrect. The given correct variation demonstrated that students used two types of reasoning more prevalently when provided with the correct outcome. Students used \( R \text{ decreases} \) reasoning to explain that the bulb is brighter because the total resistance of the circuit has decreased, which is correct reasoning. Students also used branch/flow reasoning to explain that the bulb is brighter because there are more branches and/or more flow in the circuit, which could be correct or incomplete. We are often unable to differentiate correct and incorrect responses in these types of reasoning and are forced to code these responses together.

Although these versions each provide valuable insights into students’ ideas about electric circuits through their response distributions and provided reasoning, more interesting results stem from looking across question versions.

4.2.2 Results by Response

The original and eliminate versions included three responses as possible options; this section discusses their results with each possible response. The consider increase and given correct variations focus on the response increase and will only be discussed in the section for that particular response. The pilot study did not include enough student participants to provide statistically significant results, so these results are only discussed as indicators of results we could potentially find in a larger study.
4.2.2.1 Decrease

*Decrease* was an option on the *original* and *eliminate* versions. The results show that 27% of students receiving the *original* version chose it as a correct response, whereas 6% of the students receiving the *eliminate* variation chose it as the best response to eliminate. Combining these results suggests that some students viewed the response as correct, but very few selected it as incorrect.

4.2.2.2 Remains the Same

*Remains the same* was also available on two questions. On the *original* version, *remains the same* was the response most commonly chosen as correct, with half of the students choosing it. On the *eliminate* variation, *remains the same* was most commonly eliminated as incorrect, with half of the students eliminating it. This interesting result offers two possible interpretations. One is that half of the class thought *remains the same* was the correct response, whereas the other half thought it was incorrect. A different interpretation is that the different versions may elicit contrasting responses, allowing for the possibility that a student may pick *remains the same* as correct or pick *remains the same* as incorrect, depending on the question asked.

Students’ reasoning for eliminating or selecting *remains the same* also indicated an interesting contrast. The most common justification for why *remains the same* is correct is what McDermott and Shaffer referred to as “local reasoning.” [8] These students’ responses indicated that they were thinking about only a small region of the circuit. As one student explained, “the brightness of Bulb A will not change because it is not part of the system with the switch.” The most common justification for *remains the same* being incorrect was a holistic reasoning, where students reasoned that, when a change is made to one part of a circuit by closing a switch, the whole circuit changes. For example, “I want to eliminate ‘remain the
same’ because I feel that the brightness would have to change due to the change in circulation.”

Combining these two question versions suggests that remains the same has a distinctly different profile from decrease. Although decrease seemed correct to a few students and was selected as incorrect by even fewer, remains the same was commonly selected as both correct and incorrect.

4.2.2.3 Increase

On the original version, 23% of students selected increase as the correct response. On the eliminate variation, 44% of students eliminated increase. These results show a difference between the responses increase and decrease as the percentage of students who choose them as correct was about the same, but increase was actively chosen as incorrect whereas decrease was not.

On the consider increase variation, only 8% of students said that the brightness of the bulb increases. There seems to be an inconsistency between 23% of students choosing increase as the correct response on the original version and only 8% saying the brightness increases on the consider increase variation.

One simple explanation for this phenomenon is that students have a bias towards answering in the negative on consider variations.

Another possible explanation is that students’ ideas about why increase is incorrect are more commonly activated by the consider increase variation. As evidenced by the eliminate variation, students have access to reasoning that explains why increase and remains the same are incorrect. On a question that does not mention remains the same or ask students to justify why something is correct, students might end up activating only ideas related to increase being incorrect.

Finally, regarding the reasoning provided to justify increase on the given correct variation, students offered valuable reasoning about why the brightness of the bulb
increases. This is highly contrasted with the other three versions, where students repeatedly indicated they did not view it as correct and, in fact, viewed it as incorrect.

Students used two common types of reasoning to justify why the brightness of the bulb increased in response to the given correct variation.

The most common type of reasoning was informal branch/flow reasoning, which was used more prevalently on the given correct variation (46%) than the original version (9%). We could not determine if students using this reasoning had a correct holistic view of the circuit or if they were incorrectly focusing only locally at the junction among the three bulbs. Therefore, we are not sure if students are applying correct reasoning to this physical situation or incorrect reasoning that happens to be consistent with the correct response. However, students were clearly more likely to use branch/flow reasoning on the given correct variation than the original version, showing that the reasoning they used differed when they were provided with the correct response.

The second type of reasoning used was formal $R$ decreases reasoning, which was used by 19% of students to justify why the brightness increases compared to only 9% of students using this correct form of reasoning on the original version. Thus, more students may have access to correct reasoning about the total resistance of the circuit than indicated by the other question versions.

One possible explanation for both of these differences is that students who are capable of identifying that resistance is lower when a bulb is added in parallel, or that adding a bulb in parallel allows for more flow, may be prevented from doing so when asked to select a correct response. They may be prevented from using this reasoning because the presence of other response options activates resources related to those not activated here or because they realize the knowledge structure or local coherence they developed is inconsistent with the information they have been given;
they are then able to bring in new and correct ideas from outside that local coherence.

When combined with the original version, these question variations give a nuanced view of what students think about increase. Instead of just knowing that some students think it is a correct response as the original version alone would tell us, we now have more detailed knowledge about what students think about the response increase: some students choose it as correct, nearly half eliminate it as incorrect, very few say it is correct, and nearly two-thirds of students can give correct reasoning when told it is correct.

4.3 Conclusion

The results from the pilot study show a lot of contrast, especially in relation to the responses remains the same and increase.

When looking at remains the same, half of the students selected it as the correct response while the other half eliminated it as incorrect. The fact that it was the most common selection on two essentially opposite questions shows that many students have ideas that support or counter this response.

Meanwhile, the increase response was rarely viewed as correct, infrequently selected as correct, and eliminated by about half of the class. Yet approximately 65% of the class came up with reasoning to support this response when it was provided for them. This result clearly shows variability in students’ ideas; in this case at least, students have productive ideas that support a correct understanding of the physics, even though it is a question they struggled to answer correctly. These ideas were not seen by the original version of the task alone, which in the past led researchers to conclude that students lacked these ideas altogether.

These differences in students’ responses and reasoning to the question variations indicate that the variations are eliciting alternate ideas.
In short, students can engage with these variations and provide different responses and reasoning to different variations. The pilot study demonstrated that our methodology can do what we intend and allowed us to learn more about the ideas students have in an electric circuits context. Furthermore, our methodology has the potential to show more about students’ thinking in a wide variety of content areas.
CHAPTER 5
STUDY DESIGN, ADMINISTRATION, TASK DESIGN, AND ANALYSIS METHODS

This chapter discusses the design of the main study conducted in this dissertation, including the details of the administration method and why it was selected. It discusses the design of question variations in general and the design of the broader study that includes these variations. Finally, this section discusses the analysis methods used to draw results from the data.

5.1 The Across-Content Study

The pilot study showed that the methodology of asking variations of questions had the potential to provide more information about students’ ideas in a specific content area. Based on the success of the pilot study, an across-content study was designed to elicit more information about students’ ideas within many content areas and determine the extent to which asking the question variations described earlier could elicit alternate ideas in many content areas. The across-content study also sought to determine if patterns in student responses would provide more information about how students respond to these types of variations in general. The data collected in this study makes up the majority of the data discussed in this dissertation.

The goal of the across-content study was to investigate all of the research questions presented in Section 1.3. Broadly speaking, the focus was to continue to test the question variations to determine if they could elicit alternate ideas consistently across many content areas. An additional goal was to learn more about the specific alternate ideas elicited by these question variations and about the question variations and how they elicited alternate ideas.
An online system was used to ask students conceptual physics questions over the weekend for each week of a one-semester introductory physics course. The online system presented students with a pretest, which is a collection of questions designed to assess students’ understanding before they receive instruction. Each pretest consisted of a set of questions about multiple physical situations, with each page of the pretest containing a graphic and text description of a physical situation, and one or more questions about the physical situation for the student to respond to. The pretest system was incorporated into the course, and the results were occasionally used to inform the instructors about students’ ideas. Within these pretests, some questions posed the same version to all students, and some were research tasks that posed different variations to randomly selected groups of students. The layout of the across-content study is shown in Table 5.1, and a description and the demographics of the course are provided in Appendix E.

The methodology describes four question variations. Although the focus is on investigating all of these variations, the decision was made to not investigate the given information variation in this study. This particular variation was removed because it was similar to the given correct variation, and we observed in the pilot study that targeting the correct response can lead to surprising results, and we wanted to continue focusing on reasoning about correct responses most of the time. Because only three question versions were asked each week, choosing to not investigate one of the variations provided the opportunity to better investigate the other three variations.

5.2 Administration Method

Because the across-content study required weekly administrations of research tasks, an online administration system was implemented to allow students to answer these questions without taking up excessive class time.
Table 5.1. Pretest designation, content area, and types of variation administered for each pretest administration. The *consider* variations that asked students to consider the correct response are indicated.

<table>
<thead>
<tr>
<th>Pretest</th>
<th>Question Name</th>
<th>Question Versions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest 2</td>
<td>Oval Track</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Greater Than(Correct)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Less Than</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Equal To</td>
</tr>
<tr>
<td>Pretest 3</td>
<td>Pendulum</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Upwards(Correct)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 4</td>
<td>Car-Over-Hill</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Greater Than(Correct)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 5</td>
<td>Newton’s 3rd Law</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Greater Than</td>
</tr>
<tr>
<td>Pretest 6</td>
<td>Gymnast</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 7a</td>
<td>1-D Conservation of Momentum A</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Greater Than</td>
</tr>
<tr>
<td>Pretest 7b</td>
<td>1-D Conservation of Momentum B</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Less Than(Correct)</td>
</tr>
<tr>
<td>Pretest 8</td>
<td>2-D Conservation of Momentum</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 9a</td>
<td>Change in Kinetic Energy</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Equal To(Correct)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 9b</td>
<td>Work-by-Wall</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 10</td>
<td>Change in Momentum</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Equal To</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 11</td>
<td>Pivoting Rod</td>
<td>Original, Consider Equal To</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 12</td>
<td>Rotating Disk/Wheel</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Greater Than(Correct)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 13</td>
<td>Simple Harmonic Oscillator</td>
<td>Original, Consider Zero</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
</tbody>
</table>
The general guidelines for developing an administration method were to get responses from many students, with each student being asked only one variation; allow students sufficient time to think about the question and provide a thoughtful response; keep any intrusion into the classroom environment as minimal as possible; and provide some benefit to the students in the class.

Responses were gathered from a large portion of the class by making the pretest a required part of the course while still allowing students to opt out of the research project. The questions were incorporated into the class by matching them to the course’s current content area, which also provided instructors with relevant results and hopefully improved the instruction students received. These tasks were administered outside of the classroom environment because, in other studies requiring repeated administrations, students resented having to give up their class time so frequently for research purposes.

Because the online administration was to be used for both for research and as pretest administration tool for the course instructors, the online system the University of Washington uses for research and pretest administration was employed [51].

The Physics Education Group at the University of Washington’s strategy of opening availability when students were done with class for the week and ending it just before students started class the following week was adopted. This strategy removes the need to account for instruction that takes place while the pretest is available and still provides a large period of time for students to take the pretest.

The pretests were made available to students via an email link that was sent to them each week. Once they started taking the pretest, they had 15 minutes to complete it. If they did not finish within 15 minutes, their responses were submitted automatically, and they were taken to a page saying the pretest was over. This was done to encourage students to answer questions without using external resources,
Figure 5.1. How the pretest uses the randomization task to break students into random groups.

such as looking up information in their textbook or searching for an answer on the internet. In addition, the time limit was imposed to ensure that students knew this pretest would never take more than 15 minutes per week.

Students were only allowed to take the pretest once. They were not allowed to go back and change their responses after they had submitted their answers.

To give different students different tasks as part of the experiment, students were split into random groups, and each group received a different version of the online pretest, as shown in Figure 5.1. The details of how students were broken into random groups are discussed in Appendix C.

Each of the randomized pretest groups received some questions that were the same as those the other groups received as well as other questions that were unique to each group. The arrangements of questions on the pretests are discussed in more detail in Section 5.4 on pretest design.

The online pretests were established as a standard part of the course. Students were informed they would be taking a weekly pretest as part of the course through the syllabus and an introduction by the course instructor.
Students were told that their pretest would be graded only on participation and that their responses to in-class clicker questions and online pretests would be combined to account for 6% of their grade.

5.3 Variation Design

Although the methodology describes the types of question variations to ask, this section discusses the process of selecting questions to make variations of and the details for transforming the original version into each one of the variations.

In order to develop question variations in many content areas, it was important to determine which types of questions to make variations of and establish the set of modifications to make to the original version to transform it into each variation. The methodology is focused on multiple-choice conceptual questions, where students are asked to explain their reasoning; therefore, every question with variations had to fit those requirements. In addition, only questions that had been previously used in physics education research studies were used.

One of the benefits of using questions from previous physics education research studies is that previous researchers had tested and often refined the questions. Thus, these questions can be implemented without having to trial the original version of questions before making variations, as students are not likely to misunderstand them. Another benefit is that the questions are not likely to be too easy for students, thus avoiding a ceiling effect where all of the students respond correctly.

Having existing data from other institutions also provided benefits for selecting variations. When available, the results from previous administrations of the questions indicated the most commonly chosen response, the least commonly chosen response, and how often the correct response was chosen. Knowing these factors helped make decisions about which variations to use and which responses to target with the variations. For example, if students often chose a particular response, a
Table 5.2. Examples of consider, given correct, and given information variations of a single question. This case asks about the direction of the acceleration of a pendulum at the bottom of its swing.

<table>
<thead>
<tr>
<th>Version</th>
<th>Question statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>When it reaches the bottom of its swing, is the direction of its acceleration upward, upward and to the right, right, downward, or zero? Explain.</td>
</tr>
<tr>
<td>Consider Upward</td>
<td>When it reaches the bottom of its swing, is the direction of its acceleration upward? Explain.</td>
</tr>
<tr>
<td>Given Correct</td>
<td>When it reaches the bottom of its swing, the direction of its acceleration is upward. Explain.</td>
</tr>
<tr>
<td>Given Information</td>
<td>When it reaches the bottom of its swing, the direction of its acceleration is not downward. Explain.</td>
</tr>
</tbody>
</table>

consider variation targeting that response should not be used because the responses would be highly similar to those elicited by the original version.

In designing the wording of our variations, the goal was to keep the variations as similar to the original version of the question as possible. This allows for comparing student reasoning on the original version to student reasoning on the variations without having to worry about differences beyond the intended variation of the task.

5.3.1 Consider, Given Correct, and Given Information Variations

For the consider, given correct, and given information variations, it was relatively straightforward to create variations in wording that were very similar to the original version. All of the original versions of the modified questions had a sentence at the end that posed the question to which students were to respond. This sentence was modified with as few changes as possible to create a variation. No changes were made to the description of the physical situation presented in the question. An example of these variations is shown in Table 5.2.
In addition to changing the question statement, the possible response options were adjusted. On original versions, the options were often listed as complete sentences. The modified response options were as similar to those original options as possible, but changed to present the analogous responses for the variation. An example is shown in Table 5.3.

For more examples, see the pilot study in Figure 4.2 and examples from the across-content study in Chapters 6 and 7.

5.3.2 Eliminate Variation

Creating eliminate variations of questions required more substantial modifications than the other variations. A simple change to the question statement could not be developed for this particular variation. One of the reasons this variation needed a larger modification was that it was intended to prompt students to think about which responses are incorrect before thinking about what is correct. The objective was not to have students saying that one response was incorrect because another was correct, providing the same information as the original version. In order to have the students approach the problem as a task in which they identified one response that is incorrect, it was necessary to add text explaining this task to them before the presentation of the question.

The added text was: “Imagine you are taking an exam with the question shown in the box below. You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate?” Students would then be shown the original version of the question. The text was bolded to make sure students read it and to ensure that it stood out as instruction rather than more discussion of the physical situation presented to the students.
Table 5.3. Examples of consider and eliminate variations that could be made of a single question and the possible responses. Again using the pendulum question as an example. Responses to the eliminate version have been condensed.

<table>
<thead>
<tr>
<th>Version</th>
<th>Question statement and response options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>When it reaches the bottom of its swing, is the direction of its acceleration upward, upward and to the right, right, downward, or zero? Explain.</td>
</tr>
<tr>
<td></td>
<td>• The acceleration of the pendulum bob at the bottom of its swing is upward.</td>
</tr>
<tr>
<td></td>
<td>• The acceleration of the pendulum bob at the bottom of its swing is upward and to the right.</td>
</tr>
<tr>
<td></td>
<td>• The acceleration of the pendulum bob at the bottom of its swing is to the right.</td>
</tr>
<tr>
<td></td>
<td>• The acceleration of the pendulum bob at the bottom of its swing is downward.</td>
</tr>
<tr>
<td></td>
<td>• The acceleration of the pendulum bob at the bottom of its swing is zero.</td>
</tr>
<tr>
<td>Consider</td>
<td>When it reaches the bottom of its swing, is the direction of its acceleration upward?</td>
</tr>
<tr>
<td>Upward</td>
<td>• The acceleration of the pendulum bob at the bottom of its swing is upward.</td>
</tr>
<tr>
<td></td>
<td>• The acceleration of the pendulum bob at the bottom of its swing is NOT upward.</td>
</tr>
<tr>
<td>Eliminate</td>
<td>Imagine you are taking an exam with the question shown in the box below. You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate? When it reaches the bottom of its swing, is the direction of its acceleration upward, upward and to the right, right, downward, or zero? Explain.</td>
</tr>
<tr>
<td></td>
<td>• I would eliminate the response “The acceleration of the pendulum bob at the bottom of its swing is [upward, right, ...].”</td>
</tr>
</tbody>
</table>
One other change made was to the possible response choices. Instead of just listing the responses in the same way as the original version, the text “I would eliminate the response...” was added to the front of each option so that students would be reminded to choose a response they wanted to eliminate. This helped reduce confusion about whether the students had selected responses they thought were correct or incorrect. An example of an eliminate variation and the possible responses provided to the students is shown in Table 5.3.

5.4 Pretest Design

All pretests were expected to fit into a few constraints. Questions were included on each pretest to ensure that all students would answer the same original version to help establish that the random groups were similar, which also allowed for the administration of extra formative assessment questions without variations. Finally, the class was broken into a consistent number of random groups each week, which required determining the optimal number of groups to use.

5.4.1 Questions Without Variations

Asking questions without variations served two purposes in this study.

The primary purpose for questions without variations in this research project was to help establish group similarity. A question where all students received the same version was a “base question.” Base questions allowed for a comparison of the different groups into which students had been randomly assigned. Because of the large number of base questions, equivalence on a single base questions did not ensure that the groups were the same, nor did nonequivalence ensure that they were different. However, looking at the base questions over all pretests prevented consistently seeing large differences between random groups and supported the argument that the differences identified were due to the variations. The analysis of
base questions is discussed in Appendix D and shows that the groups were, on average, similar.

Another purpose of these questions was to act as additional assessment questions. To enable future research and help further inform instructors, additional unvaried questions that were not part of any research study were asked. If the questions with variations were significantly shorter than the allotted 15 minutes per week, other questions were added.

5.4.2 Number of Variations Tested per Question

One of the most critical aspects of creating the pretests was determining how many variations of each question to administer. Four variations of the original version were developed, and the number of variations used had to be balanced with the number of student responses received.

This study used Fisher’s Exact Test with one degree of freedom to look for differences in reasoning between questions. The exact analysis methods are discussed further in Section 5.5. Based on this analysis method, a power analysis was performed to decide how many groups to break the class into each week.¹

The purpose of the across-content study was to find results that showed these question variations working, not to catalog all possible results. Therefore, we can use a lower power than normal, and a wider variety of variations, in order to look for results. This may cause us to miss results, but allows us to perform a study that covers more question variations.

Based on the enrollment in the course, responses from approximately 200 students per week were expected. As this was the first large-scale investigation of these question variations, differences of at least 15% in student use of a particular

¹Power ranges from 0 to 1 and can be described as the odds of finding a statistically significant effect if one is present to be found. Typical power is usually around .8, meaning there is an 80% chance of finding a statically significant effect if it is present.
Table 5.4. The statistical power of Fisher’s Exact Test. Calculated for 2 to 5 question variations, assuming a student population of 200 students and a 15% difference in reasoning.

<table>
<thead>
<tr>
<th>Number of question versions</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students per question (approximate)</td>
<td>100</td>
<td>65</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>Power for rare (10%) higher prevalence</td>
<td>0.76</td>
<td>0.54</td>
<td>0.41</td>
<td>0.33</td>
</tr>
<tr>
<td>Power for common (40%) higher prevalence</td>
<td>0.52</td>
<td>0.32</td>
<td>0.27</td>
<td>0.21</td>
</tr>
<tr>
<td>Power for common (40%) lower prevalence</td>
<td>0.56</td>
<td>0.38</td>
<td>0.29</td>
<td>0.22</td>
</tr>
<tr>
<td>Power for frequent (70%) higher prevalence</td>
<td>0.66</td>
<td>0.46</td>
<td>0.35</td>
<td>0.28</td>
</tr>
<tr>
<td>Power for frequent (70%) lower prevalence</td>
<td>0.53</td>
<td>0.37</td>
<td>0.27</td>
<td>0.21</td>
</tr>
</tbody>
</table>

type of reasoning across groups were identified. The smallest difference in reasoning commonly discussed in papers is 10%-15%, and this study targeted effect sizes at least this large.²

In all studies, the original version of the question was asked along with at least one variation in order to be able to make comparisons back to the original version. Therefore, the minimum number of questions was two, and the maximum was five, given that four variations were created. The power of the study was calculated as a function of the number of variations administered, using effect sizes equivalent to a 15% difference in the usage of a particular type of reasoning.³ Because the power of detecting a 15% difference in the prevalence of reasoning depends on how often that reasoning is used, the power was determined for reasoning that was rare (10% of students), common (40% of students), and frequent (70% of students). The power calculations assumed an evenly distributed number of people in each group and were calculated for finding both an higher and lower prevalence of a particular type of reasoning. The results are shown in Table 5.4.

The results show a clear trade-off between the number of questions and power. Two questions were not considered as an option; because the original version had to

²For examples of the discussion of effects of this size, see [8, 52, 53, 54].
³Calculations were performed using the statistics program G*Power-3 [55].
be asked every week, there would only be time to ask each variation three or four times within a semester, which did not complete the objective of administering each variation in many content areas.

The power seemed too small for five question versions, with less than 33% probability of finding an effect, even if an effect existed to be found. In this case, although it was more likely for there to be effects to be found, it would be very unlikely they were detected.

Eliminating those two possibilities, it was decided to ask three or four versions per week. However, shortly after starting to run pretests, the total number of responses dropped to around 170 students per week. In light of the all-around decrease in power due to the decrease in the number of accessible students, only three questions per week were asked to maintain the higher than 33% odds of finding the type of effect sizes sought.

Most studies have a power of at least .8, but a lower power was used in this study because of its exploratory nature.

5.4.3 Pretest Schedule

The across-content study was designed to apply these question variations throughout a semester in order to explore if these variations could elicit alternate ideas many content areas; therefore, the variations were spread evenly across the semester.

Fourteen questions with multiple variations were asked over 11 weeks of the semester. The consider variation was asked 13 times, the eliminate variation 6 times, and the given correct variation 10 times. In two instances, a single pretest contained two questions with variations. In these cases, the base questions between the variations were used to determine if the groups were affected by the first variation; possible interactions will be discussed when the data from those variations
are presented. An overview of the questions asked on the pretest each week is shown in Table 5.1. The question versions chosen were distributed as evenly as possible, but were also chosen based on the attributes of the question being varied.

Not included in these counts is the first pretest administered, which was a study within itself and was published separately by Wittmann and Hawkins [56]. The design of the first pretest was substantially different from the standard pretest design. It is not part of the across-content study, and for this reason will not be discussed further here.

5.5 Data Analysis

The goal of the data analysis was to compare students’ responses and reasoning between variations and the original version—specifically, to determine if the question variations elicited different reasoning, or if particular types of responses or reasoning were used more or less often in response to question variations. Identifying these differences showed the alternate ideas students used to respond to question variations. The following analysis methods were employed on either students’ selected choice of response or their reasoning as coded by a researcher.

As discussed in Chapter 4, the data analysis focused on student reasoning rather than just the responses they chose, because reasoning provides more information about the ideas students used. Therefore, the investigation explored differences in reasoning, but also talked about differences in responses when they occurred.

5.5.1 Reasoning as Evidence of Ideas

Chapter 2 discussed students’ ideas and how this investigation attempts to explore students’ alternate ideas. To investigate students’ ideas, this study looked at the reasoning they used to justify their responses. In all question variations except given correct, students were asked to select a response and provide reasoning for
their response, with students only providing reasoning on given correct variations. Students selected a response from a drop down menu on the pretest page, and reasoning was entered into the page in a text box. The reasoning students provided was used to learn about their ideas, but was an indirect probe as reasoning is students’ expression of their ideas, not the ideas themselves.

It was assumed that students’ reasoning was an accurate representation of at least part of the ideas they used to select a response. Students who expressed similar types of reasoning were assumed to use similar ideas while different types of reasoning represented different ideas. In this way, asking students to explain their reasoning helped learn about their ideas, and determine if students were using alternate ideas to respond to the variations.

5.5.2 Coding Data

In order to be able to run statistical tests on student responses and reasoning they had to be coded into a set of categories. Most student responses did not need to be coded because they were selected from a drop-down box, but in a few instances students entered their response into a text box, and those responses needed to be coded. Student reasoning was always entered into a text box and had to be coded.

The responses and reasoning were all coded by a single researcher, with the exception of the data from the pretests discussed in Chapters 6 and 7, which were checked for inter-rater reliability. In a few other instances, another researcher and the main coder worked in unison to code particular subsets of the data and engaged in discussions about how the data should be coded. The single coder was familiar with previous research in all content areas in which these questions were asked and was often familiar with previous student responses to the original version of these exact questions.
Responses were coded into categories developed by the single coder and were based on previous research, if available. Content-area-specific codes were developed for each question, and were used across all versions. In addition to the content-area-specific codes, some general codes were developed that were used across all content areas. These codes were developed by the single coder in the same way that content-area-specific codes were developed.

When students submitted neither a response nor reasoning, they were removed from the dataset for that particular question. When students provided a response, but no reasoning, their reasoning was coded as blank. These students were then removed from the analysis of student reasoning, but were kept for the analysis of responses. Blank reasoning also included instances where students started writing something but did not convey any meaningful information.

*No reasoning* codes were given to responses where students did not discuss the given problem or only restated the response they had chosen. Because instances of no reasoning provided potentially useful information, these students were not removed from the dataset.

The code *other* served as a catch-all for types of reasoning used by only a few students on any of the question versions. If a particular type of reasoning was commonly used on any question version, it would not be coded as *other* on any other question version. *Other* also included reasoning that could not be understood by the single coder, but seemed to convey some reasoning while not fitting into any non-*other* coding category.

The goal of the coding scheme was to be able to identify large differences of particular types of reasoning across question versions. Because the types of reasoning coded as *other* were only used by a few students, there would not be large differences in these types of reasoning; therefore, they were not investigated further.
Table 5.5. A contingency table for comparing students’ use of correct reasoning on the original and given correct versions.

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Given Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Reasoning</td>
<td>12 (21%)</td>
<td>24 (48%)</td>
</tr>
<tr>
<td>Other Reasoning</td>
<td>44 (79%)</td>
<td>26 (52%)</td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>50</td>
</tr>
</tbody>
</table>

5.5.3 Differences in a Particular Reasoning or Response Between Question Versions

One of the primary goals of this study was to find instances where particular responses or types of reasoning were used with different prevalence on one of the variations other than on the original version. In order to inform instruction, the data analysis needed to be able to identify instances where a particular, and identifiable, type of reasoning was used more or less prevalently across question versions. Although a $\chi^2$ test across the groups and types of reasoning can find differences in reasoning, it can not indicate what types of reasoning were used differently. To identify the types of reasoning that were used differently, a Fisher’s Exact Test was used.

In order to determine if a difference existed in responses or reasoning on two different question versions, a 2x2 contingency table was created for any particular reasoning that appeared to be used more often on one of the tasks and that might also have a substantial effect size. The 2x2 contingency table was across question type and usage (or not) of a particular type of reasoning or selection of a particular response. As an example, Table 5.5 shows a contingency table for students’ use of correct reasoning on the original version and given correct variation. Because the groups were assumed to be random, students’ responses on previous questions were not included in the analysis.
Table 5.6. A table comparing *increase* responses on different question variations. It compares students responding *increase* or *increases*, to students giving any other response, on the *original* and the *consider increase* variation of the question used in the pilot study, as shown in Figure 4.2.

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Consider Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase</td>
<td>5 (23%)</td>
<td>2 (8%)</td>
</tr>
<tr>
<td>Other</td>
<td>17 (77%)</td>
<td>24 (92%)</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>26</td>
</tr>
</tbody>
</table>

As another example, Table 5.6 shows a contingency table of students’ responses for the *consider increase* variation from the pilot study. In the *original* version, students selected from *increase*, *decrease*, or *remains the same*, while in the *consider increase* variation, students chose from *increases* or *does not increase*; thus, all the non-*increase* response options in either version were labeled as *other* to combine them. *Other* for the *original* version referred to students who chose *decrease* or *remains the same*; for the *consider increase* variation, *other* referred to students who chose *does not increase*.

Although the meaning of the *other* category varied, this variation did not affect the results because only one particular response or reasoning was analyzed at a time.

This analysis method offers two advantages over running a $\chi^2$ test on a contingency table of all responses or types of reasoning. First, as shown in the example in Table 5.6, it was often not possible to condense the table down while keeping all responses or reasoning present. Second, the goal of the study was not to find results that determined a difference existed between responses or reasoning, without also being able to identify what that difference was.

Full sets of contingency tables are presented in Chapters 6 and 7, which discuss two pretest questions from the across-content study in depth. Further details of the subtleties of some forms of the contingency tables will also be discussed there.
The Fisher’s Exact Test with a two-way tail was used to run a statistical analysis of these contingency tables. The $p$ value and odds ratio for the test are reported in each case. In addition, the $\chi^2$ test was used to find the effect size. The $p$ value of the $\chi^2$ test was not reported because the Fisher’s Exact Test found the $p$ value with a higher statistical power.\(^4\)

As is common in this research field, an $\alpha$ of .05 was used for the Fisher’s Exact Test and the $\chi^2$ test, meaning that a $p$ value of less than .05 indicated a statistically significant result.

When reporting effect size, Pearson’s $\phi$ was used; it has a range from $-1$ to $1$. The larger the absolute value of the $\phi$, the larger the effect size, with a minimal result being around a $\phi$ of .1, a typical result being around a $\phi$ of .3, and a substantial result being around a $\phi$ of .5 [57, p. 108]. This measure of effect size was chosen because it is a standard way to represent the effect size from a $\chi^2$ test.

The odds ratio in addition to effect size was also reported as a more accessible measure. The odds ratio is the odds of one population doing something divided by the odds of a different population doing that same thing. If $p_1$ is the proportion of people in population 1 using a certain type of reasoning and $p_2$ is the proportion of people in population 2 using that same type of reasoning, the odds ratio can be calculated as shown in Equation 5.1.

\[
\text{odds ratio} = \frac{p_1/(1 - p_1)}{p_2/(1 - p_2)}
\]  

(5.1)

An odds ratio of 1 corresponds to no effect, a odds ratio above 1 means the type of reasoning is more common in population 1, and a odds ratio between 1 and 0 means the that type of reasoning is more common in population 2. This study selected $p_1$ and $p_2$ so the odds ratio was always greater than one for consistency.

\(^4\)Both the Fisher’s Exact Test and the $\chi^2$ test were run in the statistics program R with the “fisher.test” function in the “stats” package and the “assocstats” function in the “VCD” package.
Further examples will be presented in detail in Chapters 6 and 7, which discuss the two implementations of the question variations in detail.
CHAPTER 6
ACROSS-CONTENT STUDY EXAMPLE 1 - WORK-BY-WALL
QUESTION

This chapter, along with Chapter 7, discusses two administrations of question variations from the across-content study. The general results of the across-content study are discussed in Chapter 9.

This chapter is intended to serve several purposes. It demonstrates how asking multiple variations of a question provides insights into students’ reasoning and ideas about a particular content area, and it provides an example of how variations of an existing question were designed, the students’ responses were coded and analyzed, and conclusions drawn from the data.

This chapter starts with a review of previous research on students’ understanding of work. It then presents the variations of a previously used research task that were developed and administered, followed by a discussion of the analysis and results. Lastly, it discusses the importance of these results and implications for instruction and future research.

The results indicate more prevalent usage of correct reasoning on the given correct variation, evidence that students are conflicted about the correct response on the eliminate variation, and two types of reasoning not seen in prior research.

6.1 Previous Research on Students’ Understanding of Work

This section presents findings from previous research on students’ understanding of work. The concept of work is directly related to that of energy, which is an overarching and emphasized part of the curricula in physics and other scientific disciplines. Energy has also been labeled a cross-cutting concept and core idea in new science standards in the United States [58]. Energy is becoming an increasingly important area of science education due to the increased public attention to issues
such as energy consumption and global warming. The increased importance of
assisting students in understanding energy places an increased importance on
helping them understand work, which is key to understanding how energy is
transferred. Work is also used extensively in both the undergraduate- and
graduate-level physics curricula in courses such as thermodynamics and quantum
mechanics.

Extensive research has been conducted on students’ understanding of work and
energy, but this discussion focuses on a few results that center around students’
difficulties using the definition of work to determine the amount of work being done.
This subset of the broader literature on students’ understanding of work relates
directly to the results discussed in this chapter and Chapter 7. This literature
informed this investigation of students’ understanding of work and provided a
developed research task for which to make variations. Previous researchers’
identifications of student difficulties were used to assist in coding the data gathered.

6.1.1 Students’ Difficulties with the Definition of Work

Previous studies identified several difficulties that students have in applying the
definition of work \( W = \int \vec{F} \cdot \Delta \vec{x} \). There are three critical elements students need
to be able to accomplish to find the amount of work done by a force: They must be
able to find the displacement of the point of application of the force, find the
magnitude and direction of the force, and quantitatively or qualitatively find the
dot product of those two vectors.

6.1.1.1 Inappropriate Use of an External Coordinate System

One difficulty students have when applying the definition of work is keeping the
sign of work independent from the coordinate system. This difficulty was observed
by both Loverude et al. [24] and Lindsey et al. [59].
Loverude et al. asked written questions about the work being done on a block and on a hand as the hand pushed the block up an inclined plane. Their findings showed that some students having this difficulty only used the sign of the applied force, displacement, velocity, or acceleration to determine the sign of the work. Meanwhile, other students treated work as a vector quantity.

The same coordinate system-dependent reasoning was also observed by Lindsey et al. in their research [59]. Lindsey et al. asked about the net work being done on two blocks being pushed toward each other with the same force exerted on them and for the same distance. The two blocks do not interact as they are pushed toward one another. Lindsey et al. found that students responding to this question used an external coordinate system to say the net work done on the two-block system was zero because the magnitude of the work done on each block was the same, but the work done on one block was positive while the work done on the other block was negative. This is incorrect reasoning, as each force is exerted in the same direction as the displacement of the block the force is being exerted on, which makes the work done on each block positive and the net work positive.

6.1.1.2 Failing to Use the Appropriate Displacement

Another difficulty related to applying the definition of work was a “failure to consider the displacement of the point at which the force is applied.” [59, p. 1003] Research by Lindsey et al. found that, when asked the question about the two blocks being pushed together without interacting, some students considered different displacements, such as the displacement of the center of mass of the system instead of the displacement of each block when finding the work. This difficulty was found in 5% to 10% of their students responding to a few different questions.

Another question used to identify students’ difficulties related to using the appropriate displacement was the Work-by-Wall question shown in Figure 6.1,
which asks about the work performed by a wall on a system consisting of a block and spring [60]. The correct response to this question is that the wall does no work on the spring block system because the wall does not exert a force over a distance; thus, the displacement in this case is zero, and no work is done. Lindsey found that many students (> 75%) did not mention the lack of displacement of the wall as being important in this question with most students reasoning about the direction of the forces instead. [60].

Some other authors have noted additional difficulties students have with using displacement correctly in the definition of work. Sherwood discussed the difficulty of dealing with systems that are not point particles when finding work, pointing out that these concepts are difficult even for senior faculty members and graduate teaching assistants [61]. Furthermore, based on informal observations of his students, Jewett [62] echoed concerns about students having this difficulty and called for an emphasis on the displacement in the definition of work always being defined as the displacement of the “point of application of the force” [62, p. 39].

6.1.1.3 Distinguishing Between a Force and Work

One last difficulty applying the definition of work was identified by Singh and Rosengrant [63], who found that students often have difficulty distinguishing between a force, and the work done by a force. They also found that students often invoke a colloquial understanding of work when answering questions about work, saying that they do not know how to use the definition of work, but they do know the correct response from past experience. For example, “it would be easier to pull which implies less work for you” [63, p. 610].

6.1.2 Difficulties with Zero Displacement Forces and Systems

In addition to having difficulties using the definition of work, students have been found to have difficulties figuring out the work done by forces that are not exerted
A block of mass $m$ is attached to an ideal (massless) spring with coefficient $k$, as shown below. The friction between the block and the surface is negligible. Initially, the block is pulled to point $P$, stretching the spring. At time $t = t_0$, the block is released from rest. When the block is released, it moves back toward point $Q$, where the spring is at its equilibrium length. Consider System A, consisting of the block and the spring.

Does the wall do positive, negative, or zero work on System A? Explain your reasoning.

Figure 6.1. A question used by Lindsey to assess students’ understanding of work and systems [3, 60]. This is also the original version of the Work-by-Wall question discussed in this section.

over a distance. These difficulties are different from those discussed in Section 6.1.1.2 because they are difficulties in which students do not appear to be using the definition of work as part of their reasoning.

By using the Work-by-Wall Question, Lindsey found that students sometimes think it is impossible for a wall to do work on a system because the wall is not the cause of the movement [60]. This incomplete reasoning does not satisfactorily explain why the wall does not do work.

Arons [64] noted that the work-energy theorem can often create misconceptions about forces that do no work when they act on deformable bodies. He observed that, in cases like a person jumping, the work-energy theorem implies that—because the kinetic energy of the person is changing—work must be performed on the person; however, because the floor does not exert a force over a displacement, it does not do any work.
Sherwood also noted that students have difficulties with zero displacement forces doing no work, such as when a rock climber climbs up a vertical cliff [61]. Students often think that the climber does work on the cliff and vice versa, but there is no force exerted over a distance by the climber or cliff in this case because the points where the climber and cliff are in contact remain stationary.

6.2 Designing Variations of a Research Task

In order to build upon previous research on students’ understanding of work, the current study developed question variations of an existing research task. The task modified was the Work-by-Wall question (Figure 6.1) developed by Lindsey [60].

This particular question was chosen for several reasons. First, this question is a useful tool for learning about students’ understanding of work, as demonstrated in the previous section. The question has been vetted and refined by previous research, and existing results are available for consideration. Lastly, it is a question with three choices and no obviously wrong distractors, which is useful because having no obviously wrong distractors helps prevent students from eliminating the same response on an eliminate variation.

6.2.1 Previous Responses to the Work-By-Wall Question

The results of previous administrations of this question were used to develop question variations in the current study. To demonstrate how these results were useful, the results are presented here and discussed in Section 6.2.2 to demonstrate how they influenced the design of the variations.

A summary of responses from 399 undergraduate students at the University of Washington (UW) is shown in Tables 6.1 and 6.2 [60]. These data were collected before work was introduced in lecture, but after a lab on work and energy. With only 20% of students providing correct and complete reasoning, it is clear from these
Table 6.1. Results of Lindsey’s administration of the Work-by-Wall question at the University of Washington (N = 339) [60]. Reasoning percentages are of all students answering the question, not just those responding zero.

<table>
<thead>
<tr>
<th>Response</th>
<th>Percentage of UW Students Providing Each Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>35%</td>
</tr>
<tr>
<td>Negative</td>
<td>25%</td>
</tr>
<tr>
<td>Zero</td>
<td>35%</td>
</tr>
</tbody>
</table>

Table 6.2. Results of Lindsey’s administration of the Work-by-Wall question at the University of Washington (N = 339) [60]. Reasoning percentages are of all students answering the question, not just those responding zero.

<table>
<thead>
<tr>
<th>Reasoning for Zero</th>
<th>Percentage of UW Students Providing Each Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct &amp; Complete</td>
<td>20%</td>
</tr>
<tr>
<td>Wall is Not Cause of Movement</td>
<td>5%</td>
</tr>
</tbody>
</table>

results that this question is difficult for students. Lindsey also showed that this is a difficult question for TAs, with only 40% of them responding correctly and only 30% responding correctly with correct reasoning [60].

6.2.2 Creating Variations

The original version of the question was used as one of the versions in the current study in order to be able to compare the results of the variations created to an unmodified version of the question. This original version of the Work-by-Wall question, shown in Figure 6.1, was developed by Lindsey and published as part of the Tutorials in Introductory Physics [3].

Given Correct was selected as one variation to use because students often responded incorrectly. Previous results (Tables 6.1 and 6.2) show that only 35% of students chose the correct response and only 20% chose the correct response and provided correct reasoning. The difficulty of this question provides an opportunity to probe the extent to which students can justify a correct response when it is
provided to them, especially in a case in which few students give themselves the opportunity to provide such justification. This question variation was also expected to prevent students from expressing the common difficulties discussed in Section 6.1.

To create the given correct variation of the question, the question statement was rephrased as described in Section 5.3.1. The given correct variation is shown in Figure 6.2.

Lastly, an eliminate variation was created because the results in Tables 6.1 and 6.2 showed that both incorrect responses were chosen by at least a quarter of the class, indicating that this question had no obviously wrong distractors. Obviously wrong distractors are responses that are easy to identify as an incorrect response, and are not commonly chosen as the correct response. Obviously wrong distractors do not work well with eliminate variations because the majority of students select the obviously wrong distractor as the best response to eliminate, and very few people on the original version select it as correct, thereby indicating that students think that this particular response is the least correct. Because questions often have at least one obviously wrong distractor, this was a good opportunity for an eliminate variation.

Designing the eliminate variation involved making more drastic changes to the question than creating the given correct variation did. Using the general guidelines described in Section 5.3.2, the eliminate variation shown in Figure 6.3 was created.

A consider variation was not administered in order to limit the number of variations being asked and to increase the number of students responding to each version. In addition, there was no particularly strong reason to create a consider variation of this question.
A block of mass \( m \) is attached to an ideal (massless) spring with coefficient \( k \), as shown below. The friction between the block and the surface is negligible. Initially, the block is pulled to point \( P \), stretching the spring. At time \( t = t_0 \), the block is released from rest. When the block is released, it moves back toward point \( Q \), where the spring is at its equilibrium length. Consider System A, consisting of the block and the spring.

The wall does zero work on System A. Explain.

Figure 6.2. The given correct variation of the question. Students only have one task to complete, to explain their reasoning. The description of the physical situation is exactly the same as it was in the original version.

6.2.3 Design of the Whole Pretest

All versions of this question were asked as part of the pretest for the 9th week of class. This pretest had two questions with multiple variations on it. The layout of the pretest is shown in Figure 6.3. The class was split into three groups, and both Questions 2 and 7 had multiple versions. The other question with variations came first, as Question 2 on the pretest; the Work-by-Wall question was Question 7.

Question 2 on this pretest will be discussed in Chapter 7.

Because this is the second varied question on the pretest, the focus was on evidence indicating that the groups remained the same.

Several points of evidence indicated that the groups were still equivalent despite receiving different versions of Question 2. When looking for differences in groups, as described in Appendix D, the average \( p \) value for all questions between 2 and 7 was
A block of mass \( m \) is attached to an ideal (massless) spring with coefficient \( k \), as shown below. The friction between the block and the surface is negligible. Initially, the block is pulled to point \( P \), stretching the spring. At time \( t = t_0 \), the block is released from rest. When the block is released, it moves back toward point \( Q \), where the spring is at its equilibrium length. Consider System A, consisting of the block and the spring.

Imagine you are taking an exam with the question shown in the box below. You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate? Explain your reasoning.

Does the wall do work positive, negative, or zero work on System A?

Figure 6.3. The eliminate variation of the question.

Table 6.3. Layout of Pretest 9, which had three groups and two questions with multiple versions.

<table>
<thead>
<tr>
<th>Pretest 9</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>Only One Version</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 2 - ‘Change’ in Kinetic Energy</td>
<td>Original</td>
<td>Given Correct</td>
<td>Consider Equal To</td>
</tr>
<tr>
<td>Question 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 7 - Work-By-Wall</td>
<td>Original</td>
<td>Given Correct</td>
<td>Eliminate</td>
</tr>
</tbody>
</table>
Of these questions, the one most similar in content to the Work-by-Wall question was Question 5, which also asked about work done on a system. The $p$ value for the comparison of groups answering Question 5 was .81. Furthermore, none of Questions 3 to 6 had statistically significant differences in reasoning between the groups. All of these indications suggested that the groups were still similar, especially when it came to their ideas about work being done on a system, which were particularly relevant for Question 7.

6.3 Administration

Administration of the pretest was the same as for all other administrations in the across-content study described in Section 5.2. This pretest was administered before the concept of work was introduced in lecture.

6.4 Discussion of the Coding

Student responses did not need to be coded as they selected one of three provided responses, but student reasoning did need to be coded. The coding categories were determined by a single researcher familiar with the literature in this content area. The codes were developed by either using existing categories from the literature or creating new codes for similar types of reasoning.

The application of the categories was then tested on all student responses to determine the inter-rater reliability of the coding scheme. The two coders agreed on 94% of codes, with most disagreement coming from instances where there was not enough information in the student response to place it definitively into one particular coding category. Where possible, disagreements between coders were settled by picking the code that would lead to more conservative $p$ values and effect sizes.

$^1$This $p$ value was found by comparing the groups’ responses on each question using a $\chi^2$ test and then taking the average of the resulting $p$ values.
6.4.1 Content-Specific Reasoning Codes

This section discusses codes for common types of reasoning seen in the data.

6.4.1.1 Coordinate System Dependent

*Coordinate system dependent* reasoning is where students use an external coordinate system to determine the sign of the work. Reasoning where students used an external coordinate system to determine the relative directions of two vectors was not coded as *coordinate system dependent* reasoning because this is analogous to computing the dot product on this question. This type of reasoning often shows up as students describing the positive direction as being to the right, or the negative direction as being to the left, and saying that the force or displacement being in one of these directions indicates the sign of the work. However, we also apply this code to instances where students simply said that one direction was positive or negative but did not link this to a force, displacement, velocity, or acceleration. This type of reasoning is the same type of coordinate system dependent reasoning identified by Loverude et al. [24], although coding responses not referring to the direction of a particular vector includes a wider range of reasoning than they did.

This type of reasoning was used on the original and eliminate versions of the question and was used to justify or eliminate positive and negative.

Examples of Coordinate System Dependent Reasoning

- Original Version - *Negative*
  
  - “I’m assuming the positive x direction is to the right, and it is pushing it to the left.”
  
  - “In the negative direction.”
• Original Version - *Positive*
  
  – “I chose the right to be the positive x-direction.”

• Eliminate Variation - *Negative*
  
  – “Cannot do negative work because its [sic] force is in the positive x direction.”

6.4.1.2 Pushes Back

*Pushes back* reasoning is where students describe the spring or the wall resisting the motion of the block. After the block is released, it will head to the right; at some point the spring will start slowing the block down. At this point students appear to reason that the work must be negative because the wall is pushing back against the block, slowing it down. This type of reasoning may be indicative of confusion between a force and the work done by the force, as noted by Singh and Rosengrant [63].

This type of reasoning was only found on the *original* version of the question and only to justify why the wall does negative work on System A. Examples are shown below.

**Examples of Pushes Back Reasoning**

• Original Version - *Negative*
  
  – “I think it will do negative work as system A is moving in one direction and the wall is pushing in the other direction.”

  – “There is a force that halts the system so it is negative”

\(^2\)In this dissertation [sic] is used to acknowledge incorrect spelling, but not other grammar mistakes such as capitalization, as that would be too distracting.
– “If the force the system exerts is in a positive direction and the wall is pushing back then the force is negative.”

6.4.1.3 Always Positive

Always positive reasoning occurs when students state that work is always a positive quantity. This type of reasoning was used as justification on the original version for why the wall does positive work on System A or for why the work cannot be negative on the eliminate variation. Examples are shown below.

Examples of Always Positive Reasoning

- Original Version - Positive
  – “Work is always positive.”

- Eliminate Variation - Negative
  – “work is a scaler and cant [sic] be negative”
  – “You cannot have a negative amount of work done on a system.”

6.4.1.4 Force Means Work

In force means work reasoning, students equate a force being exerted with a work being done. These students do not mention displacement when finding the work. When using this type of reasoning, students do not specify a sign for the work or reference an external coordinate system, which makes this code distinct from coordinate system dependent.

Some students coded as using coordinate system dependent reasoning may also think that only a force is required for work to be done. This potential overlap will be discussed further in the results.
This type of reasoning was used to justify why the wall does positive work on the original version of the question as well as why the work done by the wall is not zero on the eliminate variation. Examples are shown below.

Examples of Force Means Work Reasoning

- Original Version - Positive
  
  “I am assuming where there is force applied to an object there is also work on the system and rt [sic]”

- Eliminate Variation - Zero
  
  “There is a force applied, so work is a non-zero quantity.”
  
  “system a touches the wall with a compressed spring, causing some force between them.”

6.4.1.5 Displacement Means Work

Displacement means work reasoning is highly similar to force means work reasoning. In displacement means work reasoning, students equate an object being displaced with work being done without explicitly stating there is also a force (or push or pull) exerted.

Again, the lack of reference to an external coordinate system is what keeps this code distinct from coordinate system dependent reasoning, although some students coded as using coordinate system dependent reasoning may think that only a displacement is required for work to be done. The implications of this potential coding overlap will also be discussed as part of the results.

This type of reasoning was only used on the eliminate variation. Examples are shown below.
Examples of Displacement Means Work Reasoning

- Eliminate Variation - *Zero*
  - “The wall caused some displacement to the system therefore it can not have done zero work.”
  - “The block is moving so there is work being done to it”

6.4.1.6 Force and Displacement

*Force and displacement* reasoning is where students argue that because there is a force and a displacement, the work done cannot be zero. Students using this type of reasoning appear to be using the definition of work correctly, but fail to use the appropriate displacement in this instance.

These students are either failing to identify the displacement of the point of application of the force as the displacement that needs to be used to find the work done by the wall or not realizing that this displacement is zero. This is similar to the failure to identify the correct displacement discussed in Section 6.1.1.2.

This type of reasoning was used by very few students on the *original* and *given correct* versions, and used slightly more often on the *eliminate* variation. Examples are shown below.

Examples of Force and Displacement Reasoning

- Eliminate Variation - *Negative*
  - “If the wall does work, it would be pulling on the spring with a force in the positive direction, and the spring would also move in the positive direction. When multiplied together, the work cannot be negative.”
• Eliminate Variation - Zero

  “Because there is a force and a distance, so there is definitely work done to the block.”

6.4.1.7 No Displacement (Correct Reasoning)

With no displacement reasoning, students correctly indicate that, if the point of application of the force does not move, then there is no work done. No displacement reasoning is a correct type of reasoning and is similar to displacement means work reasoning, but instead states that no displacement means no work. This code was given when students did or did not mention the force, but it was not used when students said the force and displacement were both zero or that the net-displacement was zero.

This type of reasoning was the only type of reasoning used by multiple students on all three versions of the question, although the students using it to eliminate a response did so by justifying why zero was the correct response. Examples are shown below.

Examples of No Displacement Reasoning

• Original Version - Zero

  “There is no distance involved in the force that the wall provides”

  “The wall applies a normal force back on the spring, but for no distance and by \( W=F*d \), zero distance gives zero work.”

• Given Correct Variation

  “the distance the wall moves is zero so force x zero=0”

  “The wall does zero work on system A due to the fact that it does not displace the system in any way.”
• Eliminate Variation - *Positive*

  – “The wall doesn’t apply a force over a distance, so it does zero work.”

• Eliminate Variation - *Negative*

  – “Work is defined as force times distance. While the wall applies a force to the system, I do not think it applies this force over a distance. Therefore, I would eliminate negative work as an option.”

### 6.4.1.8 Spring Does the Work

*Spring does the work* reasoning is where students indirectly say the work done by the wall is zero because it is the spring that does the work in this situation. Students here are not explaining why the wall is not doing any work; rather, they are talking about what does do work. This type of reasoning is not necessarily incorrect, but it is incomplete. It is also similar to the reasoning noted by Lindsey where students think a wall cannot be the cause of movement, and related to their research indicating that students have difficulties identifying the correct systems to use when responding to energy questions [60].

This type of reasoning was used on both the *original* and *given correct* versions of the question.

#### Examples of Spring Does the Work Reasoning

• Original Version - *Zero*

  – “the spring is doing the work, not the wall, therefore the amount of work done by the wall is zero.”

  – “zero work because the spring does the work”
• Given Correct Variation
  
  – “Because the spring is doing all the work”

6.4.2 General Coding Categories - No Reasoning, Unclear, and Other

Many responses did not fit into the content area specific codes discussed thus far. These other types of reasoning were assigned into one of the general coding categories described in Section 5.5.2, which offers the general criteria for these codes. Below are examples of each code within this content area.

Examples of No Reasoning Reasoning

• Original Version - Positive
  
  – “It does positive because”

• Original Version - Zero
  
  – “The wall does no work on system A”

• Given Correct Variation
  
  – “I don’t know how work connects to force yet.”

• Eliminate Variation - Zero
  
  – “The wall is doing some sort of work on the system, weather [sic] is positive or negative. This is the answer that should be eliminated first”

  – “I am assuming that the wall does do work on the system, so therefore this question could easily be eliminated from the answers.”
Examples of Other Reasoning

- Original Version - Zero
  - “How does a wall do work? Maybe if it’s a force field”
  - “It doesn’t affect the length of the spring, just how it can compress.”

- Given Correct Variation
  - “it is not acting on the spring or block”
  - “Because it is not doing anything.”

- Eliminate Variation - Zero
  - “Work is done by the system on the wall. this must be opposed by the wall doing work on system A”
  - “i know that every action has an equal reaction”

6.4.3 Coded Responses for All Versions

The data analysis involved counting how many students gave each response and used each type of reasoning and then running statistical tests on those counts to determine how often each response and reasoning was used as well as testing for statistically significant differences in these counts across question versions.

Table 6.4 shows how many students gave each type of response and, of those, how many students used each type of reasoning. The content area-specific types of reasoning that appear on multiple versions or as justifications for multiple responses are color-coded consistently across versions and responses for easier identification.

In some instances, students responding to the given correct variation stated that they disagreed with the provided outcome and justified an alternate outcome instead (see Section 8.5.2 for a more detailed discussion). Because disagreement was sometimes a common phenomenon, results from the given correct version are
separated by students who disagreed with the provided response and those who did not disagree.
Table 6.4. Reasoning for all three versions of the question, with common color coding.

<table>
<thead>
<tr>
<th></th>
<th>Original (N=57)</th>
<th>Eliminate (N=49)</th>
<th>Given Correct (N=43)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Negative</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coor. System</td>
<td>4(7%)</td>
<td>1(2%)</td>
<td>Force &amp; Displacement 1(2%)</td>
</tr>
<tr>
<td>Dependent</td>
<td>4(7%)</td>
<td>5(10%)</td>
<td>No Reasoning 2(5%)</td>
</tr>
<tr>
<td>Pushes Back</td>
<td>1(2%)</td>
<td>4(8%)</td>
<td>No Reasoning 2(5%)</td>
</tr>
<tr>
<td>Force &amp; Displacement</td>
<td>2(4%)</td>
<td>3(6%)</td>
<td>Force Means Work 1(2%)</td>
</tr>
<tr>
<td>Other</td>
<td>6(11%)</td>
<td>1(2%)</td>
<td>Force Means Work 1(2%)</td>
</tr>
<tr>
<td><strong>Positive</strong></td>
<td>18(32%)</td>
<td>15(30%)</td>
<td><strong>Did Not Disagree</strong> 40(93%)</td>
</tr>
<tr>
<td>Coor. System</td>
<td>6(11%)</td>
<td>5(10%)</td>
<td>No Displacement 20(47%)</td>
</tr>
<tr>
<td>Dependent</td>
<td></td>
<td>1(2%)</td>
<td>spring Does the Work 4(9%)</td>
</tr>
<tr>
<td>Always Positive</td>
<td>3(5%)</td>
<td>4(8%)</td>
<td>Other 14(33%)</td>
</tr>
<tr>
<td>Force &amp; Displacement</td>
<td>1(2%)</td>
<td>1(2%)</td>
<td>No Reasoning 2(5%)</td>
</tr>
<tr>
<td>Force Means Work</td>
<td>2(4%)</td>
<td>2(4%)</td>
<td>Other 2(4%)</td>
</tr>
<tr>
<td>Other</td>
<td>5(9%)</td>
<td>1(2%)</td>
<td></td>
</tr>
<tr>
<td>No Reasoning</td>
<td>1(2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Zero</strong></td>
<td>24(42%)</td>
<td>28(56%)</td>
<td></td>
</tr>
<tr>
<td>No Displacement</td>
<td>8(14%)</td>
<td>6(12%)</td>
<td></td>
</tr>
<tr>
<td>Spring Does the Work</td>
<td>6(11%)</td>
<td>8(16%)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>9(16%)</td>
<td>8(16%)</td>
<td></td>
</tr>
<tr>
<td>No Reasoning</td>
<td>1(2%)</td>
<td>3(6%)</td>
<td></td>
</tr>
</tbody>
</table>
6.5 Analysis and Discussion

The analysis of the data indicated some small and large differences in how students used particular types of reasoning to answer each version of the question. One of the goals of this study was to learn more about students’ ideas in many content areas — here focusing on work-related ideas; and although not all types of reasoning evident in previous research using this question emerged in this study, unique types of reasoning and interesting results were identified when comparing responses to different question versions. These novel results helped expand understanding of students’ ideas about work and are the types of results targeted by this study. This section discusses the differences between students’ responses to different question versions and why these differences are important.

In general, our results were comparable to those from undergraduate students at the University of Washington (UW), with 42% of students in this study selecting the correct response (30% at UW), but only 14% providing correct reasoning (20% at UW). This suggests that our results may be generalizable to other student populations.

6.5.1 No Displacement Reasoning on Given Correct

A comparison of the original and given correct versions of the question showed differences in how often students use particular kinds of reasoning. Looking at the reasoning used by these two groups, no displacement reasoning was used more often on the given correct variation. The number of students using no displacement reasoning on each question version, shown in Table 6.5, represents a statistically significant difference.

Although students chose to use no displacement reasoning only 14% of the time, if provided with the correct response, 47% of students were capable of using no displacement reasoning to justify why the work done is zero. Thus, about half of the
Table 6.5. Number and percentage of students using *no displacement*, or any other type of reasoning, in response to the *original* and *given correct* versions. The bottom row shows the total number of students in each group. \[ p \text{ value}=0.000594, \text{ odds ratio}=5.23, \phi=.358 \]

<table>
<thead>
<tr>
<th>All Students</th>
<th>Original</th>
<th>Given Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Displacement</td>
<td>8(14%)</td>
<td>20(47%)</td>
</tr>
<tr>
<td>All Other</td>
<td>49(16%)</td>
<td>23(53%)</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>43</td>
</tr>
</tbody>
</table>

students in the class knew, or could figure out, that the displacement of the wall being zero caused the work done to be zero, but this reasoning did not show up as often when students were asked to select a correct response.

One way students may be able to come up with correct *no displacement* reasoning when provided with the correct outcome is to combine this outcome with the definition of work. If students use the definition of work \( W = \int \vec{F} \cdot \Delta \vec{x} \) and know the work done must be zero, then they may know that either the force or displacement must be zero or the force and displacement must be perpendicular to each other. It is easy to rule out that the force and displacement are perpendicular as this is essentially a one-dimensional problem. Students then only need to look for something that is not exerting a force, such as the wall not exerting a force on the block, or find something that is not moving, such as the wall. The fact that 33% more students referenced the wall on the *given correct* variation indicates that it is a noticeable piece of information, and/or that *no displacement* is an accessible type of reasoning.

There are many possible reasons as to why *no displacement* reasoning was less prevalent on the *original* version than on the *given correct* variation. The *specific difficulties* framework suggests that this result may be a result of particular specific difficulties being elicited by the *original* version of the question, which might have suppressed, or not created a need for, *no displacement* reasoning. By providing the
response zero incorrect ideas may be suppressed, creating an opportunity for
students to use no displacement reasoning. Additionally, informing the students
that the wall does zero work may bring up resources related to zero, such as zero
displacement. It may even further activate multiplicative resources that link some
variable being zero to the final result being zero. Student reasoning about zero is
discussed further in Section 9.6.

The difference in the prevalence of the correct no displacement reasoning
between the original and given correct variations was one of the larger statistically
significant effects in the across-content study. Students' responses to the given
correct variation clearly indicated that almost half of the students were capable of
using no displacement reasoning and related ideas, which was not indicated by the
original version of the question. Thus, no displacement reasoning can be used by a
large portion of the class, and many students have useful ideas to use when
analyzing this physical situation, which can be valuable information for instructors.

But what caused students to not use this reasoning as often on the original
version of the task? Investigating whether no displacement reasoning must be cued,
or if blocking incorrect types of reasoning is sufficient to get students to use no
displacement reasoning would be informative [65].

The increased usage of no displacement reasoning on the given correct variation
is consistent with the work by Heckler and Bogdan [45]. No displacement reasoning
was the reasoning used most commonly to justify the correct response on the
original version of the task, and therefore appears to be the most highly accessibly
type of reasoning for justifying the correct outcome. When the correct outcome is
provided to students no displacement reasoning is the only type of reasoning we see
a larger portion of the class using on the given correct variation than the original
version. This suggests that the reason more students are using no displacement
reasoning is because it is the most accessible type of reasoning that supports the response zero.

It appears no displacement reasoning was available to a large portion of the class, it was not accessible to them on the original version. There are a few different hypothesis we have for why asking a given correct variation of the task makes no displacement reasoning accessible to more students. It is possible that the given correct variation removes, or does not invoke, other types of reasoning that would have been more highly accessible options on the original version, or that contradict the provided outcome; and this causes students to seek out alternate explanatory factors or ideas – such as Heckler and Bogdan showed students might do more commonly when not presented with a highly accessible explanation. We also hypothesize that the additional information in the given correct variation may changes what is accessible to students or the accessibility of certain explanatory factors or ideas.

6.5.2 Zero - Most Eliminated and Most Chosen Response

When comparing the distribution of responses across the original and eliminate versions, as shown in Table 6.6, the distributions of responses do not look very different. This lack of a difference across groups is of interest because students were essentially asked to do opposite things, which would lead to expectations of different responses.

Comparing these question versions shows that the response most often selected as correct is also the response most often eliminated as incorrect, suggesting that the class as a whole seemed to have ideas to support the response zero and reject the response zero. This result is similar to the pilot study results, where the response remains the same was also the most chosen and most eliminated response. However, this result is distinct from the pilot study in that, on this question, the
Table 6.6. Number and percentage of students selecting or eliminating each response on the *original* and *eliminate* versions respectively. Responses where no reasoning was provided were not removed from this dataset.

<table>
<thead>
<tr>
<th>Response</th>
<th>Original Version (%)</th>
<th>Eliminate Variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The wall does negative work on System A.</td>
<td>16 (25%)</td>
<td>16 (31%)</td>
</tr>
<tr>
<td>The wall does positive work on System A.</td>
<td>19 (30%)</td>
<td>7 (13%)</td>
</tr>
<tr>
<td>The wall does zero work on System A.</td>
<td>29 (45%)</td>
<td>29 (56%)</td>
</tr>
</tbody>
</table>

most prevalently selected and eliminated response was the correct response.

However, *remains the same* (zero change) and *zero* both involve ideas related to zero, so they are similar in that regard.

The sum of students choosing or eliminating this response was greater than 100%, but only by 1%. If the sum of these two percentages was clearly greater than 100%, it would be evidence that some students either chose *zero* as correct or eliminate it as incorrect, depending on the question version they received. This would indicate that a single student would reason that *zero* is correct and incorrect and provide very specific evidence of the effect of question phrasing on student reasoning. However, the error in this experiment is greater than 1%, which means the class could potentially have been divided into two groups that either exclusively believe *zero* is the correct response or that *zero* is the best response to eliminate.

Thus, although the class as a whole has ideas about why *zero* is correct or incorrect, it is not clear whether this is the case for any individual student.

This class’s responses to the *original* and *eliminate* versions show that the class was confused about the response *zero*, and within the class there were many ideas to support *zero* being a correct and an incorrect response. There may have been individual students who were confused about this response and could come up with
reasons to both justify and eliminate this response, although no direct evidence exists to support this.

When comparing these results to those of the original version alone, knowing that half of the students chose the correct response as the best one to eliminate paints quite a different picture from simply knowing that half of the students chose it as the correct response.

As Chapter 9 will discuss, having the same response be the most often selected and the most often eliminated response is not a common occurrence.

However, it is common to have responses that are selected as correct as often as they are eliminated as incorrect. These responses, that are chosen as correct as often as they are eliminated as incorrect, we refer to as “conflicted responses” because the class as a whole seemed to have ideas about why they are correct and why they are incorrect. The criteria used to identify conflicted responses was that the $p$ value is greater than .2 for the Fisher’s Exact Test comparing the number of students selecting that particular response compared to the number of students selecting any other response on the original and eliminate versions.\(^3\) Table 6.7 shows the number of students selecting the response zero on the original and eliminate versions and the associated $p$ value.

One reason students may have been conflicted about the response zero is that it may lend itself to being easy to choose and easy to eliminate. In order to determine that the work done is not zero, students needed to be able to justify that some work

\(^3\) Conflicted responses and the criteria used to identify them are discussed further in Section 9.3.1.
Table 6.8. Number and percentage of students giving the response negative or a different response to the original and eliminate versions. \[ p \text{ value} = .54 \]

<table>
<thead>
<tr>
<th>Response</th>
<th>Original</th>
<th>Eliminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>16(25%)</td>
<td>16(31%)</td>
</tr>
<tr>
<td>All Other</td>
<td>48(75%)</td>
<td>36(69%)</td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td>52</td>
</tr>
</tbody>
</table>

is done, but did not need to worry about the sign or directionality, as they might when eliminating positive or negative. Likewise, when justifying zero, students again did not have to worry about the sign or directionality. This, along with a similar result for remains the same in the pilot study, provided some evidence that these non-directional type responses may be easier for students to both justify and eliminate.

As shown in Table 6.8, negative is also a conflicted response, although it was not chosen or eliminated as often as the response zero. This again indicated that as many students had ideas about why this response was correct as had ideas about why this response was incorrect. Negative and positive were chosen equally as often on the original version, but negative was eliminated more often. Thus, positive and negative were not viewed equally by the class even though the original version indicated that the class thought they were equally correct.

In terms of the types of reasoning used to eliminate negative, students used always positive, coordinate system dependent and force and displacement reasoning. It is apparent how always positive reasoning, and the lack of any corresponding always negative reasoning, would have a greater contribution to the elimination of negative than positive; however, it is less clear why coordinate system dependent and force and displacement reasoning would be more commonly used to eliminate negative. One possible explanation is that students more commonly picked coordinate systems that would lead to positive values for force and/or displacement.
6.5.3 Force Means Work and Displacement Means Work Reasoning

This investigation also turned up new types of reasoning that were not seen in previous research on work. *Force means work* and *displacement means work* reasoning both showed up commonly, but only on the *eliminate* version. Only 2 out of 12 uses of *force means work* reasoning occurred on the *original* version. All six uses of *displacement means work* reasoning occurred on the *eliminate* version. This means that these types of reasoning made up 4% (2/57) of responses to the *original* version, and 32% (16/49) responses to the *eliminate* variation. These types of reasoning together indicate that some students had the view that either force or displacement was all that was necessary for work to be done. This incorrect view may have led students to overgeneralize situations in which work was being done.

Although students using *coordinate system dependent* reasoning may be thinking that the only thing required for work to be done is force or displacement, this is not clear from their responses. Students may have used the direction of the force or the displacement to determine the sign of the work, but not to determine if the work done was zero or not. They may have thought that the force determined the sign of the work unless, for example, the displacement was zero. Students may simply have been noting that a force and displacement existed, so therefore some work was done, but only stating that they used the force or displacement to determine the sign of such work.

When presented in the literature [24, 59], researchers’ issue with students using *coordinate system dependent* reasoning was that students were using an external coordinate system where it was not appropriate. In our research, students were clearly reasoning that only a force or displacement alone indicated work being done. This issue is distinct from the issue discussed in previous research; therefore, this result provides additional insights into students’ understanding and information about the types of ideas that need to be addressed via instruction.
It might be suggested that these new types of reasoning are evident because the student population in this study differed from such populations in other studies. However, the administration of the original version of the task in this study did not yield a large number of students expressing force means work or displacement means work reasoning either. Both of these types of reasoning were found in significant numbers only on the eliminate variation of the question, indicating that the eliminate variation—not the student population—is what caused these to be elicited as common types of reasoning.

6.6 Conclusion

The comparison of students’ responses on the different versions of the Work-by-Wall question showed differences, similarities, and unique types of reasoning.

The more prevalent usage of correct reasoning on the given correct variation demonstrated that many students were capable of justifying the correct response, when provided. Knowing that students are able to access or figure out correct reasoning provides more information about students’ ideas and capabilities, which is useful for researchers and instructors.

The lack of differences in responses to the original and eliminate variations offers a better picture of how students view particular responses. By identifying responses where students commonly selected the same response as correct, and the best response to eliminate, it as demonstrated that the class was conflicted about these responses. Although the class clearly had conflicting ideas about some responses, more research is needed to determine if this conflict also exists within individual students. With approximately half of the students responding to the original version selecting zero as correct and approximately half responding to the eliminate variation by eliminating zero, it is unclear if half of each group thought
zero was correct and the other half thought it was incorrect or if individual students thought zero was correct or incorrect depending on the question version asked.

Identifying instances where a class is conflicted about a response being correct or incorrect can provide a great opportunity for peer instruction. When a class is conflicted about a response, they have ideas about why that particular response is correct or incorrect. Asking the class to consider if that response is correct or not could provide an opportunity to elicit these conflicting ideas, which could lead to a productive class discussion. An ability to identify responses students are conflicted about is one example of how using multiple versions of a question can lead to improved formative assessment.

Finally, asking students to eliminate a response and explain why it was incorrect revealed new types of reasoning being used by many students. Knowing that students used force means work and displacement means work reasoning, both of which were somewhat prevalent, led to documenting and understanding these ideas as well as how they are part of an incorrect understanding of work. This furthers research on students’ ideas about work and highlights an area where a tailored instructional intervention may be useful.

Overall, asking students to justify the correct response or explain why a response of their choice was incorrect showed more about the ideas they had or had access to in the context of work.
CHAPTER 7
ACROSS-CONTENT STUDY EXAMPLE 2 - CHANGE IN KINETIC ENERGY QUESTION

This chapter again looks at one particular administration of question variations from the across-content study. As in the previous chapter, this discussion will demonstrate how asking multiple variations of a question provides insights into students’ ideas about a particular content area, and the chapter offers an example of how variations of an existing question were designed, the students’ reasoning was coded and analyzed, and conclusions were drawn from the data. What is not seen in the previous chapter, but is included here, is a discussion of large differences in incorrect reasoning along with the design and analysis of a consider variation. The previous chapter compiled a review of the literature from multiple sources and selected a question from those resources. This chapter discusses a study similar to that performed by Lawson and McDermott [23], using their study and results as the basis for the literature review and question variations.

The results indicate more prevalent usage of incorrect reasoning on the given correct variation along with unique and less prevalent incorrect reasoning on the consider variation.

7.1 Student Understanding of the Work-Energy Theorem

Section 6.1 looked at the literature on students’ understanding of work; the review drew from many sources. That literature is also relevant here, but of particular relevance is the original physics education research publication on the work-energy theorem upon which this study expands.

In 1987, Lawson and McDermott published a study investigating students’ understanding of the work-energy theorem and the impulse-momentum theorem. This section focuses on discussing this one publication regarding the work-energy
Figure 7.1. A diagram of the demonstration used by Lawson and McDermott [23].

Lawson and McDermott used interviews in which they performed a demonstration for the students and then asked them questions about the demonstration. In the demonstration, two dry ice pucks are pushed across a glass table by the air blowing out of a reversed vacuum hose. The vacuum hose is kept a constant distance away from the pucks so that it exerts the same constant force on each puck. The force from the vacuum is applied, in turn, to each puck while they are between two marks on the table. This creates a situation whereby two pucks of unequal mass are pushed for the same distance, with the same constant force, but spend an unequal amount of time between the lines and therefore have the force applied for an unequal amount of time. A diagram of their setup is shown in Figure 7.1. Puck A is the lighter puck, Puck B is the heavier puck.

To investigate students’ understanding of the work-energy theorem, Lawson and McDermott asked students if “the two pucks have the same the same or different kinetic energy during their free motion beyond the [second] line.” The correct reasoning is that the two pucks each have the same force applied to them over the same distance and, therefore, the work done on them is the same. Using the work-energy theorem, this means that the change in kinetic energy of the two pucks
must be the same. As each puck started from rest, their final kinetic energy, after the force has stopped being exerted, will be equal to their change in kinetic energy, which is the same for both pucks. Therefore, the kinetic energies of the two pucks are equal.

Correctly comparing the kinetic energies of the two pucks was difficult for the majority of the interview subjects. As shown in Table 7.1, only 50% of the honors calculus physics students answered correctly with correct reasoning, and 0% of the non-calculus physics students answered correctly with correct reasoning. Both of these student populations consisted of above-average students in their respective courses, and both populations had already been introduced to the concept of work.

Lawson and McDermott commonly observed two types of incorrect responses. *Equal applied force* is a type of reasoning where students state that the change in kinetic energies of the two pucks are equal because the force applied to each puck is the same. They also observed *compensation* reasoning, wherein students either said the mass and velocity balanced out to make the changes in momentum equal or, on the kinetic energy task, said that the velocity had a larger impact than the mass in determining kinetic energy and, therefore, the puck with more velocity had more kinetic energy. When using *compensation* reasoning, students were focusing on a different formula for kinetic energy instead of trying to figure out the work done on the pucks. There was not enough information to solve the problem using only the

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Explanation</th>
<th>Honors Calculus Physics (N=12)</th>
<th>Non-Calculus Physics (N=16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_A = T_B$</td>
<td>Same Work Done</td>
<td>50%</td>
<td>0%</td>
</tr>
<tr>
<td>$T_A &gt; T_B$</td>
<td>Compensation Argument</td>
<td>33%</td>
<td>25%</td>
</tr>
<tr>
<td>$T_A = T_B$</td>
<td>Equal Applied Force</td>
<td>0%</td>
<td>19%</td>
</tr>
<tr>
<td>None Specified</td>
<td>Confused Discussion</td>
<td>17%</td>
<td>56%</td>
</tr>
</tbody>
</table>

Table 7.1. Table from Lawson and McDermott showing the percentage of students giving each particular response and reasoning before an intervention from the interviewer [23].
definition of kinetic energy and, therefore, students were not able to reason correctly about the situation using this method.

Lawson and McDermott ultimately determined that this was a difficult task for students to complete and that students often lacked the conceptual understanding of the work-energy theorem to correctly apply it in this situation. However, one thing that did allow some additional students to come up with the correct reasoning was the interviewer prompting discussion of the starting conditions with them or asking them if they knew what work was.

7.2 Research Task Design

In order to further investigate student understanding in this area, the current study used variations of a question highly similar to Lawson and McDermott’s demonstration. The original version of the question, shown in Figure 7.2, was part of a pretest from *Tutorials in Introductory Physics* [3], which asked students to choose the correct response or indicate that there was not enough information to decide which response was correct. This is a slightly different approach than asking students if the kinetic energies are equal or different, as Lawson and McDermott did.

The key factor from previous research used to design the question variations was that this was a difficult question for students. As with the Work-by-Wall question from Chapter 6, this meant that a given correct variation of this question could be particularly enlightening as students would be provided with a response not many of them would have chosen.

7.2.1 Design of the Variations

This study used three different question versions. As always, the original version was used so that all variations could be compared to it. The two variations used here were the given correct and consider variations.
Two carts, A and B, are initially at rest on a frictionless, horizontal table. The mass of Cart A is less than that of Cart B. The same constant force, $F_0$, is exerted on each cart, in turn, as it travels between the two marks on the table. After they pass the second mark, the carts are allowed to glide freely.

After the carts have passed the second mark, is the kinetic energy of Cart A greater than, less than, or equal to the kinetic energy of Cart B? Explain your reasoning.

Figure 7.2. The original version of the Change in Kinetic Energy question [3]. When selecting a response, students could select from greater than, less than, equal to, or not enough information.

The given correct variation was used because previous research showed that the original version of this question was often answered incorrectly. Thus, asking a given correct variation could show what reasoning students had about the correct response, which was not something seen often in the original version. To create the given correct variation of the question, shown in Figure 7.3, the question statement was simply rephrased, as described in Section 5.3.1.

A consider variation was used instead of an eliminate variation because of the presence of four options instead of three. In more recent versions of this question, its developers have added an option to select “can’t tell from the information given” which we refer to as not enough information. This option was retained in the current study, as researchers often work to improve their research tasks and the latest iteration should be used. It was also assumed that the researchers who developed it added this option because it was a type of response they saw often and
Two carts, A and B, are initially at rest on a frictionless, horizontal table. The mass of Cart A is less than that of Cart B. The same constant force, $F_0$, is exerted on each cart, in turn, as it travels between the two marks on the table. After they pass the second mark, the carts are allowed to glide freely.

After the carts have passed the second mark, the kinetic energy of Cart A is equal to the kinetic energy of Cart B. Explain.

Figure 7.3. The given correct variation of the Change in Kinetic Energy question. Students only have one task to complete, to explain their reasoning. The description of the physical situation is exactly the same as it was in the original version.

wanted to separate that type of response from the others. Having four options, including not enough information, might present students with a response that is easy to eliminate. As this would potentially make an eliminate question uninformative, it would be a good opportunity to ask a consider variation instead.

Designing the consider variation only involved minor modifications to the question statement and options. However, the consider variation targets one response; this study targeted the response equal to as this was the correct response and students had been seen to have difficulty answering correctly. Asking this specific variation could indicate why students thought that the correct response was incorrect, if they did in fact think it was incorrect.

Furthermore, the not enough information response was kept separate from the NOT equal to response to keep this question analogous to the original version.

Students were not often given an option to select not enough information on pretests we administered in this study, so it was made very clear that this was an
Two carts, A and B, are initially at rest on a frictionless, horizontal table. The mass of Cart A is less than that of Cart B. The same constant force, $F_0$, is exerted on each cart, in turn, as it travels between the two marks on the table. After they pass the second mark, the carts are allowed to glide freely.

After the carts have passed the second mark, is the kinetic energy of Cart A equal to the kinetic energy of Cart B? Explain your reasoning.

Figure 7.4. The consider equal to version of the Change in Kinetic Energy question. When selecting a response, students could select from equal to, NOT equal to, or not enough information.

option they were allowed to choose, as was done in the original version. Removing this option on the consider equal to version could prevent students from expressing not enough information reasoning, potentially forcing them to use other reasoning instead of compromising the comparison to the original version.

Still following the general guidelines described in Section 5.3.1, the consider equal to variation shown in Figure 7.4 was created.

7.2.2 Design of the Pretest

All versions of this question were asked as part of the seven-question pretest given for the 9th week of class (Table 7.2). Variations of the Change in Kinetic Energy question were asked as the second question and variations of the Work-by-Wall question discussed in Chapter 6 were the final question on the pretest. Questions 3-6 were used as base questions, but are not discussed further in this dissertation. Question 1’s role as a base question is discussed further in this section.
Table 7.2. Layout of Pretest 9, which had three groups and two questions with multiple versions.

<table>
<thead>
<tr>
<th>Pretest 9</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>Only One Version</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 2 -</td>
<td>Original</td>
<td>Given Correct</td>
<td>Consider Equal To</td>
</tr>
<tr>
<td>Change in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kinetic Energy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 3</td>
<td>Only One Version</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 4</td>
<td>Only One Version</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 5</td>
<td>Only One Version</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 6</td>
<td>Only One Version</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 7 -</td>
<td>Original</td>
<td>Given Correct</td>
<td>Eliminate</td>
</tr>
<tr>
<td>Work-By-Wall</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.3. Number and percentage of students selecting each response. This data is from the question asking students to compare the acceleration of Cart A to the acceleration of Cart B. Greater than is the correct response.

<table>
<thead>
<tr>
<th>Response</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater Than</td>
<td>50(76%)</td>
<td>43(90%)</td>
<td>49(91%)</td>
</tr>
<tr>
<td>Less Than</td>
<td>8(12%)</td>
<td>3(6%)</td>
<td>3(6%)</td>
</tr>
<tr>
<td>Equal To</td>
<td>4(6%)</td>
<td>1(2%)</td>
<td>2(4%)</td>
</tr>
<tr>
<td>Not Enough Info.</td>
<td>4(6%)</td>
<td>1(2%)</td>
<td>0(0%)</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>48</td>
<td>54</td>
</tr>
</tbody>
</table>

Questions 1 and 2 both ask about the two carts being pushed with the same force between two lines. Students’ responses to the first question on this pretest (shown in Table 7.3 and Table 7.4), which asks about the acceleration of the two carts, indicate that the groups responded to this question differently. The difference in students responses to the first question is problematic when seeking to attribute the differences to the question variations, not student populations. However, these groups are discussed as equivalent despite their different responses to the acceleration question for three distinct reasons.
To determine if there are differences between the groups shown in Table 7.3, a $\chi^2$ test is run comparing the number of students giving a correct response and any incorrect response. The incorrect responses were grouped to minimize the error due to small cell values. [$p$ value=0.041, $\phi = .195$]

<table>
<thead>
<tr>
<th>Response</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>50(76%)</td>
<td>49(91%)</td>
<td>49(91%)</td>
</tr>
<tr>
<td>Any Incorrect</td>
<td>16(24%)</td>
<td>5(10%)</td>
<td>5(9%)</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>48</td>
<td>54</td>
</tr>
</tbody>
</table>

- Re-running all statistics using only students who responded to the acceleration question correctly, which gives equivalent groups again, shows all but two of the results discussed in this chapter are still statistically significant. The two results that are not have $p$ values of .06 and .07. This small increase in $p$ value on those results could solely be due to a decrease in statistical power from the reduced N of this selected group.

- Plausible explanations can be offered for why all of the statistically significant results found should be occurring.

- Group 1, who received the original version of question 2, was less correct on the acceleration question than the other groups. However, the results show that Group 2 and Group 3 were more likely to use incorrect or mathematically flawed reasoning on the variations they received.

The first point is the strongest, and the discussion will mention those results that did not hold with this smaller sample at the end of the results section.

7.3 Administration

Administration of the pretest was the same as for all other administrations in the across-content study described in Section 5.2. This pretest was administered before the concepts of work and kinetic energy were introduced in lecture, but after
momentum was introduced and a homework assignment about momentum was completed.

7.4 Discussion of the Coding

As in all of the studies in this dissertation, the reasoning students used was coded into a smaller set of categories for analysis. The coding categories were created by a single researcher familiar with the literature in this content area. The application of the categories was then tested to determine the inter-rater reliability of the coding scheme. After their discussion, the two raters agreed on the coding of 100% of students’ responses.

7.4.1 Content-Specific Reasoning Codes

This section discusses codes for types of reasoning that are common within a single version of this question or that show up on multiple variations.

7.4.1.1 Compensation

Compensation reasoning is a type of reasoning seen in the previous literature by Lawson and McDermott, although this study uses a narrower definition of compensation reasoning than they do. Whereas they considered compensation reasoning to apply to students who argued for responses other than equal to, such as by explaining that the kinetic energy of Cart A is bigger because it has more velocity, this study only codes reasoning that justifies equal to as compensation reasoning. This is done because the types of reasoning used to justify these different responses can be considered distinct from each other.

In the current study’s definition of this type of reasoning, students acknowledged that one car has a larger mass and a smaller velocity (or vice-versa), and concluded that the kinetic energies would be equal to each other because those two factors compensate. In order to be coded as using this type of reasoning the student had to
mention both mass and velocity. This code was also applied to students who
compensated mass and acceleration, which all students using this reasoning on the
original version did.

Students using compensation reasoning were considered to have provided
incorrect or incomplete reasoning, as compensation reasoning does not describe a
mechanism or explain in sufficient detail why the kinetic energies are equal.

This reasoning was used on all three versions of the question to justify why the
kinetic energies of the two carts are equal.

Examples of Compensation Reasoning

• Original Version - Equal To
  
  “even though one has a larger acceleration and the other has a larger
  mass, they will cancel out”

• Given Correct Variation

  “The kinetic energy would be the same because the kinetic energy is
  related to the velocity and the mass of the object so the object with the
  larger mass would have a smaller velocity while the object with the
  smaller mass would have a greater velocity”

• Consider Equal To Variation - Equal To

  “Their respective products of mass and velocity cancel each other out to
  make them equal.”

7.4.1.2 More Velocity

More velocity reasoning argues that the kinetic energies of the carts are unequal
because the velocity of one car is greater or less than that of the other. Lawson and
McDermott considered this type of reasoning as *compensation* reasoning, but the current study separated it out because students used it to claim the kinetic energy of one cart was larger than the kinetic energy of the other. Although velocity was not given directly, based on Lawson and McDermott’s research, and the responses received in this study, students seemed to reason based on velocity rather than mass because it was squared in the kinetic energy equation [23], which led them to believe that it has a greater impact when determining the kinetic energy.

*More velocity* reasoning was used on the *original* and *consider equal to* variations to justify the responses *greater than* and *NOT equal to*.

**Examples of More Velocity Reasoning**

- Original Version - *Greater Than*
  
  “this is because KE=1/2mv^2, and the velocity will be greater since the acceleration is greater so, then the KE will also be greater.”

- Consider Equal To Variation - *Greater Than*
  
  “KE=1/2mv^2 -> A has less mass, greater velocity so more KE”

7.4.1.3 More Mass

*More mass* reasoning is reasoning that argues the kinetic energies of the carts are unequal because one car has more or less mass than the other. This reasoning is very similar to the *more velocity* reasoning except that it uses mass instead of velocity. Such reasoning could have also been considered part of the *compensation* reasoning defined by Lawson and McDermott, although again it is coded it separately here. This reasoning is always used to argue that the kinetic energy of Cart B is greater than that of Cart A, or that their kinetic energies are not equal, because Cart B has more mass. All students coded as using this type of reasoning
talked about one cart being more or less massive than the other rather than simply stating that their masses were not the same.

*More mass* reasoning was used on the *original* and *consider equal to* variations to justify the responses *less than* and *NOT equal to*.

**Examples of More Mass Reasoning**

- Original Version - *Less Than*
  
  “because B has a greater mass and K=1/2mv^2”

- Consider Equal To Variation - *NOT Equal To*
  
  “the cart with more mass (B) will develop more momentum and therefore with likely have a greater [sic] kinetic energy.”

**7.4.1.4 Equal Applied Force**

*Equal applied force* was the other type of reasoning Lawson and McDermott identified. The same categorization for this reasoning was used in the current study. *Equal applied force* is a type of reasoning where students say that, because both carts receive an equal applied force $F_0$, they will end up with the same kinetic energy. This code was applied even if students also stated that the time the carts spend between the two marks was the same in addition to stating that the forces were the same. No students combined their reasoning about the forces being the same with reasoning about the displacements being the same, which is required to provide “complete” correct reasoning.

This reasoning was used on all three versions to justify why the kinetic energies of the carts were the same.
Examples of Equal Applied Force Reasoning

- Original Version - *Equal To*

  - “The kinetic energy of the two carts is equal because they have different masses and different accelerations but since they experience the same force they’ll have equal kinetic energies.”

- Given Correct Variation

  - “Because Carts A and B were both provided the same amount of force, the kinetic (moving) energy they have is the same even though one is going faster than the other. The energy for the movement comes from the force and the force on each was the same, so they have the same energy.”
  - “The same force has been applied so they have the same kinetic energy.”

- Consider Equal To Variation - *Equal To*

  - “The kinetic energy will be the same, because the carts have had the same force applied to them.”

7.4.1.5 **Conservation**

*Conservation* reasoning occurred when students said that the kinetic energies of the two carts were the same because some quantity was being conserved. All but one student claimed it was energy that was conserved; the other student claimed momentum was conserved.

This reasoning was used on all three versions to justify why the kinetic energies of the two carts were the same.
Examples of Conservation Reasoning

- Original Version - Equal To
  - “Conservation of momentum.”

- Given Correct Variation
  - “This must be true due to the conservation laws of energy.”

- Consider Equal To Variation - Equal To
  - “Energy is conserved in a system.”

7.4.1.6 Unequal Elements

Unequal elements is a type of reasoning used to say that the kinetic energies of the carts are not equal because both the velocity and the mass are unequal, but it does not mention which mass or velocity is larger or smaller. It is unclear whether the students believed that both the velocities and the masses would need to be equal for the kinetic energies to be equal or if they believed that either mass or velocity must be equal for the kinetic energy to be equal. Either possibility indicates that students were having mathematical difficulty.

This reasoning was only used to explain why the kinetic energies of the carts were not equal and, therefore, showed up only on the consider equal to variation. All five instances of this type of reasoning, which makes up 9% of the responses to the consider equal to variation, are listed.

Examples of Unequal Elements Reasoning

- Consider Equal To Variation - NOT Equal To
  - “They’re not equal because 1/2mv^2. Cars A and B do not have the same mass nor do they have the same velocity.”
– “Their kinetic energy is not equal because the carts are moving at different speeds and have different masses.”

– “They are not equal in mass or velocity so therefore the kinetic energy must not be equal.”

– “Since both carts have different accelerations, by Newton’s second law, they will have different velocities when they pass the second mark. They also have different masses. By the equation \( \frac{1}{2}mv^2 \), the kinetic energies will not be equal.”

– “No, because K.E. = 1/2 * m(v^2) and they are not equal in mass of [sic] velocity and no ratios are given to us to compare”

7.4.1.7 More Mass or Unequal Elements

In some instances, it could not be determined if a student’s response was consistent with more mass or unequal elements reasoning. Such instances were classified in the more mass or unequal elements category. Not many students were coded as using this type of reasoning. Section 7.5.2 discusses why these responses were kept as a separate category as well as the coding of more mass, more velocity, and unequal elements reasoning in more detail.

7.4.1.8 Need to Know

Need to know was the main type of reasoning students used to justify the response not enough information (NEI).\(^1\) This reasoning was categorized very generally, and this code was used for any reasoning where a student mentioned one or more specific pieces of information they needed to answer this question (e.g., exact masses or velocities, a ratio of the masses, the difference between the masses, and the time each cart spent between the lines). This reasoning is only a

\(^1\)Not enough information is a shorted version of can’t tell from the information given.
justification for the response *can’t tell from the information given*. This type of reasoning was used on the *original* and *consider equal to* versions of the question.

**Examples of Need to Know Reasoning**

- Original Version - *Can’t tell from the information given*
  
  – “K.E = (1/2)m.v^2. None of the velocity is given to calculate the K.E.”
  
  – “kinetic energy is equal to 1/2mv^2. Don’t know mass difference or velocity to determine one larger than *sic* other.”

- Consider Equal To Variation - *Can’t tell from the information given*
  
  – “we would need to know the times.”

**7.4.2 General Coding Categories - No Reasoning, Unclear, and Other**

Many responses did not fit into the content-area-specific codes discussed above. These other types of reasoning were assigned into one of the general coding categories described in Section 5.5.2. Examples of these codes are not presented here. For examples in another content area, see those presented in Section 6.4.2.

**7.4.3 Coded Responses for All Versions**

The data analysis created counts of how many students gave each response and used each type of reasoning in order to determine how often each response and reasoning was used as well as test for statistically significant differences in these counts across question versions.

Table 7.5 shows how many students gave each type of response and, of those, how many students used each reasoning. The content-area-specific types of reasoning that appear on multiple versions or as justifications for multiple responses are color-coded consistently across versions and responses for easy identification.
Coding for the *given correct* version is separated by students who disagreed with the provided response and those who did not disagree.
Table 7.5. Reasoning for all three versions of the question with common color coding.

<table>
<thead>
<tr>
<th>Original (N=66)</th>
<th>Consider Equal To (N=54)</th>
<th>Given Correct (N=48)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Equal To</strong> 25(38%)</td>
<td><strong>Equal To</strong> 24(44%)</td>
<td><strong>Disagreed</strong> 3(6%)</td>
</tr>
<tr>
<td>Compensation</td>
<td>Compensation</td>
<td>More Mass or Unequal Elements</td>
</tr>
<tr>
<td>Equal Applied Force</td>
<td>Equal Applied Force</td>
<td>Other</td>
</tr>
<tr>
<td>Conservation</td>
<td>Conservation</td>
<td>Did Not Disagree</td>
</tr>
<tr>
<td>Other</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>No Reasoning</td>
<td>No Reasoning</td>
<td></td>
</tr>
<tr>
<td>Greater Than 19(29%)</td>
<td>NOT Equal To 20(37%)</td>
<td></td>
</tr>
<tr>
<td>More Velocity</td>
<td>More Velocity</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>More Mass</td>
<td></td>
</tr>
<tr>
<td>No Reasoning</td>
<td>Unequal Elements</td>
<td></td>
</tr>
<tr>
<td>Less Than 12(18%)</td>
<td>More Mass or Unequal Elements</td>
<td></td>
</tr>
<tr>
<td>More Velocity</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>More Mass</td>
<td>No Reasoning</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Need to Know</td>
<td></td>
</tr>
<tr>
<td>No Reasoning</td>
<td>Not Enough Information</td>
<td>Need to Know 5(9%)</td>
</tr>
<tr>
<td>Not Enough Information</td>
<td></td>
<td>Other 1(2%)</td>
</tr>
<tr>
<td>Need to Know</td>
<td></td>
<td>No Reasoning 4(7%)</td>
</tr>
</tbody>
</table>
7.5 Analysis and Discussion

In looking at the results of all three question variations, similar results emerged as those observed in Chapter 6. The results show students using particular types of incorrect reasoning more often on the *given correct* variation, and a unique form of reasoning on the *consider equal to* variation. This section walks through the results, first looking at the *given correct* variation and then looking at the *consider equal to* variation.

7.5.1 More Common Incorrect Reasoning on Given Correct

When comparing responses to the *original* and *given correct* versions of the question, students used *compensation* and *equal applied force* reasoning more prevalently on the *given correct* variation. Both of these are incorrect types of reasoning. Students justifying a correct response were also more likely to use *compensation* reasoning when that response was provided for them, showing that the reasoning they chose to use was influenced by whether they selected, or were provided with, the correct outcome.

7.5.1.1 Equal Applied Force Reasoning on Given Correct

In terms of the relative proportion of students using *equal applied force* reasoning on the *original* and *given correct* versions, students were statistically significantly more likely to use *equal applied force* reasoning on the *given correct* variation. The data comparing the usage of *equal applied force* reasoning on these two question versions are shown in Table 7.6.

This result is similar to the results for the Work-by-Wall question discussed in Chapter 6, but this time the more commonly used reasoning was not correct. Equal applied force reasoning is incomplete because it only mentions the force exerted on
Table 7.6. Number and percentage of students using equal applied force reasoning, or any other type of reasoning, in response to the original and given correct variations. [\( p \text{ value}=0.0305, \text{ odds ratio}=2.96, \phi=.213 \)]

<table>
<thead>
<tr>
<th>All Students</th>
<th>Original</th>
<th>Given Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal Applied Force</td>
<td>8(12%)</td>
<td>14(29%)</td>
</tr>
<tr>
<td>All Other</td>
<td>58(88%)</td>
<td>34(71%)</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>48</td>
</tr>
</tbody>
</table>

the two objects being the same, but does not mention that the displacement of the two objects is the same, which is required for complete correct reasoning.

Students may have used equal applied force reasoning more often because when they were told that the kinetic energies of the two carts were equal, they looked for something done equally to both carts and were able to identify the fact that the force exerted on both carts was the same. This method may be productive though; had these students identified that the displacement was also the same, they would have been coded as providing correct reasoning.

No student said that the kinetic energies were equal solely because the carts were pushed for the same distance, which was likely due to the force equivalence being more salient than the displacement equivalence. We draw this conclusion based on research by Heckler et al. [65], who studied how cue salience can influence responses. In the question, student were told that “The same constant force \( F_0 \), is exerted on each cart, in turn, as it travels between the two marks on the table.” Explicitly stating “the same constant force” without using the phrases “same distance” or “same displacement” could have made the force equivalence more salient. It is possible that more students might have used correct reasoning on the original version, and especially the given correct variation, if the phrasing of the question were changed to be something like “the same constant force \( F_0 \), is exerted on each cart, in turn, as they travel an equal distance between the two marks on the
Table 7.7. Number and percentage of students using compensation reasoning, or any other type of reasoning, in response to the original and given correct variations. \[ p \text{ value}<.0001, \text{odds ratio}=14.2, \phi=.391 \]

<table>
<thead>
<tr>
<th></th>
<th>All Students</th>
<th>Original</th>
<th>Given Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compensation</td>
<td>2(3%)</td>
<td>15(31%)</td>
<td></td>
</tr>
<tr>
<td>All Other</td>
<td>64(97%)</td>
<td>33(69%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>48</td>
<td></td>
</tr>
</tbody>
</table>

This higher salience of the equal displacements for each cart might help students identify that the displacements are the same.

7.5.1.2 Compensation Reasoning on Given Correct

This section discusses the more prevalent usage of compensation reasoning on the given correct variation. This incorrect type of reasoning was more prevalent on the given correct variation than the original and was used as an explanation of the correct response more often when the correct response was provided.

Table 7.7 shows the number of students using compensation reasoning on the original and given correct variations and demonstrates that students who received the given correct variation were more likely to use compensation reasoning.

Furthermore, selecting only the students who chose the correct response on the original version to compare with all the students who received the given correct variation shows that, out of only those students who justified the correct response, compensation reasoning was used more often by the group of students answering the given correct variation than the group of students answering the original variation. This comparison is shown in Table 7.8.

It is important to note the distinction between the results in Tables 7.7 and 7.8. The results in Table 7.7 comparing all students responding to the original version to all students responding to the given correct variation indicate that more students had access to this type of reasoning than the original version indicated. The results
in Table 7.8 comparing only students selecting the correct response to the original version to all students responding to the given correct variation shows that students were more likely justify the correct response with this type of reasoning when the correct response was provided for them.

Both of these results seemed to have the same root cause—namely, the correct response provided by the given correct variation verified that the mass and velocity did in fact compensate to give the two carts equal kinetic energies.

When asked if the kinetic energies were equal, students did not have enough information to determine if the mass and velocity compensated to make the kinetic energies equal without using the work-energy theorem. When provided with the correct response, students had enough information to know that the mass and velocity compensated, thereby verifying that a compensation argument led to the correct response in this situation. Using the outcome as part of the argument to justify the outcome is circular reasoning, and this reasoning is still not complete and correct reasoning, even on the given correct variation.

Although providing students with the correct outcome does not inherently block them from using other types of reasoning to support equal to, it may effectively do this for some students. If students initially think of compensation reasoning as an argument, they might move on to other types of reasoning when compensation reasoning is unclear, but might not when the given correct variation specifies that mass and velocity do compensate. In the same way, students responding to the original version may have used other reasoning or selected other responses because compensation reasoning was incomplete.

7.5.1.3 Accessibility of Reasoning for Given Correct

As in Section 6.5.1, we see these two types of reasoning appear to be available to a large portion of students, but were often not accessible on the original version.
Table 7.8. Number and percentage of students using *compensation* reasoning, or any other type of reasoning, when justifying the correct response *equal to* on the *original* and *given correct* variations. Students disagreeing with the provided response on the *given correct* variation are not considered to be justifying the correct response and are not included in this table. \[ p \text{ value}=0.0201, \text{ odds ratio}=5.63, \phi=.283 \]

<table>
<thead>
<tr>
<th>Correct Only</th>
<th>Original</th>
<th>Given Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compensation</td>
<td>2(8%)</td>
<td>15(33%)</td>
</tr>
<tr>
<td>All Other</td>
<td>23(92%)</td>
<td>30(67%)</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>45</td>
</tr>
</tbody>
</table>

In Section 6.5.1 we saw that the most accessible type of reasoning, as measured by the *original* version, was used much more often on the *given correct* variation. In this instance, we see similar results with *equal applied force* reasoning, but also see a larger increase in the prevalence of *compensation* reasoning that was used by very few students on the *original* version.

This suggests that *compensation* reasoning is influenced by the question version, specifically that it’s accessibility or availability is changed by the provided outcome of the *given correct* variation. Our assertion that the *given correct* variation confirming that the mass and velocity balance in a way that leads to equal kinetic energies allows more students to use this type of reasoning is consistent with the *given correct* variation making *compensation* reasoning more available in addition to making it more accessible.

Further research is needed to determine if a *given correct* variation can influence the availability of student’s ideas, and this is discussed in the conclusion of this dissertation.

### 7.5.1.4 Incorrect Reasoning on Given Correct Conclusion

Comparing the reasoning of the entire class suggests that students can use both *equal applied force* and *compensation* reasoning more often when they are provided with the correct outcome. This suggests students that have access to more ideas
about the physical situation than the results of the original version alone would indicate. In addition, the more common use of compensation reasoning on the given correct variation when looking only at students who justified the correct outcome shows that the reasoning students used to justify the correct response differed when they were provided with the correct outcome.

The difference in the use of compensation is consistent with the fact that the provided response on the given correct variation told students that compensation reasoning was correct in this physical situation, but the original version did not convey this information. The data suggest that compensation may be students’ first line of reasoning, since it supports the given correct response, providing them no motivation to seek out other types of reasoning. With equal applied force reasoning, it is equally clear on both versions that the forces are equal and, when looking at only the justifications of a correct response, a similar percentage of students used equal applied force reasoning on both question versions.

7.5.2 Novel Unequal Elements Reasoning on Consider Equal To

The last result comes from reasoning for incorrect responses to the original and consider equal to versions. The majority of students answering these versions incorrectly used reasoning referencing the two carts having different masses or velocities, as shown in Table 7.9.

This section primarily discusses unequal elements reasoning, which was only given in response to the consider equal to variation; it stated that the carts having different masses and velocities led to them having different kinetic energies.

The key differences between more mass, more velocity, and unequal elements reasoning are shown in Table 7.10. In more mass or more velocity reasoning, students specified whether the masses or velocities were unequal and stated which cart had a lower or higher mass or velocity. In unequal elements reasoning, students
Table 7.9. Reasoning for students providing the incorrect responses greater than, less than, or NOT equal to on the original and consider equal to versions. The percentages are of the students responding greater than, less than, or NOT equal to.

<table>
<thead>
<tr>
<th>Original Version (N=66)</th>
<th>Consider Equal To Variation (N=54)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The kinetic energy of Cart A</strong></td>
<td><strong>The kinetic energy of Cart A</strong></td>
</tr>
<tr>
<td>is greater/less than that of Cart B (N=31)</td>
<td>is NOT equal to that of Cart B (N=20)</td>
</tr>
<tr>
<td>More Velocity</td>
<td>12(39%)</td>
</tr>
<tr>
<td>More Mass</td>
<td>8(26%)</td>
</tr>
<tr>
<td>Other</td>
<td>9(29%)</td>
</tr>
<tr>
<td>No Reasoning</td>
<td>2(6%)</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>No Reasoning</td>
<td></td>
</tr>
</tbody>
</table>

simply stated that the mass and/or velocity was unequal and did not specify which one was greater for which cart. The types of reasoning coded as more mass or unequal elements reasoning are those stating that the mass was different, but not specifying which cart’s mass was greater.²

All of these types of reasoning are incomplete. They all, to some extent, argue that the kinetic energies are not equal because the masses or velocities are not the same for both objects. These types of reasoning are concerning because students are arguing that different masses, different velocities, or different masses and velocities make it so the carts cannot, or do not, have the same kinetic energies. All of these types of reasoning are mathematically flawed, as the product of mass and velocity can end up being equal to each other in any of these conditions. This demonstrates that some students had issues with conceptually understanding physics formulas or mathematical principles.

Two statistically significant differences were observed in the use of these types of reasoning. This section presents these two results together because they are related. One difference is that unequal elements reasoning was used only on the consider

²Although unequal elements reasoning is a different form of reasoning from more mass reasoning, it is not clear which one more mass or unequal elements reasoning is more closely related to.
Table 7.10. Key differences between the types of reasoning used to justify incorrect responses on the original and consider equal to versions.

<table>
<thead>
<tr>
<th>Reasoning Type</th>
<th>Specify if mass or velocity is different</th>
<th>Specify which cart has larger or smaller mass or velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>More Mass</td>
<td>Mass</td>
<td>Yes</td>
</tr>
<tr>
<td>More Velocity</td>
<td>Velocity</td>
<td>Yes</td>
</tr>
<tr>
<td>Unequal Elements</td>
<td>Either/Both</td>
<td>No</td>
</tr>
<tr>
<td>More Mass or Unequal Elements</td>
<td>Mass</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 7.11. Number and percentage of students using unequal elements reasoning, or any other type of reasoning, in response to the original and consider equal to variations. [p value=0.0165, odds ratio=∞, \( \phi = .231 \)]

<table>
<thead>
<tr>
<th>Reasoning Type</th>
<th>Original</th>
<th>Consider Equal To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unequal Elements</td>
<td>0(0%)</td>
<td>5(9%)</td>
</tr>
<tr>
<td>All Other</td>
<td>66(100%)</td>
<td>49(91%)</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>54</td>
</tr>
</tbody>
</table>

equal to version, as shown in Table 7.11. The second is that more velocity reasoning was used less often on the consider equal to variation, as shown in Table 7.12.

Knowing that unequal elements reasoning appears on the consider equal to variation—and only on this variation—is important because it shows that students had unique forms of reasoning that they could use to justify why a response was incorrect. In addition, knowing that unequal elements is a type of reasoning students will use provides more information about how students reason about this physical situation.

Table 7.12. Number and percentage of students using more velocity reasoning, or any other type of reasoning, in response to the original and consider equal to variations. [p value=0.02, odds ratio=5.7, \( \phi = .224 \)]

<table>
<thead>
<tr>
<th>Reasoning Type</th>
<th>Original</th>
<th>Consider Equal To</th>
</tr>
</thead>
<tbody>
<tr>
<td>More Velocity</td>
<td>12(18%)</td>
<td>2(4%)</td>
</tr>
<tr>
<td>All Other</td>
<td>54(82%)</td>
<td>52(96%)</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>54</td>
</tr>
</tbody>
</table>
In this case, this type of reasoning is highly similar to *compensation* reasoning, but instead argues for the opposite response. When using *compensation* reasoning, students argued that, although the mass and velocity were not equal across the two carts, these differences would compensate and lead to the carts having the same kinetic energy. When using *unequal elements* reasoning, students argued that, because the mass and velocity were not equal across the two carts, the carts would not have the same kinetic energy. Both types of reasoning rely on the same fundamental observation—namely that both mass and velocity are different between the two carts—but arrive at opposite results.

*Unequal elements* reasoning also has the same mathematical informality that *compensation* reasoning does. Just because the masses and velocities are unequal does not guarantee that the kinetic energies will be unequal, just as it does not guarantee they will compensate.

Therefore, while *unequal elements* reasoning is unique to the *consider equal to variation*, it also has some distinct similarities to reasoning found on the *original* version.

The difference in prevalence of *more velocity* reasoning suggests that some students who would use *more velocity* reasoning on the *original* version of the question are using *unequal elements* reasoning instead.

Such a result is interesting because *unequal elements* reasoning involves a discussion of both mass and velocity where *more velocity* reasoning only discusses velocity. In addition, there is no similar effect for *more mass* reasoning, indicating that students were not using *unequal elements* reasoning instead of *more mass* reasoning.

Thus, the *consider equal to variation* somehow prompted students to talk about mass in addition to velocity, but did not prompt them to talk about velocity in addition to mass. As the description of the physical situation was the same, the
difference must lie in responding to “is the kinetic energy of Cart A greater than, less than, or equal to the kinetic energy of Cart B?” instead of “is the kinetic energy of Cart A equal to the kinetic energy of Cart B?”

One likely explanation for the difference in more velocity reasoning, without a corresponding difference in more mass reasoning, is that the mass difference was explicitly discussed in the problem statement, but velocity was not. This made it easy for students who figured out that the velocities were different to add that the masses were different, but it was not as trivial to identify that the velocities were different, making it more challenging to use unequal elements reasoning instead of more mass reasoning. Students using more velocity reasoning would not discuss mass even if they were aware of a mass difference because a lower mass would be evidence of a lower kinetic energy and that does not support the claim they would be trying to make. However, when using unequal elements reasoning, students apparently thought the carts had different masses, which provided additional support to their claim that the kinetic energies were different.

Although using different masses and different velocities to support the claim that the kinetic energies are unequal may be a mathematical difficulty for students, as already discussed, it may also have come from students looking for properties of the carts that supported the response they were trying to justify, and not considering the full scenario, similar to students use of local reasoning instead of holistic on the electric circuits question discussed in Section 4.2.2.2. A simpler explanation than the mathematical difficulty is that students were simply ignoring the math and pointing out variables that differed in order to justify the outcome being different. In this case different masses and different velocities were used to argue for different kinetic energies. This pattern also holds for more mass, more velocity, and more mass or unequal elements reasoning. More/less mass was used to argue for more/less kinetic energy, more/less velocity was used to argue for more/less kinetic
energy, and different masses was used to argue for different kinetic energies. These responses are also all indicative of students paying attention to only one variable instead of both mass and velocity when developing their reasoning, which is another way to describe these explanations of student reasoning.

These two differences show a unique type of reasoning being used by students and provide a plausible explanation for why they might have used this reasoning over other types of reasoning used on the original version. They can identify multiple variables that are different and use those to support their claim that the kinetic energies are not the same. Further study is required to distinguish if these types of reasoning rely on flawed mathematical understanding or if students were not thoroughly considering mathematics when developing these justifications.

7.5.3 Interaction with Acceleration Question

Section 7.2.2 mentioned that groups showed statistically significant differences in their responses to the first question on this pretest, which asked about the acceleration of the carts. Selecting only students who answered the first question correctly (90% of all students) we still find that most results discussed thus far hold. The two exceptions are the difference in compensation reasoning among students justifying the correct response on the original and given correct versions and the difference in equal applied force reasoning across all students answering the original and given correct versions.

Despite concerns about the validity of these two results because of the differences in responses on the first question, which may have affected students’ responses to later questions [15], there are reasons to believe the effects identified are valid. Both of these results are very close to being statistically significant when looking only at students that answered the first question correctly, and the loss of statistical significance could be due to a decrease in the number of responses being
considered rather than differences in reasoning. We have also been able to develop plausible explanations for why these results would occur and have presented them throughout this chapter. Finally, equal applied force reasoning should be relatively independent of students’ reasoning about acceleration, although they may potentially be connected.

However, we allow that there is potential for students who answered the first question about acceleration correctly to be able to use that information to express compensation reasoning, and that a difference in the groups responses to the acceleration question may have had an effect on the differences in the usage of compensation reasoning between students justifying the correct response on the given correct and original versions. Therefore, for this one result, we have reasons to both think it is valid, and to doubt it’s validity, and it should be treated as a preliminary finding.

Although none of the reasons are conclusive, they are provided as arguments for why the results discussed in this section should viewed as very likely being real effects.

7.6 Conclusion

Comparing the results of the original version in this study to those of Lawson and McDermott shows many similarities to those found when interviewing non-calculus physics students despite the current sample population being from a calculus-based course. In both studies, students had difficulty answering this question correctly, with no students in the current study being able to provide correct reasoning. The fact that most students have difficulty answering this question correctly is consistent with research by Sabella and Redish [15], who suggested that prior questions about acceleration led students into a local coherence about kinematics (mass and velocity) instead of about work (force and
displacement). Although Lawson and McDermott were able to prompt some non-calculus physics students to use correct reasoning by discussing the initial conditions of the setup, and even more by explicitly asking them if they knew about work, asking students a given correct variation in the current study did not enable students to provide correct work-energy theorem-based reasoning.

One reason for this difference is that students in the current study were given these variations before instruction on work, while the students in Lawson and McDermott’s study were more familiar with the concept of work. Additionally, this difference may have occurred because the Lawson and McDermott intervention of calling attention to the starting conditions is more effective at getting students to arrive at the correct reasoning than the current study’s modification of informing them that the kinetic energies of the two carts are equal. Because of these large differences in student population and prompting, it is unsurprising that we see different levels of student success across these two studies. However, a future study that could provide more insight into the different levels of student success would be to compare two versions of the unmodified question—one pointing out that the displacements are equal for the two carts and one providing students with the correct outcome—to see which question leads to correct reasoning more often.

The results of this question demonstrate that providing students with the correct outcome does not always enable them to arrive at the correct reasoning. This case may have been a particularly difficult one for students to develop correct reasoning on because the proceeding acceleration question may have cued mass and velocity-based reasoning, which was seen to interfere with work and energy-based reasoning in work by Sabella and Redish [15]. Instead, more students inappropriately use compensation and equal applied force reasoning when provided with the correct outcome. Although seeing more prevalent usage of these types of incorrect/incomplete reasoning may be viewed negatively, there are reasons to think
that these are good signs. More prevalent usage of *equal applied force* reasoning suggests that students are productively identifying that the two forces have the same magnitude, which is part of the correct reasoning. Similarly, the more prevalent usage of *compensation* reasoning on the *given correct* variation shows that, although many students are able to use this idea, they often use other reasoning on the *original* version; indicating that students may have some awareness that *compensation* reasoning is not a complete justification for why the kinetic energies are equal.

Furthermore, the way in which students justified the correct outcome between the *original* and *given correct* versions differed. Students were more likely to use *compensation* reasoning when the correct outcome was provided for them, compared to when they selected the correct outcome. This is likely due to the fact that the provided correct outcome let students know that the mass and velocity do compensate in the definition of kinetic energy.

In all reasoning provided for any question version, no student used correct and complete reasoning. This could be due to incorrect/incomplete reasoning being easily accessible or correct reasoning being inaccessible — which is likely given that no instruction on work had yet taken place. However, the use of a *given correct* variation in the current study helps provide an argument for the inaccessibility of correct reasoning being the primary factor, as no student managed to come up with correct reasoning despite being provided with the correct outcome. This contrasts with the example discussed in the previous chapter, where more students were able to arrive at the correct reasoning when provided with the correct outcome. This demonstrates the complexity of students’ thinking and the content dependence of the types of reasoning students use in response to these variations.

As in Chapter 6, this chapter has demonstrated that allowing students to justify why a particular outcome is incorrect leads to reasoning seen on the *original* version.
of the question as well as reasoning that is different from that observed on the original version. When comparing the current findings to those of Lawson and McDermott, some of the same difficulties were identified as well as some other difficulties Lawson and McDermott did not discuss. By finding types of student reasoning they did not discuss and learning more about how many students have access to the types of reasoning they did discuss, the results of this study can expand what is known about students’ understanding of work.

The unique type of reasoning identified herein is unequal elements reasoning, which is closely related to compensation reasoning seen on the original version. Uncovering this type of reasoning helped identify that some students use incorrect mathematical arguments when justifying why the kinetic energies of the two carts are not equal to each other or overlook the mathematics and simply identify different quantities in order justify why the kinetic energies are different. This potential overlooking of the mathematics involved, which was only identified by having students explain why a response was incorrect, may also be present in other types of reasoning, such as more mass or more velocity, which raises new concerns about those incorrect types of reasoning also.
CHAPTER 8
STUDENT RESPONSES TO GIVEN CORRECT VARIATIONS

8.1 Introduction

This chapter discusses students’ responses to the *given correct* variation in detail, with a focus on understanding what responses the *given correct* variation elicits and why. Although providing a similar discussion focusing each variation would be a beneficial addition to this dissertation, only the *given correct* variation is discussed in-depth because the data collected in this study provided insights into how students viewed and interacted with this variation whereas similar data were not received from any other variation.

A common thread to all of these responses is that they are the reactions exhibited in response to anomalous data, such as that from an experiment. Specifically, a few types of responses where students did not engage with the *given correct* variation in the anticipated way will be discussed. For example, students stated that they did not think the information provided to them on the *given correct* variation was correct, or they interpreted the *given correct* variation of the question differently than students answering the original version.

To demonstrate that these reactions appear to be responses to anomalous data, the discussion first examines research on anomalous data in science education and then presents a framework developed by Chinn and Brewer [17] that describes all possible reactions to anomalous data. The discussion then shows that the results from the *given correct* variation align with the framework created by Chinn and Brewer, with students disagreeing, reinterpreting, and changing in response to the *given correct* variation. The features of the questions that led to these different responses are also examined.
Finally, two findings emerged from our analysis of student responses to anomalous data: the number of students disagreeing with a question increases with difficulty and students appear to potentially have a hierarchy for the response types presented in Chinn and Brewer’s framework. These results are discussed explicitly.

8.2 Research on Anomalous Data

Research on the use of anomalous data has been a focus of science education research primarily because of its prevalence as a tool in modern science education methods. Anomalous data are data that conflict with a person’s existing set of beliefs, and many instructional methods use anomalous data to cause students to change their beliefs\(^1\) and move toward a correct understanding. However, students will often not accept anomalous data presented to them in order to maintain their pre-instructional beliefs, which makes understanding students’ reactions to anomalous data an important area of science education research.

Students’ pre-instructional beliefs do not always align with accepted scientific understanding [66], and many modern instructional methods use anomalous data as a way to cause students to change their beliefs to align with accepted understanding. For example, methods based on the theory of conceptual change [67] attempt to get students to modify their beliefs by presenting them with anomalous data to make them dissatisfied with their current set of beliefs and cause them to modify their existing understanding [68, 69]. Other examples include bridging analogies [70], which utilize students’ current understanding to direct them to anomalous data; interactive computer simulations, which present students with anomalous data through experimentation [71]; and group discussion-based methods that have students perform experiments and discuss the often anomalous results they receive [3].

\(^1\)The use of “beliefs” is taken from Chinn and Brewer [17]. This language is used for now as framework-agnostic terminology; theoretical frameworks will be discussed in the next section.
When studying how students respond to anomalous data through a wide variety of instructional methods, it is not uncommon to find cases where students choose to hold on to their pre-instructional beliefs despite having received data that conflict with those beliefs. A few examples of this behavior are students persisting in believing that the earth is flat despite being shown evidence that it is round [72], students measuring a change in mass of steel wool after it is burned by arguing that the scale must be broken or rigged upon seeing no difference [73], and students holding on to pre-Newtonian beliefs about motion even after completing a university physics course [74].

Responses that do not result in a change of belief are witnessed across a wide variety of natural science topics, a variety of instructional techniques, and in both novice and expert scientists [17, 75]. Because these reactions are so prevalent and disruptive to developing a correct understanding, understanding these responses—and what leads students to maintain their beliefs in the face of anomalous data—is of critical importance to science education. To further research the understanding of reactions to anomalous data, Chinn and Brewer [17] developed a theoretical framework that describes the types of responses people have to anomalous data and what factors influence their responses.

The following section describes the consistency between Chinn and Brewer’s theoretical framework and the theoretical framework used in this dissertation and discusses why their framework is appropriate to use for the current study’s dataset. The discussion will then describe their framework in detail and show how many of the responses to the given correct variation in this study fit into their framework, suggesting that students were responding to this question variation as if it were anomalous data.

2 For a more thorough review of the research on anomalous data and examples of students maintaining pre-instructional beliefs, see Chinn and Brewer [17, 75].
8.3 Anomalous Data in the Current Study

Before presenting the framework Chinn and Brewer developed, two issues must be addressed: 1) why this framework is applicable to the current study, which was not developed to study students’ responses to anomalous data, and 2) the compatibility between the theoretical framework of the current study and the framework Chinn and Brewer used.

8.3.1 Why This is an Appropriate Framework

Although this study did not directly attempt to gather data on students’ responses to anomalous data, it is appropriate to apply this framework here because the data collected align very well with the framework (as will be shown) and because the given correct variation presented students with data about a physical situation, which could potentially be anomalous to them.

As discussed in Section 3.3, the given correct variation is the only variation in this study that provided students with the outcome of a physical situation. Providing students with a physical situation and outcome means providing them with data similar to that of a description of an experiment; for some students, these data may conflict with their pre-instruction beliefs and are anomalous. Because previous research has shown that students may not accept the data provided to them through experimentation, readings, or instruction [76, 77], it would be reasonable to see some students treating the given correct variation as anomalous data.

We make the presumption that the provided outcome can be perceived by some students as anomalous data because a provided outcome is something that can conflict with their prior beliefs or ideas. As we will discuss in this chapter, some students directly express that the provided outcome is something they disagree with, clearly showing that these students view the provided outcome as anomalous data. This framework is not applicable to the other question versions administered.
herein because those do not provide students with an outcome; therefore, no elements of those versions are potentially anomalous data.\textsuperscript{3}

It is important to note that not every student receiving the \textit{given correct} variation would perceive the provided outcome as anomalous data. Only those students who failed to develop and accept reasoning to support the provided outcome were presented with anomalous data. In some cases, it was easy to identify which students received anomalous data based on their response; in other cases, estimates were made about the portion of the class that received anomalous data based on other students’ performance when responding to the \textit{original} version of the question. The methods used to determine which or estimate how many students were presented with anomalous data will be discussed alongside the presentation of the data.

For these reasons, this is an appropriate framework to use to look at responses to the \textit{given correct} variation.

\textbf{8.3.2 \ Theoretical Framework Compatibility}

The goal of Chinn and Brewer’s framework is to categorically describe, based on history and previous research, “\textit{how} students respond to contradictory data and \textit{why} they respond as they do” [17, p. 3]. Because they based their framework on previous research across science education, the overarching theory in their framework is kept vague and describes only student theories, ideas, concepts, beliefs, etc., as something that can be altered in response to anomalous data or remain unchanged. In this way, they can incorporate a wide variety of studies that use many different frameworks that meet this simple criteria.

\textsuperscript{3}All other question variations besides \textit{given correct} ask students to select an outcome from two or more possible responses; therefore, on those variations, they are not provided with any data about the situation.
In Chinn and Brewer’s model, an individual holds a “theory”; if that theory is inconsistent with the data being presented, those data are considered anomalous data.

“In idealized form, we conceptualize the situation in which anomalous data occur as follows: An individual currently holds theory A. The individual then encounters anomalous data, data that cannot be explained by theory A.” [17, p. 4]

Although they use the term “theory,” there is nothing in their framework that requires a theory-driven view of student learning. They are simply looking for evidence of a change in students’ thinking caused by anomalous data or a lack thereof. The theoretical framework used in this dissertation is compatible with Chinn and Brewer’s framework because it also allows the anomalous data to cause changes in students’ ideas or to change the ideas they use to respond.

Because the framework of the current study, presented in Section 2.6, is agnostic about the stability of students ideas beyond answering a single physics question, we are open to the possibility that students may either be using stable cognitive structures, such Chinn and Brewer’s “theories,” or may be developing reasoning on-the-fly response to one particular question. Therefore, we discuss possible explanations for our results from either perspective.

Additionally, our framework does not posit that students have theories and we could not look for students changing from one theory to another. Instead looked for evidence of students’ reactions to anomalous data by looking for differences between what students answered on a version that did not provide the outcome (such as the original version) and a question that did potentially provide anomalous data in form of an outcome (the given correct variation). Therefore, will compare the use of particular types of reasoning on the original and given correct versions, as previously done in Section 6.5.1 with results from the Work-by-Wall question in
order to look for evidence that students reasoning was different due to the inclusion of potentially anomalous data. To identify evidence that students did not change their ideas in response to anomalous data, responses where their reasoning did not support the provided outcome will be highlighted.

8.4 Chinn and Brewer’s Theoretical Framework

The framework Chinn and Brewer developed categorizes the types of responses people have to anomalous data, and the factors they find influence the type of response people have. It was developed using historical evidence as well as evidence from previous psychological and science education research.

8.4.1 Responses to Anomalous Data

In their framework, Chinn and Brewer [17, 75] presented eight ways in which students can respond to anomalous data. This is meant to be a complete set of responses that covers all possible reactions. This section presents an overview of these eight categories and then compares the result received to the categories they developed.

(a) *ignore* the anomalous data

(b) *reject* the data

(c) profess *uncertainty* about the validity of the data

(d) *exclude* the data from the domain of theory A

(e) hold the data in *abeyance*

(f) *reinterpret* the data while retaining theory A

(g) reinterpret the data and make *peripheral changes* to theory A, possibly in favor of theory B
(h) accept the data and change theory

Of these eight responses, only the change and peripheral change categories involve accepting the data as correct and changing ideas.

The following sections discuss each response category, provide an example of the evidence used to develop each category, and determine how the current data fit those particular categories.

8.4.2 Conditions Leading to Particular Responses

In addition to describing the categories of responses individuals may have in response to anomalous data, Chinn and Brewer developed a list of characteristics influencing how individuals respond to the anomalous data. Although all of these characteristics could have affected how students chose to respond to given correct variations, discussing how each of these characteristics could play a role in the current results is beyond the scope of this dissertation. Instead, a few elements with relevant data to address or that played a significant role in influencing how students responded to the questions asked herein (the characteristics in bold) will be discussed. These characteristics are:

1. Characteristics of prior knowledge

   (a) **Entrenchment of the prior theory** - Entrenched theories are those “deeply embedded in a network of other beliefs” [17, p. 15]. Entrenched beliefs are more resistant to being replaced by alternate theories.

   (b) Ontological beliefs - “Beliefs about the fundamental categories and properties of the world” [17, p. 17].

   (c) Epistemological commitments - “Beliefs about what scientific knowledge is and what counts as a good scientific theory” [17, p. 17].
(d) Background knowledge - Scientific knowledge that is accepted as valid but is not related to the theory being used.

2. Characteristics of the new theory

(a) **Availability of a plausible alternative theory** - Awareness of an alternate theory to explain the anomalous data.

(b) Quality of the alternative theory - Consists of five characteristics: accurate, explains a broad scope of data, consistent internally and with other theories, simple, and fruitful.

3. Characteristics of the anomalous data

(a) **Credibility** - Aspects of the source of the data that are used to determine if the data are believable or not.

(b) **Ambiguity** - The extent to which data may be reinterpreted.

(c) Multiple Data - Many sources of data that support a single theory.

4. Processing strategies

(a) Deep Processing - “Attending carefully to the contradictory information, attempting to understand the alternative theory, elaborating the relationships between the evidence and competing theories and considering the fullest available range of evidence” [17, p. 29].

Because these selected characteristics each affect multiple response categories, they will be discussed first.

**8.4.2.1 Entrenchment of Ideas**

Chinn and Brewer defined an entrenched theory as one that “...contains one or more deeply entrenched beliefs. An entrenched belief is a belief that is deeply
embedded in a network of other beliefs.” They also postulated several sources of entrenchment, such as evidentiary support, explanatory power across many domains, and compatibility with personal or social goals [17].

In our framework, entrenched theories would be equivalent to ideas that are very stable and pervasive. They could be global specific difficulties that are particularly difficult to address and common for students to express or knowledge structures with strong connections between resources and features of a wide variety of physical situations so as to be frequently activated.

If students have these very stable or global ideas and associate them with the physical situation the question is asking about, they will be more likely to not use ideas consistent with the anomalous data to develop their reasoning as they will use these entrenched ideas instead.

8.4.2.2 Availability of an Alternative Set of Ideas

Chinn and Brewer also proposed that, in order to change theories in response to anomalous data, students must have an available theory capable of explaining the anomalous data.

In the framework employed herein, in order for students to use ideas capable of addressing the anomalous data, they must have or be able to generate such ideas. If they cannot access or generate these ideas, they will likely use ideas that can not account for the anomalous data and develop incomplete or incorrect reasoning.

Students might not be able to use ideas able to account for the anomalous data because they do not have the right associations between the features of this physical situation and a knowledge structure that allows them to make sense of the data; in addition, they might not be able to generate an appropriate knowledge structure at all.
8.4.2.3 Credibility of the Data

In order for the anomalous data to be believed they must have credibility. The credibility of the data presented to students when administering given correct variations came largely from the researcher’s authority as an instructor of a course and professional physicist. Although students are often aware that professors and physicists make mistakes, it can be assumed that they usually take the information provided as correct unless they have a strong reason to believe otherwise. The credibility of the data provided to students should be consistent from question to question, as the data were always presented to the students from the same authority.

8.4.2.4 Ambiguity of the Data

Ambiguity in the data opens up the potential for students to change their interpretation of the data to make them consistent with a wider range of ideas. If students are able to reinterpret the data in a way that makes the data consistent with global difficulties or stable and highly associated knowledge structures, they may choose to use those over ideas that can account for the data without requiring reinterpretation.

8.5 Students Disagreeing with Given Correct

This and the following three sections, discuss relevant response categories developed by Chinn and Brewer, the responses received that fit into those categories, and the results.

This section covers responses where students disagreed with the data provided to them by the given correct variation and shows that the percentage of students who disagreed increased with the difficulty of the question.
8.5.1 Ignore and Reject Response Categories

The first two possible responses to anomalous data discussed here are the *ignore* and *reject* responses. The *ignore* category is used to describe responses where the individual does not accept the data, but also does not address the anomalous data. The *reject* category consists of responses where the individual does not accept the data, but does respond to the anomalous data.

A few historical examples of an *ignore* responses described by Chinn and Brewer are physicists ignoring reports of perpetual motion devices, psychologists ignoring reports of ESP, and biologists ignoring reports of Loch Ness Monster sightings. In all of these cases, the scientists did not accept the data as valid and did not respond to the data to explain why they were not accepting them.

A historical example of a *reject* response based on a perceived methodological error is that of Galileo’s first telescopic observations. When Galileo presented data that were inconsistent with the current Aristotelian view, many astronomers initially rejected the data as an artifact of the telescope.

These types of responses are easy to identify in the current data, as they are the responses where students indicated that they disagreed with the outcome provided. For example, students who were told “the magnitude of the momentum of Cart A is less than the magnitude of the momentum of cart B” may respond by justifying a different response (*ignore*) or saying they did not agree with the response *less than* (*reject*). Two examples of this kind of response are:

- *Ignore or Reject* - When provided with the response *less than* - “since the energies are equal the momentum is equal.”

- *Reject* - “this is a lie. there is not enough information to determine which has the greater momentum. they could possibly be equal”
Although Chinn and Brewer clearly distinguished the *ignore* and *reject* categories, it is difficult to determine if many responses of this nature from the current data fit into the *ignore* category or *reject* category. In the examples shown above, it is clear that the student saying “this is a lie” is rejecting the anomalous data and not simply ignoring them. However, although it seems like the student saying “since the energies are equal the momentum is equal” is simply ignoring the anomalous data, it cannot be ruled out that they may be rejecting the provided outcome by explaining why another response is correct but not explicitly stating the rejection in their response. Because ambiguous responses like this are common, no attempt was made to distinguish between responses that could be categorized as *ignore* or *reject*. Instead all *ignore* or *reject* responses were categorized into a single category: “disagreeing.” Any student who expressed not agreeing with the information provided was coded as disagreeing.\(^4\)

### 8.5.2 Ignore and Reject Responses

Despite knowing that students would sometimes disagree with data that conflict with their current understanding, it was not anticipated that students would reject or ignore the information given to them as part of a *given correct* variation. Disagreement was not expected because the researchers were not aware of any instance in previous research in which students had rejected elements of a question being asked. This trend was expected to continue despite the fact that students received additional information that may be difficult for them to explain. Faced with this conflict, students were expected to use different physics knowledge to answer the question or state that they do not know why this is the outcome that would occur. However, students commonly rejected the outcome provided by the

\(^4\)Responses where individuals directly questioned the validity of the anomalous data they received will be discussed in Section 8.7.
Table 8.1. The percentage of students who disagreed with the information presented as part of the *given correct* variation for all *given correct* variations. The *given correct* variations of all questions are presented in Appendix B.

<table>
<thead>
<tr>
<th>Pretest</th>
<th>% Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest 3: Pendulum</td>
<td>2%</td>
</tr>
<tr>
<td>Pretest 4: Car-Over-Hill</td>
<td>22%</td>
</tr>
<tr>
<td>Pretest 6: Gymnast</td>
<td>3%</td>
</tr>
<tr>
<td>Pretest 8: 2-D Change in Momentum</td>
<td>4%</td>
</tr>
<tr>
<td>Pretest 9a: Change in Kinetic Energy</td>
<td>6%</td>
</tr>
<tr>
<td>Pretest 9b: Work-by-Wall</td>
<td>7%</td>
</tr>
<tr>
<td>Pretest 10: Change in Momentum</td>
<td>24%</td>
</tr>
<tr>
<td>Pretest 11: Pivoting Rod</td>
<td>0%</td>
</tr>
<tr>
<td>Pretest 12: Rotating Disk/Wheel</td>
<td>20%</td>
</tr>
<tr>
<td>Pretest 13: SHO</td>
<td>0%</td>
</tr>
</tbody>
</table>

*given correct* variation and instead used ideas that did not account for the provided outcome.

The prevalence of disagreement was surprisingly high. Within the *given correct* data, more than 20% of students disagreed with the provided outcome on three questions, between 2% and 7% of students disagreed on four questions, and two questions had no students disagree. The percentage of students categorized as disagreeing on each question is shown in Table 8.1.

The rest of this section briefly discusses each of the three questions that had greater than or equal to 20% of students disagreeing with the outcome provided by the *given correct* variation.

### 8.5.2.1 Pretest 4: Car-Over-Hill Question

For the Car-Over-Hill question shown in Figure 8.1, 22% of students disagreed with the provided outcome that stated the gravitational force exerted by the earth on the car is greater than the magnitude of the force exerted by the hill on the car. The provided outcome is correct because, as the car rounds the hill, it will need a downward acceleration to continue along the arc of the hill, meaning that the
A car drives over a hill as shown above.

The magnitude of the gravitational force exerted by the earth on the car is greater than the magnitude of the force exerted by the hill on the car. Explain.

Figure 8.1. The given correct variation of the Car-Over-Hill question administered as part of Pretest 4.

downward force must have a larger magnitude than the upward force. Students who disagreed with this outcome usually stated that the forces were equal because the car was neither floating upwards into the air nor sinking into the hill, which would be the result of a vertical acceleration due to unbalanced forces. (Student responses to other versions of this question are discussed further in Chapter 9.)

8.5.2.2 Pretest 10: Change in Energy and Momentum Question

On the change in energy and momentum question shown in Figure 8.2, 24% of students disagreed with the provided outcome that stated the magnitude of the momentum of Cart A is less than the magnitude of the momentum of Cart B. The provided outcome is correct because, although both carts have the same force applied to them over the same distance, Cart A is lighter and accelerates faster, which causes it to spend less time between the two marks than Cart B does. Because Cart B spends more time between the two marks, the force is exerted on Cart B for a longer period of time. The impulse-momentum theorem tells us that applying the same force for a longer period of time will lead to a larger change in momentum. Because both Cart A and Cart B started with zero momentum and
Two carts, A and B, are initially at rest on a frictionless, horizontal table. The mass of Cart A is less than that of Cart B. The same constant force, \( F \), is exerted on each cart, in turn, as it travels between the two marks on the table. The carts are then allowed to glide freely.

![Diagram of carts on a frictionless table](image)

After the carts have passed the second mark the magnitude of the momentum of Cart A is less than the magnitude of the momentum of Cart B. Explain.

Figure 8.2. The given correct variation of the Change in Momentum question from Pretest 10.

Cart B had a larger change in momentum than Cart A, Cart B will have a larger final momentum when it reaches the second mark.

Students who disagreed with this outcome used a variety of types of reasoning to justify the responses equal to or not enough information. Many students did not provide any reasoning as to why they disagreed, but simply stated that they did not agree. The two most common types of reasoning are using compensation reasoning to say that the momenta were equal and arguing that values for mass and velocity needed to be known to determine how the carts’ momenta compared [23]. Both of these types of reasoning were described in more detail in the discussion of the Change in Kinetic Energy question in Chapter 7.

8.5.2.3 Pretest 12: Rotating Disk/Wheel Question

On the Rotating Disk/Wheel question from Pretest 12 shown in Figure 8.3, the results show that 20% of students disagreed with the provided outcome that the angular acceleration of the disk+clay system is greater than the angular acceleration...
An aluminum disk and an iron wheel (with spokes of negligible mass) have the same radius $R$ and mass $M$ as shown below. Each is free to rotate about its own fixed horizontal frictionless axle. Both objects are initially at rest. Identical small lumps of clay are attached to their rims as shown in the figure.

The angular acceleration of the disk+clay system is greater than the angular acceleration of the wheel+clay system. Explain.

Figure 8.3. The given correct variation of the Rotating Disk/Wheel question from Pretest 12.

of the wheel+clay system. The provided outcome is correct because, although the clay lumps exert equal torques on the disk and wheel, the moment of inertia for the disk and wheel are not the same and will therefore have different angular accelerations. Because $\tau = I\alpha$, the system with the lower moment of inertia will have a larger angular acceleration. In this case, the aluminum disk has more mass concentrated closer to the axis of rotation and will therefore have a smaller moment of inertia than the iron wheel and, thus, a larger angular acceleration. Students who disagree with this outcome nearly all use the reasoning that the important features of the wheels are all identical and should lead to the same angular acceleration. This indicates that students did not recognize that the moment of inertia differed for these two wheels, even when they were told that the angular acceleration was not the same.
8.5.3 Correlation Between Difficulty and Disagreement

In addition to finding that students would disagree with information provided to them as part of a question, it was determined that the number of students who rejected the provided outcome appears to be correlated with the difficulty of the question.

A plot of the percentage of students disagreeing with the provided outcome on the given correct variation and the percentage of students answering incorrectly on the original version for each pretest question is shown in Figure 8.4. This plot highlights three groups of data points. Three levels of question difficulty were defined based on the percentage of students who selected an incorrect response on the original version of the question. Hard questions were defined as those where more than $\approx 70\%$ of students answered the original version incorrectly. Medium questions were those in which between $\approx 40\%$ and $\approx 70\%$ of students answered the original version incorrectly. Easy questions were those in which fewer than $\approx 40\%$ of students answered the original version incorrectly. These ranges are rough estimates as the data to define the edges of the clusters exactly were not available.$^5$

This plot also shows an apparent correlation between the percentage of students answering incorrectly on the original version and the percentage of students disagreeing on the given correct variation. Part of the correlation between difficulty and the prevalence of disagreements may be explained by simply accounting for the fact that the easier the question is, the lower the percentage of students we are confronting with an outcome that is different from the one they would presumably select if given the original version instead. For example, if only 20\% of students answer the original version incorrectly, at most 20\% of students would be expected to disagree with the given correct variation because the other 80\% of students are not being presented with information that conflicts with the response they would

$^5$The pilot study data are not used as part of this analysis because that study utilized a different administration method.
Figure 8.4. A plot of the percentage of students who disagree with the provided outcome on the given correct variation versus the percentage of students who answer the original version of the question incorrectly. Hard, medium, and easy question difficulty zones are shown as well.

have chosen. This effect should result in a linear increase in the number of disagreements as the percentage of students answering incorrectly on the original version increases, assuming that the ratio of students who disagree to those who do not when presented with an outcome that differs from the one they would have chosen remains constant.

To attempt to account for this expected increase in the percentage of students disagreeing due to the increased number of students receiving anomalous data, a new variable was created called the rate of disagreement. The rate of disagreement is the percentage of students disagreeing divided by the percentage of students answering incorrectly on the original version of the question. This provides a ratio of the number of students who disagree to the number of students expected to be receiving anomalous data.
Table 8.2. The percentage of students who choose an incorrect response on the original version and disagree with the provided outcome on the given correct variation, and the rate of disagreement, for each given correct variation.

<table>
<thead>
<tr>
<th>Pretest</th>
<th>% Incorrect</th>
<th>% Disagree</th>
<th>Rate Dis.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest 10: Change in Momentum</td>
<td>79%</td>
<td>24%</td>
<td>.304</td>
</tr>
<tr>
<td>Pretest 4: Car-Over-Hill</td>
<td>78%</td>
<td>22%</td>
<td>.282</td>
</tr>
<tr>
<td>Pretest 12: Rotating Disk/Wheel</td>
<td>78%</td>
<td>20%</td>
<td>.256</td>
</tr>
<tr>
<td>Pretest 9b: Work-by-Wall</td>
<td>54%</td>
<td>7%</td>
<td>.13</td>
</tr>
<tr>
<td>Pretest 9a: Change in KE</td>
<td>62%</td>
<td>6%</td>
<td>.097</td>
</tr>
<tr>
<td>Pretest 8: 2-D Cons. of Momentum</td>
<td>46%</td>
<td>4%</td>
<td>.087</td>
</tr>
<tr>
<td>Pretest 6: Gymnast</td>
<td>67%</td>
<td>3%</td>
<td>.045</td>
</tr>
<tr>
<td>Pretest 3: Pendulum</td>
<td>59%</td>
<td>2%</td>
<td>.029</td>
</tr>
<tr>
<td>Pretest 11: Pivoting Rod</td>
<td>24%</td>
<td>0%</td>
<td>0</td>
</tr>
<tr>
<td>Pretest 13: SHO</td>
<td>31%</td>
<td>0%</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 8.2 shows the results of calculating the rate of disagreement for each pretest question. Plotting the rate of disagreement in Figure 8.5 gives a very similar plot to the one shown in Figure 8.4.

This plot clearly indicates that the more difficult questions had a higher rate of disagreement that is not simply due to the more common availability of students who would have selected a different response as correct.

8.5.4 Discussion of Disagreement

Investigating student disagreement on the given correct variation led to two important results. First, in some cases, students commonly disagreed with the information given to them as part of the given correct question statement. Second, student disagreement was more common on more difficult questions.

8.5.4.1 Existence of Disagreement

Students sometimes disagreed with the information provided to them through a given correct variation. This provides insights into how they responded when receiving anomalous data in this setting.
Although it is common for students to initially disagree with the information provided to them through a physics course, it was not expected that so many students would ignore or reject the data provided to them through given correct variations. The fact that students were willing to ignore or reject information as part of their final response to these questions suggests a level of disagreement with the data being presented as well as their lack of willingness or ability to use different ideas in response. In addition, the large number of students who did this across a wide variety of content areas shows how universal these effects can be. In looking at the characteristics that Chinn and Brewer suggested might lead to students having particular responses to anomalous data, it seems that entrenchment of the original set of ideas and availability of an alternate set of ideas are what would be expected to be the most influential here.

Based on these findings, although it is clear students were using ideas that did not account for the anomalous data, it is unclear if they were doing so because those
ideas were entrenched or because there were no other available ideas to access to explain the data. However, as there were a multitude of other ways to respond to anomalous data, as will be shown, for students to disagree with the question statement in this manner there needed to be a high level of entrenchment.

One explanation consistent with these findings is that students stuck to their answers in such a robust fashion because they held misconceptions. Misconceptions are widely held and stable beliefs that early research in science education used to study students’ thinking [78]. The data presented herein are consistent with this model of students’ thinking, with a larger number of students expressing highly entrenched beliefs. The given correct variation may in fact be testing for “misconceptions” or misconceptions-like reasoning by identifying areas in which students have such strongly held beliefs that they challenge the question statement, and where these are also commonly held beliefs. However, future research in this area would need to address other factors that may be causing this result, such as a lack of availability of alternate ideas, which cannot be ruled out here.

In the frameworks of this dissertation, these results could be explained as strongly connected and widely activated resources or knowledge structures or a commonly held and difficult-to-address specific difficulty. Both of these would be seen as misconception-like reasoning.

Students disagreeing with the provided data showed that they did not find them completely credible, yet this does not seem to be a highly influential characteristic because the data have consistent credibility across all questions and disagreement is only common on some questions. However, the fact that students did not accept the data as correct in all cases shows they may have rejected information provided to them through assessment materials and possibly through other instructional materials as well. This adds to the research demonstrating that students may not accept the data provided to them through readings or instruction [76, 77].
Because of the similarities between given correct variations and information provided by reading or instruction, this result demonstrates the importance of presenting alternate (and correct) ways of thinking about physics content so that students have them accessible as well as the importance of not relying solely on authority to convey anomalous data. Alternate theories must be presented in a way that makes those ideas accessible to students and shows how they account for data that might otherwise appear anomalous to help students make sense of data they might otherwise dismiss [79, 80]. If steps are not taken to demonstrate multiple pieces of evidence in support of correct ideas or to use other techniques to reinforce the presentation of the anomalous data, students may disagree with the data received and avoid changing their thinking to accommodate the anomalous data.

8.5.4.2 Correlation with Question Difficulty

The correlation between question difficulty and disagreement is an important finding because it can help explain what makes some questions particularly difficult for students.

Regardless of what is causing students to disagree with the question statement, the correlation between difficulty and disagreement suggests that the same factors influence both of these variables. Knowing this provides a new line of investigation for understanding why students find these questions from the PER literature so challenging.

It appears that entrenched and widely held ideas may be one of the main causes for disagreement; if true, this would suggest that some PER questions are more difficult than others because they address ideas that are strongly entrenched whereas other questions do not. This consistent with the fact that many of the questions we used are research tasks designed to assess student understanding of topics that students have difficulty learning. In addition, entrenchment appears to
be binary in these results, with disagreement occurring with either \( \sim 20\% \) of students or \( \sim 0\% \). In both this study’s framework and the one presented by Chinn and Brewer, there is no specific reason to believe that entrenchment should be binary. Our results could indicate that there may be some particular level of entrenchment that needs to be meet for students to disagree, or that entrenchment does have some binary properties.

However, there are possibilities beyond entrenchment that could potentially explain this result. Lack of available alternate theories could also explain these results, and because Chinn and Brewer’s set of factors that influence students responses to anomalous data is not comprehensive, there may be other explanations for this result beyond their framework. Further research into what exactly causes students to disagree with the anomalous data provided by given correct variations should be further investigated to determine the exact causes.

Although there are reasons to believe this correlation may hold for all questions, our questions were selected from previous research studies, which may have introduced a selection bias. All of the questions used herein developed publishable results about students’ understanding of physics. The questions the field chooses to publish on are typically those for which many students have difficulty learning the material and choosing the correct response, and for which they may have entrenched beliefs. Therefore, the observed trend between difficulty and entrenchment may result from the motivations of physics education research.

### 8.6 Students Reinterpreting Given Correct

This section discusses instances where students reinterpreted the data provided to them by the given correct variation.
8.6.1 Reinterpret Response Category

Chinn and Brewer described the *reinterpret* category as occurring when individuals accept the validity of the data, but resolve the inconsistency between their theory and the anomalous data by changing their interpretation of the data.

An example of this type of response, discussed by Chinn and Brewer, is from the introduction of Alvarez’s impact theory of Cretaceous extinctions. Alvarez identified a band of iridium at the K-T boundary and claimed that it came from a comet impact. Some scientists proposed a different interpretation of these data, saying that the iridium must have seeped down to the K-T boundary through the limestone above and then been prevented from seeping further down. In this case, the scientists accepted the data showing a band of iridium at the K-T boundary, but reinterpreted them to keep their initial set of ideas.

8.6.2 Reinterpret Responses

Although the *reinterpret* response did not show up in large numbers on multiple pretest questions like the *ignore* and *reject* responses did, one pretest question provided very clear evidence that many students reinterpreted the data in order to use ideas that would not have accounted for the anomalous data.

Results from the *original* version of the question are presented here to demonstrate a standard interpretation of the question and then show how the responses to the *given correct* variation demonstrate the question being interpreted differently.

These data come from Pretest 6, which contained a question from Flores et al. that asked about the tension in a rope holding up a gymnast [54]. The *original* version and the *given correct* variation of this question are shown in Figures 8.6 and 8.7, respectively.
A gymnast who weighs 500N is suspended by two ropes as shown. Is the magnitude of the tension in the left rope greater than, less than, or equal to 250N? Explain your reasoning.

Figure 8.6. The original version of the Gymnast question [54].

A gymnast who weighs 500N is suspended by two ropes as shown. The magnitude of the tension in the left rope is greater than 250N. Explain your reasoning.

Figure 8.7. The given correct version of the Gymnast question.

The correct response to the original version is greater than because the vertical component of the tension in each rope must be 250N; because the ropes also have a horizontal component of tension, their total tension must be larger than 250N.

In their study, Flores et al. [54] found that a common incorrect response to the original version of this question is that the tension is equal to 250N. Flores et al. found the response equal to was selected 70% of the time in two of their three administrations, showing close agreement between their results and the current study’s results. The most common reasoning for this response was that the angle of
the ropes is the same, so they must have the same tension, and each rope must support exactly half the gymnast’s weight. This reasoning relies on the symmetry of the situation.

Similar results emerged in the current administration of the original version of the question: 67%\(^6\) of students selected an incorrect response and most commonly used the same symmetry-based incorrect reasoning as Flores et al. observed.

However, these results show a different kind of common reasoning on the given correct variation. Here, when students were told the tension in the left rope was greater than 250N, many students stated that the tension in the left rope was greater because there was something differentiating that rope from the right rope. This asymmetry-based reasoning is in direct contrast with the assumption that the tension in each rope must be equal, which is the intent of the question and something students often relied upon in explaining their reasoning on the original version.

Students cite many sources of asymmetry:

**Length Asymmetry** “The length of the left rope might be shorter or longer than the length of the right rope.”

**Force Asymmetry** “This means that the gymnast is pulling on the left rope with more force then he/she is pulling on the right rope with.”

**Angle Asymmetry** “the angles of his arms are not exactly identical, therefore the 500N force is not split up evenly.”

In total, 16 out of 39 students (41%) responding to the given correct variation used an asymmetry argument, with force asymmetry being the most common type of asymmetry reasoning being, mentioned in more than half the asymmetry responses. No students used asymmetry arguments on the original version.
8.6.3 Discussion of Reinterpreting

These responses suggest that students reasoned using one-dimensional kinematics on both versions of the Gymnast question.

On the original version, the line of reasoning that the tension in the left rope is equal to 250N because the situation is symmetrical and the ropes need to each have a tension equal to half of the gymnast’s weight is essentially an inappropriate treatment of this situation as one-dimensional. Looking at only the vertical components of the tension vectors from each rope, they do need to add to 500N, but students are not addressing the horizontal component of tension in each rope.

When students responded to the given correct variation using asymmetry reasoning, it appears that they were trying to adapt the same one-dimensional kinematics reasoning to explain why the tension is greater than 250N. One-dimensional kinematics reasoning would explain why the tension would be 250N, and students brought in other factors—specific to the left rope—that would cause the tension to increase, thereby breaking the symmetry of the problem. This lack of symmetry allows the left rope to have a higher tension than the right rope while still maintaining a total of 500N between them.

The last element they needed in order to be able to use a asymmetry argument to make one-dimensional kinematics reasoning compatible with the given correct variation was to be able to come up with a reason for the situation to be asymmetrical; students seemed to be able to come up with many types of asymmetry to yield this result.

The fact that students used this asymmetry reasoning over the correct reasoning offers several insights.

This result suggests that ideas related to one-dimensional kinematics are separate from two-dimensional kinematics. If they were closely related, students who largely use one-dimensional reasoning to solve this problem would have been
expected to switch to two-dimensional reasoning when provided with the outcome. Instead, they stuck with one-dimensional kinematics reasoning and interpreted the problem differently to keep that reasoning and the outcome consistent.

Furthermore, two-dimensional kinematics ideas do not appear to be accessible. Students did not display an ability to extend from one-dimensional to two-dimensional kinematics when faced with anomalous data. This could also be described as vector ideas not being accessible, with students only using scalar ideas to reason about the situation.

The anomalous data provided by the given correct variation also clearly demonstrate ambiguity in this case, which leaves room for reinterpretation. Given a choice of response, everyone assumes symmetry, but since it is not explicitly stated and students know many ways to make this situation asymmetrical, they can use the ambiguity here to reinterpret the data. Students had a variety of changes they could suggest that would increase the tension in the rope, which is a good example of students being able to pull in knowledge from outside their physics instruction to respond to this question. Using informal knowledge to make sense of physical situations is something we want students to do, and something they have been seen to not do in several research studies [15, 16].

As reinterpretation did not commonly occur on any other given correct variation, this question is likely to have a high level of ambiguity relative to the other questions sourced from existing PER literature.

The degree of reinterpretation seen here is likely due to the focus on the left rope in the wording of the question, and it would be interesting to see how students respond if the question clarified that the situation is symmetric. Assuming this inhibited the use of asymmetry arguments, there could be numerous possible results. Would students change to using two-dimensional kinematics-based

\footnote{This could be done by asking about the tension in each rope, rather than just the left one; stating that the tension in the two ropes are equal; or explicitly stating that the situation is symmetric.}
reasoning and respond correctly more often? Would they find other ways to reinterpret the question? Would they state disagreement more often?

8.7 Students Excluding, Holding in Abeyance, or Claiming Uncertainty about Given Correct

This section discusses instances in which students presented a variety of similar responses based on the notions that they had not yet learned enough or did not know enough to respond to the *given correct* variation. These responses were very infrequent with less than 5% of students providing these responses, but they are presented to reinforce the argument that students can potentially view the *given correct* variation as anomalous data.

By presenting these responses, we further demonstrate that the responses in our dataset are those that would be expected from students who are interpreting the provided outcome as anomalous data. We also provide a complete picture of all of the types of responses to anomalous data we should expect to see when administering *given correct* variations, according to the framework developed by Chinn and Brewer [17]. Finally, we are able to discuss why these types of responses are not seen as commonly as the other types of responses.

8.7.1 Exclude, Abeyance, and Uncertainty Response Categories

The final three categories of responses to anomalous data that do not involve changing ideas to fit the anomalous data are: to exclude the data from the domain of the set of ideas, to hold the data in abeyance, or to claim to be uncertain about the validity of the data. These three responses are grouped together because they are responses indicating that students did not necessarily disagree with the data, as seen in Section 8.5, but also did not go on to justify why they were correct.
The exclude category consists of responses where the individual claims the data are not part of the set of ideas they have access to. In the current application, this may play out as students thinking that the provided anomalous data were not something that they had learned yet (the data are excluded from the physics content they have been taught so far) or that they are not something that physics explains (the data are chemistry or biology data for example).

The abeyance category consists of responses where individuals acknowledge that the data should be explained by content they know, but that their understanding may not be developed enough to explain this anomalous data.

The uncertainty category consists of responses where students explicitly express doubt about the validity of the data, but neither accept them as valid nor reject them as invalid.

For an example of an exclude response, Chinn and Brewer discussed the search for an explanation of Brownian motion. For many years it was debatable which natural science was responsible for explaining Brownian motion. Here scientists could exclude data supporting Brownian motion by claiming that the data are not something that their field should explain. For example, a physicist in the 1830s may have said that Brownian motion is not something that needs to be consistent with physics theories because there are tiny life forms pushing the molecules around and that makes it something that needs to be explained by biology theories [81].

An example of an abeyance response discussed by Chinn and Brewer is the way astronomers dealt with the observed discrepancies between the orbit of Mercury and Newtonian physics in the late 1800s. When presented with this discrepancy, astronomers accepted the data as correct, but they did not abandon Newtonian mechanics. Instead, they claimed the discrepancy could be settled within Newtonian mechanics, just not at that time.
As the *uncertainty* response was not part of the original classification scheme, the example presented here comes from a follow-up article by Chinn and Brewer [75]. In this article, they presented the example of physicists from Fermilab discussing the discovery of the top quark. They showed that there was clearly some disagreement amongst the Fermilab team responsible for the discovery as to whether they had discovered the top quark or not. At this point in their research, the team at Fermilab could be considered to be holding an *uncertainty* response toward the anomalous data they had taken.

**8.7.2 Exclude, Abeyance, and Uncertainty Responses**

Although a large number of these responses were not observed, their absence is notable as these are responses some students would be expected to express if they are treating the given correct variation as anomalous data. In fact, in this study’s dataset, students had good reason to hold data in abeyance or exclude them. The pretest questions may have required the use of ideas not discussed in class or just recently presented. This makes it reasonable for students to think either that the data could not be explained using their current ideas or that the ideas they had may not have been developed enough to explain the data.

Although no students provided responses that were clear examples of this type of reasoning (e.g., “I do not think this response is consistent with my current understanding of kinematics, but I think that kinematics should be able to explain this result” or “I do not think that I have learned the relevant physics content I need to be able to answer this question”), some responses seemed to fit these categories:

- **Abeyance** - “I can’t tell, maybe someone else can.”
- **Exclusion** - “because of a law that i dont know and havent learned yet...”
- **Exclusion, Abeyance or Uncertain** - “I do not know”
• *Abeyance* - “I don’t know how work connects to force yet”

• *Abeyance or Uncertain* - “I don’t see why it would be that way.”

Because are reasonable responses to pretest questions, and because they are part of Chinn and Brewer’s framework, we find it interesting that they were used so uncommonly in response to the *given correct* variation.

### 8.7.3 Discussion of Exclude, Abeyance, and Uncertainty

The lack of these types of responses is most likely due to the fact that these responses are not valued, as they would probably receive no credit on any sort of a graded question. In addition, although knowing that students do not feel that they understand a topic yet may be useful information for research, questions are often used to engage students in sharing what they think about the content on pretests rather than just stating they do not know the content yet. In fact, researchers and instructors often inform students that they know the students have not learned the content yet, but they want the students to talk about it anyway, which may make such responses unnecessary.

Another related explanation might be that students will choose another type of response over these and only use these responses if they are unable to address the anomalous data in another way.

These types of responses are not necessarily inappropriate responses for students to have, and are sometimes appropriate responses for scientists to have when presented with anomalous data. The fact that students rarely provide these responses could be a cause for concern as students may be having these responses to anomalous data, but not documenting these responses as part of their submission. With *abeyance* in particular, realizing that something is potentially important instead of rejecting it because it is not understood is not an inappropriate response for students to have for anomalous data when they are unable to make sense of
them. If students respond the same way when asked a question by an instructor during instruction, choosing to not express an *abeyance*, *exclude*, or *uncertain* response, it could have potentially harmful effects on the instructor’s ability to teach and on the students’ learning. Further research into students’ responses of this manner and potential impacts of the lack of these types of responses could be useful in improving instruction and learning in science.

8.8 Students Accepting Given Correct

This section examines instances in which students accepted the data and appeared to adjust the ideas they were using in response to the *given correct* variation. Because the data set did not identify individual students having this response, and because this response is so well studied, this part of the framework will only briefly be presented, and the results suggesting that students were reacting to the *given correct* variation in this manner will be reviewed.

8.8.1 Change and Peripheral Change Response Categories

The last two categories of Chinn and Brewer’s framework are those in which students accept the data and make some change to their theories. Chinn and Brewer argued that one of the goals of science instruction is to get students to change their theories to align with the current scientific understanding; in order to achieve this goal, the desired response to anomalous data is for students to change their often incorrect initial theories to ones that are consistent with the data.

The first response where students accept the data and make some change to their ideas occurs when students “reinterpret the data and make *peripheral changes* to theory” [17, p. 4]. With a *peripheral change* response, individuals may continue to hold most of an incorrect set of ideas, but may make small changes to it to reconcile it with the anomalous data. The second response is to “accept the data
and change theory” [17, p. 4]. In both of these types of responses, students accept the data they are given and change or use different ideas to account for the data.

For an example of a peripheral change response, Chinn and Brewer discussed another example from Galileo’s early telescope observations. Through the telescope, mountains on the edges of the moon could be observed, which conflicted with the theory that all celestial objects were perfectly spherical. In response to these anomalous data, one observer agreed that he saw mountains on the moon, but argued that the moon was encased in a transparent crystal sphere. This peripheral change to the theory that all celestial objects are perfect spheres allowed this observer to keep the majority of his initial theory intact.

Examples of the change response are common, but one particular example that Chinn and Brewer highlighted is the chemical revolution. In this revolution, scientists switched from major phlogiston theorists to supporters of Lavoisier’s oxygen theory.

8.8.2 Change and Peripheral Change Responses

The change and peripheral change responses are a phenomenon that occur often and are required for student learning through interaction with anomalous data.

Although the given correct variation was not designed to be a tool to cause a lasting change in students’ thinking, it might have caused students to change the ideas they use to respond and could possibly lead to a lasting change in some cases. As discussed in previous chapters, students presented with a given correct variation will sometimes use reasoning with a different prevalence than students presented with the original version of the question. This result provides evidence that some students change the ideas they use to answer the given correct variation in response to receiving anomalous data.
Chapters 6 & 7 discussed two cases where a particular type of reasoning was used more often by students when responding to the *given correct* variation than to the *original* version. One of these cases was from student responses to the *given correct* and *original* versions of the Work-by-Wall question where students used *no displacement* reasoning more often on the *given correct* variation than on the *original* version.

In the Work-by-Wall question, students were asked if a wall does positive, negative, or zero work on a system consisting of a block and a spring undergoing simple harmonic motion. When told on the *given correct* variation that the wall does zero work on the block-and-spring system, students more often used *no displacement* reasoning to argue that the work done by the wall is zero with 47% of students using this correct type of reasoning on the *given correct* variation compared to only 14% using it on the *original* version.

### 8.8.3 Discussion of Change and Peripheral Change

In these cases, as well as similar responses in other *given correct* variations, some students appeared to use different ideas than they would have when responding to the *original* version. This provides evidence that the anomalous data provided to those students by the *given correct* version of the question causes them to use different ideas than they would have if responding to the *original* version. In this chapter, we consider this use of different ideas between the *given correct* and *original* version as a change in the ideas students are using in response to being presented with anomalous data. However, because individual students’ initial ideas were not identified, it cannot be determined which students perceived the information provided by the *given correct* variation as anomalous data or which students changed their ideas in response to anomalous data presented in the *given correct* variation. However, the differences in the prevalence of particular types of
reasoning between the *given correct* and *original* versions show that some of these students must fit into the *change* or *peripheral change* categories.

Furthermore, the inability to quantify the total number of students changing their ideas in response to the *given correct* variation, rather than just the changes in one particular type of reasoning, prevents the determination of which factors affect students’ choice of this response on a particular question in this study. Identifying changes in student ideas is not something that can currently be measured consistently, and this dissertation presents a method of performing an analogous measurement of differences in student ideas precisely because changes in ideas are very difficult to measure. We suggest that further studies could potentially thoroughly catalog attributes of particular research tasks and the ideas related to them, and be able to draw conclusions about the interactions between tasks, variations and ideas that way.

8.9 Discussion

This section reviews the findings of this chapter as whole and talks about results that carry across multiple parts of the framework.

8.9.1 Students Treat Given Correct as Anomalous Data

The most overarching finding of this chapter is that some students responded to the *given correct* variation as anomalous data. This was an emergent finding in the current research and something not targeted. However, Chinn and Brewer’s framework does an excellent job of describing the identified results, thereby providing substantial evidence that students can treat the *given correct* variation as anomalous data.

Such a result is surprising given the studies and historical evidence Chinn and Brewer relied upon in developing their framework. The majority of evidence they
considered involved in-depth exposure to anomalous data, and often years of scientific experimentation. In contrast, the current study briefly exposed students to a small piece of anomalous data, giving them only a few minutes to consider it before responding. Chinn and Brewer’s framework also covered all of the responses received in this study, and there was no need to append to their framework to include new types of responses. The fact that these two different sets of exposure to anomalous data all led to the same sorts of responses is remarkable and points to the importance of understanding these reactions in as much detail as possible.

It is also interesting that students appeared to, at least temporarily, change or use different ideas than they would have used to answer this question without the outcome provided. It would be interesting to know if these changes are at all lasting modifications of knowledge structures or specific difficulties.

8.9.2 Hierarchy of Response Types

In looking across the students’ responses to all questions, each question had one common type of response to the anomalous data it presented, suggesting that students have a commonly shared hierarchy of responses to anomalous data.

Several pieces of evidence support the existence of this hierarchy. Three questions had large amounts of disagreement and one question had common reinterpret responses, but in no case was there a question where these response types strongly overlapped. In particular, approximately 40% of students reinterpreted the Gymnast question, while only 3% disagreed, despite this being one of the hardest questions, with 2/3 of students answering the original version incorrectly. Because we saw a correlation between question difficulty and disagreement, we would expect a high rate of disagreement on this question, but do not see that, and see a large percentage of reinterpret responses instead. The vast majority of students seemed to choose to reinterpret the Gymnast question instead
of disagreeing with it, suggesting that reinterpret may supersede disagreement in this case. This is again reinforced by the fact that approximately 40% of students reinterpreted the Gymnast question, and the largest percentage of students who disagreed on any question was only half that, suggesting students are potentially more willing to reinterpret than disagree.

Very few students provided abeyance or exclude responses, despite these being appropriate responses for a pretest question. This suggests that students may be responding with other reactions to anomalous data over these ones.

Combining these observations, for students responding to anomalous data from given correct variations, it appears there could potentially be an order to the responses they will consider or use. Our data suggests that this hierarchy is that disagreement is superseded by reinterpret if a reinterpretation is available, and abeyance or exclude are potentially superseded by disagreement. However, our data is a tentative suggestion of this and follow-up studies are needed to determine if a hierarchy actually exits, and how it is structured.

A hierarchy amongst the responses was not part of Chinn and Brewer’s framework, but it could be an important addition to future iterations of the framework. In studying anomalous data, it is critical to know what will get students to change instead of respond in a different way, but there is likely instructional value in, for example, getting students to respond with abeyance over ignore, reject, or reinterpret.

The specific hierarchy identified here is most likely tied to our tasks being questions. Such a hierarchy is not necessarily expected to hold when students are presented with anomalous data while completing other tasks.
8.10 Conclusion

In science instruction, students should be able to change their ideas when presented with correct anomalous data. Understanding what responses students have to anomalous data and why they chose to respond in those ways is an important part of being able to use anomalous data effectively in instructional techniques.

Some students treated the given correct variation as anomalous data; therefore this variation can be used to study how students respond to anomalous data.

In the analysis of students’ responses to given correct variations, students disagreed with the provided outcome, reinterpreted problem features, delayed accepting or rejecting the data, and changed their set of ideas. Although it was based on historical data or experiments that took place over long periods, Chinn and Brewer’s framework describes all of the types of responses received here.

It was not anticipated that students would disagree with or reinterpret the responses provided to them, and these results are consistent with C&B’s notion that students’ ideas can be entrenched.

Chinn and Brewer’s framework was further expanded upon by demonstrating that the rate of disagreement increases with question difficulty. Although difficulty is not a characteristic that Chinn and Brewer list, question difficulty may be related to an amalgamation of factors they do describe, such as the entrenchment of incorrect ideas or the lack of availability of alternate ideas consistent with the anomalous data.

Furthermore, the possible responses to anomalous data appeared as if they could be hierarchical in the current study’s administration. Students rarely disagreed with the given correct variation if they could reinterpret it and did not commonly defer on passing judgment about the anomalous data. Further study of this potential hierarchy could provide insights into how students respond to questions in general.
by helping explain how they view these types of responses, thereby potentially providing insights into their epistemological beliefs as well.

Because the *given correct* variation provided anomalous data to students as part of a physics course, this result could provide some insights into how students respond to information provided through instruction. Understanding the reasons that students provide certain types of responses when presented with anomalous data allows instruction to present anomalous data in ways that more often lead students to ideas consistent with the accepted content understanding.

This chapter has shown how factors of questions such as difficulty, entrenchment, or ambiguity can influence how students react to these variations and how the same variation can work differently based non-surface-level features of the question. Narrowing down the root cause of why students respond to these questions as they do is a very difficult task because the variations do not cause the same reaction every time. This is what makes it difficult to have data-based discussions of how these questions elicit alternate ideas. Fortunately, this study identified this trend, as well as a framework that described it, so how this variation works could be discussed. Similar discussions about the other variations would be beneficial, but unfortunately patterns in the responses to the other variations that provide insights into how they influence students’ thinking were not observed.
CHAPTER 9
GENERAL RESULTS FROM ALL CONTENT AREAS

9.1 Introduction

The goal of the across-content study was to demonstrate that more could be learned about the physics ideas students have by asking question variations. Some question variations more effectively elicited unique or alternate ideas from students. These question variations were considered more successful because they more frequently provided more information about the ideas students have than the original version alone did, highlighting other types of reasoning students can use to make sense of the physical situations presented to them.

This chapter discusses the results of all question variations that were part of the across-content study, including the different types of results and the findings of each question variation related to that result. In addition, this chapter reviews the results that span findings—namely, that targeting some particular responses led to results more often than others, that students could express unique reasoning to justify the elimination of a response, that students appeared to be able to adapt the same reasoning to different variations, and that students sometimes appeared to use the provided outcome when developing reasoning. This chapter covers all administrations of variations in any content area and discusses these results in a more general way than in Chapters 6 & 7 due to these findings being from many different content areas. This chapter is intended to show how common the types of results seen in Chapters 6 & 7 are as well as identify other patterns in the data.

All question versions used in the across-content study are shown in Appendix B.
9.2 Methods for Comparing Variations

This dissertation intended to investigate alternate ideas that are not elicited by one version of a question alone. Our goals were to demonstrate question variations can elicit those alternate ideas, catalog the alternate ideas students had, and investigate patterns in student reasoning in response to the question variations; thus demonstrating the need for more in-depth investigation to develop a complete understanding of students’ ideas.

Therefore, three different types of results were considered to help determine if variations were able to elicit alternate ideas not seen in response to the original version. Other aspects of the analysis of these questions that are deemed important are also discussed.

The primary method for checking to see if these questions are accomplishing the task for which they were designed was to look for types of reasoning given in response to a variation with different prevalence than in response to the original version.

To look for instances where the question variations elicited alternate ideas, three comparisons were made between the original version and each of the question variations:

- Prevalence of Responses - Differences in the proportion of students selecting each response regardless of reasoning.

- Presence of Unique Reasoning - A statistically significant difference in reasoning that was used by more than 10% of students on a single variation and was not used by any students on the original version.

- Prevalence of Non-Unique Reasoning - Differences in reasoning established by comparing the number of students providing each particular type of reasoning
between the original version and one of the variations. This category is broken down into two subcategories:

- Informative differences - Differences where the reasoning could be applicable to responses available on both versions.
- Expected differences - Differences in types of reasoning that argue for a response that is not an option on one of the versions.

Finally, a few important aspects of the statistical analysis should be noted. Although more than 40 students almost always answered each variation, the statistical power of this study was lower than hoped for, resulting in a much lower chance of finding statistically significant results where differences exist. Having a larger number of students answering each question may have helped find additional differences between the variations.\(^1\) In addition, many statistical tests were run on the dataset to look for differences, and readers should be aware that some results may be statistical anomalies. More details of the statistical analysis used to determine if differences are statistically significant are provided in Section 5.5.3.

The following sections discuss each of these comparisons and how often these differences appeared on each question variation. Table 9.1 shows how many of these differences were observed in each variation. Table 9.2 shows all questions and the variations made.

### 9.3 Differences in Response Prevalence

This section compares the prevalence of responses between the original version and each question variation. Because there were no response options on the given correct variation, it is not discussed in this section.

\(^1\)The low statistical power of this study is discussed further in Section 5.4.2.
Table 9.1. A summary of the results of all variations from the across-content study. This table shows a count of how many results were found for each of our analysis categories: response differences, unique reasoning, and informative differences. Each question variation is labeled with the number of variations that were administered.

<table>
<thead>
<tr>
<th></th>
<th>Eliminate (6)</th>
<th>Consider (15)</th>
<th>Given Correct (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Difference in Response</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest Showing Result</td>
<td>3 (50%)</td>
<td>1 (7%)</td>
<td>-</td>
</tr>
<tr>
<td>Total Number of Results</td>
<td>4</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td><strong>Unique Reasoning</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest Showing Result</td>
<td>2 (34%)</td>
<td>1 (7%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Total Number of Results</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Informative Differences</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest Showing Result</td>
<td>3 (50%)</td>
<td>2 (13%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Total Number of Results</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Differences in response prevalence were identified because they are an indicator that students are responding to these versions differently and using different ideas to develop reasoning.

Several instances were identified on the eliminate variation, where students commonly selected and eliminated the same response, but only one instance was found where the consider variation had a statistically significant difference in responses compared to the original version.

9.3.1 Response Similarities from Eliminate

Because students were asked to select correct and incorrect responses on the original and eliminate versions, they were not expected to have similar responses on these questions when they used the same ideas. Instead, they were expected to select common responses on only one of the two versions and uncommon ones on the other. For example, it was anticipated that the responses a large percentage of the class selected as correct would not also be selected as incorrect by a large percentage of the class. However, in a few instances, this was indeed the case; these referred to as “conflicted responses.”
Table 9.2. Pretest designation, question name, and type of variation for each question administration in the across-content study. The consider variations that asked students to consider the correct response are indicated.

<table>
<thead>
<tr>
<th>Pretest</th>
<th>Question Name</th>
<th>Question Versions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest 2</td>
<td>Oval Track</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Greater Than(Correct)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Less Than</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Equal To</td>
</tr>
<tr>
<td>Pretest 3</td>
<td>Pendulum</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Upwards(Correct)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 4</td>
<td>Car-Over-Hill</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Greater Than(Correct)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 5</td>
<td>Newton’s 3rd Law</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Greater Than</td>
</tr>
<tr>
<td>Pretest 6</td>
<td>Gymnast</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 7a</td>
<td>1-D Conservation of Momentum A</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Greater Than</td>
</tr>
<tr>
<td>Pretest 7b</td>
<td>1-D Conservation of Momentum B</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Less Than(Correct)</td>
</tr>
<tr>
<td>Pretest 8</td>
<td>2-D Conservation of Momentum</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 9a</td>
<td>Change in Kinetic Energy</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Equal To(Correct)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 9b</td>
<td>Work-by-Wall</td>
<td>Original, Eliminate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 10</td>
<td>Change in Momentum</td>
<td>Original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consider Equal To</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 11</td>
<td>Pivoting Rod</td>
<td>Original, Consider Equal To</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 12</td>
<td>Rotating Disk/Wheel</td>
<td>Original, Consider Greater Than</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
<tr>
<td>Pretest 13</td>
<td>Simple Harmonic Oscillator</td>
<td>Original, Consider Zero</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Given Correct</td>
</tr>
</tbody>
</table>
Table 9.3. Number of students selecting, or eliminating, each response on the original and eliminate versions for all conflicted responses in the across-content study.

<table>
<thead>
<tr>
<th>Conflicted Responses</th>
<th>Original Version(%)</th>
<th>Eliminate Variation(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest 5 - Equal to</td>
<td>14(32%)</td>
<td>22(38%)</td>
</tr>
<tr>
<td>Pretest 6 - Greater than</td>
<td>12(41%)</td>
<td>22(34%)</td>
</tr>
<tr>
<td>Pretest 9b - Negative</td>
<td>16(31%)</td>
<td>16(25%)</td>
</tr>
<tr>
<td>Pretest 9b - Zero</td>
<td>29(56%)</td>
<td>29(45%)</td>
</tr>
</tbody>
</table>

Conflicted responses were previously discussed in Section 4.2.2.2 of the pilot study and in Section 6.5.2 on the results of the Work-by-Wall question. Table 9.3 lists all observed instances of conflicted responses.

Finding conflicted responses was an unexpected result. The initial assumption was that what the class thinks of as correct, they do not think of as incorrect and vice versa, which should lead to responses that are either commonly selected as correct or commonly eliminated. In order to document these cases, a criterion was developed for the responses of a class to qualify as conflicted.

Students were expected to answer the eliminate and original versions of the questions differently, so the criterion we defined for being conflicted was having a \( p \) value of greater than 0.2 when comparing the usage of a particular response on the original and eliminate versions. This criterion was used because it indicates when the results are not close to finding a difference between the two groups. Although not an explicit requirement, none of the responses for which students were conflicted were chosen as correct or eliminated as incorrect by fewer than 25% of students.

Out of the six questions with eliminate variations, the Work-by-Wall question discussed in Section 6.5.2 was the only question where students were conflicted about more than one response.\(^2\) Two more questions showed students were conflicted about one of three possible responses, and the remaining three questions had no conflicted responses.

\(^2\)Students were conflicted about 2 of 3 possible responses on this question.
The existence of conflicted responses is notable because they indicate that a single student may have ideas that support and counter a particular response and that the student can be cued to use either one by a question variation. This is indicated by conflicted responses where the sum of the percentages of students justifying and arguing against a particular response is close to or above 100%, which is consistent with Steinberg and Sabella’s [43] research proposing that students can have multiple mental models simultaneously. This is also consistent with the resources framework as the different variations could activate different resources causing the assembly of knowledge structures that are in conflict with one and other.

In addition, finding conflicted responses shows that, within the classroom, there are accessible ideas for both justifying and eliminating these responses. Having class discussions about these responses could be an effective form of instruction, and these responses could provide a good opportunity for peer instruction.

9.3.2 Response Differences from Consider

When comparing the response distributions between the original and consider versions, in only one instance out of 15 were the distributions statistically different. The one observed difference was on the Oval Track question from Pretest 2 (Figure B.2), which asked students to consider a correct response (see Table 9.4). For this question, fewer students selected the response greater than on the original version of the task than they did when asked the consider greater than variation. However, no differences occurred in the prevalence of reasoning between the original and consider greater than versions of this question. Because this is only one result out of 15 variations, and no difference was found in reasoning to explain the difference in responses, it is possible that this result is simply a statistical anomaly.
Table 9.4. Number and percentage of students giving the response greater than, on the original and consider greater than versions of the Oval Track question. Students giving the responses less than or equal to are categorized as responding NOT greater than for comparison between the two variations. [$p$ value=.011, odds ratio=2.79, $\phi$=.253]

<table>
<thead>
<tr>
<th></th>
<th>All Students</th>
<th>Original</th>
<th>Consider Greater Than</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater Than</td>
<td>22(41%)</td>
<td>33(66%)</td>
<td></td>
</tr>
<tr>
<td>NOT Greater Than</td>
<td>32(59%)</td>
<td>17(34%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

This result suggests that students are likely approaching answering the original and consider versions of the question in the same way and these question versions may not be distinct enough to elicit different responses from students.

9.3.3 Differences in Response Prevalence Conclusion

In looking at responses to the variations, conflicted were often found when using the eliminate variation, but the response students chose using the consider variation often could not be changed. As the consider variation was the variation most similar to the original version and the only one that could cause a directly comparable difference in response prevalence, these variations, or similar ones, may be able to cause differences in response prevalence, but the consider variation was not different enough from the consider variation to lead students to change their responses.

9.4 Unique Reasoning in Variations

In several cases, a question variation brought out unique reasoning not seen in responses to the original version of a question. In order for reasoning to be considered unique, the reasoning must meet three criteria: it must be a statistically significant difference in reasoning; it must be used by more than 10% of students responding to the variation; and it must not be used by any students answering the original version. The 10% criterion was used to ensure the effect is large enough to
be worth discussing. With only about 50 people answering each question, dropping below this threshold includes types of reasoning used by only a few students, which would be too many potentially rare types of reasoning to discuss.

It is interesting to identify unique reasoning because it is a clear indicator of alternate ideas, with the ideas only present in response to variations. It was expected to find unique reasoning in response to all question variations, but the types of unique reasoning observed were not anticipated.

The results show that unique reasoning was found in response to all question variations. It was found on less than 10% of the consider variations, while being found on a fifth of given correct variations and a third of eliminate variations.

9.4.1 Unique Reasoning from Eliminate

Two cases of unique reasoning were identified when comparing the reasoning used on the original and eliminate versions.

It was predicted that asking students to eliminate a response would encourage students to use types of reasoning that were appropriate for eliminating a particular response and discourage students from using types of reasoning that only work as a justification for a particular response. Therefore, unique reasoning was expected to be more commonly used on the eliminate variation.

A previously discussed case of students using unique reasoning on an eliminate variation was displacement means work reasoning which was used on the eliminate variation, but not the original version, of the Work-by-Wall question, (see Section 6.5.3). A second case was velocity must change reasoning, which was used on the eliminate variation of the Pendulum question from Pretest 3, but not on the original version. Both types of reasoning were most exclusively used to eliminate the response zero. Because this second result was not discussed previously, it will briefly be discussed here.
Table 9.5. Number and percentage of students using *velocity must change* reasoning in response to the *original* and *eliminate* versions.\[p\text{ value}=8.273\text{-e-06, odds ratio}=\infty, \phi=.474\]

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Eliminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity Must Change</td>
<td>0(0%)</td>
<td>14(36%)</td>
</tr>
<tr>
<td>All Other</td>
<td>42(100%)</td>
<td>25(64%)</td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>39</td>
</tr>
</tbody>
</table>

The *original* version of the Pendulum question from Pretest 3, shown in Figure B.5, asked students which direction the acceleration of a pendulum bob was at the bottom of its swing. *Velocity must change* reasoning is where students correctly stated that the acceleration of a pendulum bob at the bottom of its swing is not zero because the magnitude or direction of the velocity of the pendulum bob is changing at the bottom of its swing. This reasoning is correct, although the nature of the variation allows it to be less specific about the direction the change is in than correct responses on the *original* version.

As shown in Table 9.5, *velocity must change* reasoning was not used on the *original* version, was used by more than 10% of students on the *eliminate* variation, and was a statistically significant difference in usage between the *original* and *eliminate* versions.

This is another example of students using types of reasoning that are specific to justifying why a particular response is incorrect.

Although both of these instances of unique reasoning can be coded as different types of reasoning from those used on the *original* version of the question,\(^3\) these unique types of reasoning may simply be less specific versions of other types of reasoning that were used on the *original* version.

Section 6.5.3 discussed that *displacement means work* reasoning may be a subset of *coordinate system-dependent* reasoning that simply does not go on to talk about

\(^3\)Inter-rater reliability was performed on the coding of reasoning in response to the Work-by-Wall question to 94% agreement as discussed in Section 6.4.
whether the non-zero work is positive or negative because that is not being asked about. Similarly, *velocity must change* reasoning may be a less specific version of *circular motion* reasoning. Students used *circular motion* reasoning on the *original* version to justify the acceleration being upwards by stating that the pendulum bob was following a circular path and, therefore, the acceleration must be toward the pivot or toward the center of the circle. *Velocity must change* reasoning could simply be a more simplistic and less complete version of *circular motion* reasoning where students do not talk about how the direction or magnitude of the velocity will change because that is not required to explain why the acceleration is non-zero.

Finding that the *eliminate* variation can cause students to use less complete forms of reasoning is an interesting result. By allowing students to eliminate one response, they may not need to use reasoning that is as thorough or complex as the reasoning needed to justify why the correct response is correct. This provides support for theoretical frameworks that allow students’ ideas to be assembled on the fly, as opposed to having stable theories or pre-assembled correct reasoning that they use to answer a question.

Another commonality between these unique types of reasoning is that both were used to eliminate the response *zero*. *Zero* was the most commonly eliminated response both times it was offered as an option. It was eliminated by 71% of students on the Pendulum question and 56% of students on the Work-by-Wall question. These were the two highest percentages of students eliminating a particular response out of all possible responses on all *eliminate* variations. As discussed in Section 6.5.2 and 9.6, *zero* may be easier to eliminate than other response options because it does not indicate any specific direction.

It is not clear if these unique types of reasoning are more likely to show up when eliminating *zero*, or if there are not enough students in this study to see unique
reasoning in the less commonly chosen responses. The guideline of not talking about types of reasoning used less than 10% of the time may be too stringent.

Because the *eliminate* variation was expected to provide the best opportunity to find unique reasoning, the types of reasoning students used on *eliminate* variations where no unique reasoning was found were also examined.

On these questions, students commonly used the same the reasoning on both the *original* and *eliminate* versions of the task, with that type of reasoning being appropriate for both justifying one response and eliminating another one.

One example of this type of reasoning is seen on the Newton’s 3rd Law question from Pretest 5, which is shown in Figure B.13. This question asked students to compare the magnitude of the normal force exerted by a block on a platform to the magnitude of the normal force exerted by the platform on the block while the block and platform are accelerating upwards. The correct reasoning is that the forces will be equal because they are a third law force pair. However, on the *original* version, many students said the upward force was greater, reasoning that the block was moving upwards. Similarly, many students eliminated the downward force as being greater because the block was moving upwards; thus, no unique reasoning was observed. The most common type of reasoning used to eliminate this option was *net force in direction of motion*, which has been seen by other researchers [74].

Finally, one last result noticed was a common type of reasoning that appeared across many *eliminate* variations. In this form of reasoning, students provide a response they think is correct, without any reasoning about the response they think is correct, as a justification for eliminating a different response they think is incorrect. For example, a student selected the response “I would eliminate ‘The magnitude of the change in momentum of glider B is greater than the magnitude of the change in momentum of glider Y’” and explained “change in momentum has to be equal.” This explanation does not provide any reasoning about the situation;
were the student to provide this reasoning on the original version, it would be coded as no reasoning. However, on the eliminate variation it provides some information beyond the chosen response, and gives some insights into what the student’s thought process was.

This type of reasoning is not perceived as good reasoning because it is analogous to stating an outcome, not explaining. The presence of this type of reasoning indicates some students may not have an understanding of what makes for a good explanation. Do students view this as equivalent to providing no reasoning, or do they think it is sufficient reasoning to explain their response?

The most interesting lesson learned about unique reasoning on the eliminate variation is that it tends to come from ideas that appear to be a subset of more complex ideas, but when pared down are sufficient to eliminate a response. Differences in students’ reasoning were identified when they were using ideas that could be simplified in this way, but not when students had to used the same reasoning to argue both directions without simplification.

This calls into question the research value of learning about the unique ideas used to respond to the eliminate variation, as they are just simpler forms of ideas the original version elicits. However, this may be a good area for future investigation of students’ thinking as it can provide insights into how students simplify their ideas when it is not necessary to use more complex versions; it could also provide some evidence that students’ ideas are malleable and allowed to be only partially activated.

Overall, fewer-than-expected cases of unique reasoning were observed eliminate variation. It was anticipated that asking students to explain why something does not happen would cause them to used different types of reasoning than they use to justify why something is correct. However, this result can also be easily accounted
for by common types of reasoning that are both useful for justifying why a response is correct and eliminating why a response is incorrect.

9.4.2 Unique Reasoning from Consider

As with the eliminate variation, it was interesting to see if the consider variation elicited any reasoning that was not also used on the original version.

Only one instance of unique reasoning was observed in all 15 consider variations. This instance is unequal elements reasoning being used on the Change in Kinetic Energy question from Pretest 9a, where students were asked to consider the response equal to. This result is discussed in detail in Section 7.5.2. Unequal elements reasoning was given to justify why the kinetic energies of the two cars were not equal after being pushed by the same force for the same distance. Because this reasoning was used to justify why something was not correct, it is similar to the types of unique reasoning observed on the eliminate variation.

This single observed case of unique reasoning is an interesting discovery, because it suggests that students have a similar, but opposite, mathematical difficulty as the one seen in compensation reasoning or that they are simply citing changes similar to the outcome they are justifying. The similarities between unequal elements reasoning and compensation reasoning are discussed further in Section 9.7. Overall, the consider variation was mostly unable to elicit alternate ideas.

9.4.3 Unique Reasoning from Given Correct

Two instances of unique reasoning emerged in the given correct variation: asymmetry reasoning, previously discussed in Section 8.6, and non-compensation reasoning, which has not been previously discussed.

Asymmetry reasoning was a unique form of reasoning used on the Gymnast question from Pretest 6. On this question, students were asked about the tension in one of two non-vertical ropes suspending a gymnast in the air. On the original
Table 9.6. Number and percentage of students using non-compensation reasoning in response to the original and given correct versions of the Change in Momentum question from Pretest 10. \( p \text{ value} = .0261, \text{ odds ratio} = \infty, \phi = .267 \)

<table>
<thead>
<tr>
<th></th>
<th>All Students</th>
<th>Original</th>
<th>Given Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Compensation</td>
<td>0(0%)</td>
<td>6(13%)</td>
<td></td>
</tr>
<tr>
<td>All Other</td>
<td>45(100%)</td>
<td>39(87%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>45</td>
<td></td>
</tr>
</tbody>
</table>

version, students were asked if the tension in the left rope was greater than, less than, or equal to half the gymnast’s weight. The correct response was that the tension was greater than half the gymnast’s weight because the ropes were not vertical. The unique reasoning seen in this instance was students responding to the given correct variation with various asymmetry arguments to explain why the tension in the left rope was greater.

As demonstrated in Section 8.6, these students used asymmetry to reinterpret the physical situation in a way that allowed them to use one-dimensional kinematics ideas to justify the provided outcome.

A second instance of unique reasoning is the use of non-compensation reasoning from the Change in Momentum question from Pretest 10, shown in Figure B.34. In this question, students were asked about the change in momentum of two carts of unequal mass that were pushed with the same force for the same distance. This question uses the same physical situation as the Change in Kinetic Energy question discussed in Chapter 7. Here the momentum of the heavier cart is larger because it has a lower acceleration, causing it to spend more time between the lines, thereby receiving a larger impulse and having a greater change in momentum.

When using non-compensation reasoning, students discussed how mass and velocity balanced in such a way that the momenta were not the same. This type of reasoning is considered non-compensation reasoning because it seems highly similar to compensation reasoning, with the twist that it is used to justify why the momenta
are unequal. This reasoning is different from *more mass, more velocity, and unequal elements* reasoning discussed in Section 6.4.1 because it focuses on a ratio of the two quantities rather than simply indicating which one is greater, moreover, it does not focus only on the fact that the two elements are unequal. Table 9.6 shows the prevalence of this reasoning on the *original* and *given correct* versions.

Here are two examples of student use of *non-compensation* reasoning from the Change in Momentum question:

- “The mass of b must be so much greater than A to make up for and then some the velocity of A.”
- “momentum is equal to mass times velocity, at this point the ratio of mass and velocity is greater for B”

In *compensation* reasoning, which students commonly used on the *original* version, they argued that the mass and velocity would compensate so that momenta would be equal. *Compensation* reasoning was commonly used on the Change in Kinetic Energy question from Pretest 9. Examples and discussion of the usage of this type of reasoning were presented in Section 7.4.1.1.

*Compensation* and *non-compensation* reasoning are similar approaches to developing a response, but the approaches are utilized differently to be able to justify different responses. In addition, students commonly used *compensation* reasoning to justify a response they chose; this was not the case with *non-compensation*, as no students used this reasoning on the *original* version of either the Change in Momentum or Change in Kinetic Energy questions. The similarities between these two types of reasoning are discussed further in Section 9.7.

As discussed in Section 7.5.1, when students were presented with a *given correct* variation of the Work-by-Wall question and asked to justify why the kinetic energies of the two carts were the same, students used *compensation* reasoning more
prevalently. In that case, students appeared to be using the fact that the kinetic energies were the same to justify their argument that the mass and velocity would compensate, which is circular reasoning. With *non-compensation* reasoning students were expected to do the same thing—this time using the different momenta of the two carts to justify why the mass and velocity did not compensate. Results in which students may have used the provided outcome as part of their reasoning are discussed further in Section 9.8.

9.4.4 Unique Reasoning Conclusion

The question variations had varying levels of success at finding unique reasoning. The *eliminate* variation was the most successful, with a third of its administrations turning up unique reasoning, while the *consider* variation only managed to find unique reasoning for one of 15 administrations.

The *eliminate* variation was often successful at eliciting simpler forms of reasoning by having the response *zero* be an option. Eliminating a response seems to allow simpler forms of reasoning to be used instead of the more complete forms of reasoning required to justify a correct response. The *zero* was a frequent target for elimination, possibly because it is a discrete value with no direction rather than a range of values or vector and therefore easier to justify its elimination.

Two types of reasoning were observed for these variations. Students only provided the response they thought was correct as reasoning when responding to the *eliminate* variation, and they also appeared to be using the provided outcome as part of their reasoning on the *given correct* variation. Both of these are undesired results that do not indicate which ideas students are using to solve these tasks, and future research with these variations should attempt reduce these types of responses or at least be aware they may show up.
The *eliminate* variation seems to have the most potential for uncovering unique reasoning and for finding patterns that can help determine why students use particular unique reasoning in response to this variation. The *given correct* and *consider* variation did not uncover a sufficient amount of unique reasoning to allow for connections to be drawn between features of the question and the successful elicitation of unique reasoning.

### 9.5 Differences in Reasoning Prevalence

This section discusses all instances where a particular type of reasoning is provided on both the *original* version and a variation, but where the prevalence of the reasoning is different. Examples of differences in reasoning are presented that seem to be due to a change in the available responses from the *original* version to the variation—referred to as “expected differences.” In addition, instances where the differences do not appear to be due to a chance in the available responses and that do not qualify as unique reasoning—called “informative differences”—are presented. Finally, the discussion reviews instances where no differences in reasoning emerged and why these differences are absent.

Although both informative and expected differences provide valuable information about students’ ideas, a distinction can be drawn between them because expected differences are predicted to occur much more frequently and have a consistent cause for the difference. An example of an expected difference would be for reasoning based on Newton’s 3rd Law, used to justify why two forces are equal on the *original* version of a question, to be effectively absent on a *given correct* variation that states the two forces are not equal. Because this difference in reasoning is consistently due to the reasoning not being applied to the available response choices, and because these expected differences occur frequently, they are grouped into a distinct category.
The goal was to identify differences in reasoning because such differences show that a larger number of students have particular ideas than the original version indicates; therefore, those ideas are alternate ideas for some students.

Informative differences were common and were found on half of the eliminate variations and 20% of the consider and given correct variations.

9.5.1 Informative Differences - Eliminate

Three instances of informative differences were found in the prevalence of a particular type of reasoning in response to eliminate variations. These differences are discussed here because they provide insights into how students’ ideas change in response to the version of the question being asked.

One of the informative differences identified was in the usage of force means work reasoning on the Work-by-Wall question (see Section 6.5.3). Here students said that a wall did non-zero work on an oscillating block-spring system because there was a force acting on the block. This is an incorrect application of the definition of work. Knowing that this type of reasoning is prevalent helps better understand how students misuse the definition of work and increases the importance of addressing this set of incorrect ideas.

The two other informative differences found were on Pretest 4 & 6; both involved differences in the prevalence of the same type of reasoning, which is referred to here as otherwise reasoning.

Otherwise reasoning occurs when students make an argument that one response must be correct because one, or both, of the other responses cannot be correct. This type of reasoning was more common on the eliminate variation than on the original version for two questions: Pretests 4 & 6. The questions asked on these pretests were the Car-Over-Hill question (see Figure B.10) and the Gymnast question (see
Table 9.7. Number and percentage of students using otherwise reasoning in response to the original and eliminate versions of the Car-Over-Hill question from Pretest 4. [$p$ value=.0338, odds ratio=2.61, $\phi=.225$]

<table>
<thead>
<tr>
<th>All Students</th>
<th>Original</th>
<th>Eliminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otherwise</td>
<td>11(24%)</td>
<td>23(46%)</td>
</tr>
<tr>
<td>All Other</td>
<td>34(76%)</td>
<td>27(54%)</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 9.8. Number and percentage of students using otherwise reasoning in response to the original and eliminate versions of the Gymnast question from Pretest 6. [$p$ value=.0303, odds ratio=8.78, $\phi=.22$]

<table>
<thead>
<tr>
<th>All Students</th>
<th>Original</th>
<th>Eliminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otherwise</td>
<td>1(1%)</td>
<td>5(12%)</td>
</tr>
<tr>
<td>All Other</td>
<td>68(99%)</td>
<td>38(88%)</td>
</tr>
<tr>
<td>Total</td>
<td>69</td>
<td>43</td>
</tr>
</tbody>
</table>

Figure B.17). The prevalence of otherwise reasoning on these questions is shown in Tables 9.7 and 9.8.

For the Car-Over-Hill question, students were asked to compare the gravitational force exerted by the earth on a car to the normal force exerted by the ground on the car as the car crested the top of a circular hill. The correct reasoning for this question is that the gravitational force is greater, since at the top of the hill the car needs a downward acceleration to continue following the curve of the hill. The most common incorrect response on the original version was for students to state that the two forces are equal. In justifying this response, students used otherwise reasoning to explain that the forces must be equal in order for the car to not float up into the air or sink into the hill. Although these examples mention both sinking and floating, some students only mentioned one or the other.

Here are two examples of otherwise reasoning from the Car-Over-Hill question:
• “The magnitudes are equal because if either were different then the car would be pushed up off the ground or down into the ground. Since they are equal the car can continue to drive down the hill.”

• “The magnitude is the same because if they were different the car would sink into the hill or the car would be pushed up by the hill.”

Similarly, on the Gymnast question (see Section 8.6.2), students used otherwise reasoning to express that the magnitude of the upward net-force had to be equal to the magnitude of the gymnast’s weight; otherwise, the gymnast would accelerate upward or downward. Although this is correct reasoning, students often failed to acknowledge that there were x-components to the tension force from each rope, causing the tension in each rope to be greater than half the gymnast’s weight.

This reasoning is based on a process of elimination. As expected, it is sometimes used in response to consider variations, as that is the other variation where students can reason about why an outcome does not occur, but it is used much more prevalently on the eliminate variation and is not prominent enough to discuss in the consider section of this chapter. Despite being a reasoning focused on the elimination of possibilities, this reasoning is also used on the original version, although not as commonly as on the eliminate variation.

The more common use of otherwise reasoning on the eliminate variation indicates that students answering the original version of the question may have had access to otherwise reasoning, but chose to use other reasoning instead. One possible explanation is that students preferred a reasoning that did not rely on eliminating the alternate responses to one that justified the correct response. Another explanation could be that some students may only have been able to eliminate one response with this type of reasoning, which allowed them to use it on the eliminate variation, but not on the original version.
To the researcher’s knowledge, no existing physics education research literature has discussed students using this type of process of elimination reasoning to justify why a response is correct. Research in other areas, such as that on “test-wiseness,” indicate that students will eliminate “incorrect or absurd responses” [82, pg. 716] in order to improve their test scores, which could be applicable here. In general, we view otherwise reasoning as a complex type of reasoning with students extrapolating over time and using limiting cases to develop reasoning. It seems that otherwise involves the interaction of many different resources, and more study is needed to investigate how students develop and use this reasoning.

One last observation about otherwise reasoning is that it was only used on Pretests 4 through 6, which cover two-dimensional kinematics. It was used most commonly on the Gymnast question from Pretest 4, then was used less commonly on question variations in the following two consecutive pretests. Both of these instances of otherwise reasoning are also consistent with students having issues applying two-dimensional kinematics instead of one-dimensional kinematics, a difficulty these research tasks were designed to elicit. Although it seems like otherwise reasoning could be seen in response to questions about other content areas, two-dimensional kinematics may be a content area that elicits this type of reasoning particularly often as students have extensive everyday experience with force and motion and may be more aware of “what would happen otherwise” in this content area than in others.

9.5.2 Expected Difference Example - Eliminate

Expected differences were also common in the eliminate variation. An example of this type of reasoning is from the Pendulum question on Pretest 3 (see Figure B.6), where students were asked about the direction of the acceleration of a pendulum bob at the bottom of its swing. Many students responding to the original
version incorrectly used *gravity only* reasoning to explain that the acceleration must be downwards because gravity was the only force acting on the pendulum bob. However, *gravity only* reasoning was rarely used to eliminate a response. This response justifies why the acceleration is down, but it does not help students select or explain why any of the other directions—*upward*, *upward and to the right*, or *right*—is better to eliminate than any of the others.

Thus, this type of reasoning does not assist students in selecting a response or explaining why it is the best one to eliminate, and it is therefore not surprising that this type of reasoning is not used on the *eliminate* version. In contrast, *net force in direction of motion* reasoning, which was discussed in the previous section, works for eliminating the downward force being greater because that is the only other direction that can be chosen. It seems that, in these cases, providing students with more response options may prevent them from using the same reasoning they would use on the *original* version when eliminating responses.

Seeing *gravity only* reasoning not being commonly used on the *eliminate* variation is a positive sign that students are engaging with the question and understanding the content, but it is not really useful information as students are not expected to use this reasoning to eliminate a response. Therefore, these and other types of expected reasoning are not discussed further, except to note that they are a common occurrence.

### 9.5.3 Absence of Differences - Eliminate

In some instances of *eliminate* variations, no difference was found in the prevalence of reasoning. This is often due to the most common types of reasoning being appropriate to use on both versions of the task or students using many different types of reasoning, leading to only a few students using any particular type.
An example of a question where no differences in reasoning were found is the Newton’s 3rd Law question from Pretest 5. On this pretest, students commonly used \textit{net force in direction of motion} reasoning on both the \textit{original} and \textit{eliminate} versions of the task, as discussed in the previous section. Using this common reasoning on both versions means there are fewer students who might be using a reasoning that is not appropriate for both question versions, making differences in the prevalence of those types of reasoning harder to detect.

In cases where students use many different types of reasoning on the \textit{eliminate} variation leading to no significant differences in reasoning, the large variety of reasoning may be due to the fact that the \textit{original} version of the question was designed with the goal of investigating particular students’ ideas. When changed to an \textit{eliminate} variation, it may no longer target the same ideas or may target them less effectively, leading to the elicitation of many different ideas.

### 9.5.4 Informative Differences - Consider

Three informative differences in reasoning were found on two pretest questions. One question was previously discussed, and the other is discussed in this section.

The previously discussed result was from Section 7.5.2, where \textit{more velocity} reasoning was used less commonly on the \textit{consider equal to} variation of the Change in Kinetic Energy question. More velocity reasoning is a type of reasoning where students argue that a faster cart will have more velocity and, therefore, the carts will not have equal kinetic energies.

Another example of an informative difference is from the Simple Harmonic Oscillator question from Pretest 13, where there were differences in the prevalence of two types of reasoning. One type of reasoning was more common on the \textit{consider} variation; the other was more common on the \textit{original} version.
On the Simple Harmonic Oscillator question, which asked students to consider an incorrect response (see Figure B.42), students were asked about the acceleration of the mass at the end of a spring undergoing simple harmonic motion when the mass is at its maximum displacement from equilibrium.

When responding that the acceleration is not zero on the consider zero variation, many students stated that it was not the acceleration that was zero, but the velocity. Such reasoning is called velocity is zero reasoning. This was an unexpected type of response because the students were not reasoning about the acceleration, but rather about velocity. As shown in Table 9.9, this reasoning was given by seven students on the consider zero variation and only one student on the original version.

A similarity was noted between this type of reasoning and unequal elements reasoning, discussed in Section 7.5.2, where students were potentially only identifying quantities that changed and using those as justifications for different results. These are both very associative types of reasoning that point out quantities that are zero to justify zero or quantities that are different to justify NOT equal to, which are not complete.

Students responding to this question also used equilibrium reasoning, which states that the acceleration will point toward the equilibrium position; however, it was used by fewer students on the consider zero variation than on the original version.

Table 9.9. Number and percentage of students using velocity is zero reasoning in response to the original and consider zero versions of the Simple Harmonic Oscillator question from Pretest 13.$[p \text{ value}=0.0207, \text{ odds ratio}=9.83, \phi=0.271]$

<table>
<thead>
<tr>
<th>All Students</th>
<th>Original</th>
<th>Consider Zero</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity is Zero</td>
<td>1(2%)</td>
<td>7(18%)</td>
</tr>
<tr>
<td>All Other</td>
<td>46(98%)</td>
<td>32(82%)</td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>39</td>
</tr>
</tbody>
</table>
Table 9.10. Number and percentage of students using *equilibrium* reasoning in response to the *original* and *consider zero* versions of the Simple Harmonic Oscillator question from Pretest 13. [$p$ value=.025, odds ratio=4.51, $\phi=.255$]

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Consider Zero</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equilibrium</td>
<td>13(28%)</td>
<td>3(8%)</td>
</tr>
<tr>
<td>All Other</td>
<td>34(72%)</td>
<td>36(92%)</td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>39</td>
</tr>
</tbody>
</table>

The more prevalent use of *velocity is zero* reasoning and the less prevalent use of *equilibrium* reasoning on the *consider zero* variation may be directly linked, but these differences could also come from smaller differences in many other types of reasoning. It is intriguing that the *consider zero* variation is better at eliciting *velocity is zero* reasoning than *equilibrium* reasoning as both are very closely related to the response zero. This result seems to indicate that ideas related to *velocity is zero* reasoning are more easily cued by this variation than ideas related to *equilibrium* reasoning. As previously discussed, *velocity is zero* reasoning is a very simple and associative but incomplete form of reasoning that simply identifies something that is zero and uses that as a justification. *Equilibrium* reasoning appears to be a more complex form of reasoning, which may explain why it is not as easily cued.

9.5.5 Expected Differences Example - Consider

An example of an expected difference from the *consider* variation was observed on the one-dimensional Conservation of Momentum A question from Pretest 7a, where there was less common usage of Newton’s 3rd Law based reasoning on the *consider greater than* question. Newton’s 3rd Law is a justification for why two forces are equal and is not helpful in justifying or eliminating the *greater than* response. In this way, Newton’s 3rd Law based reasoning is not useful on the *consider greater than* variation and is expected to be used by fewer students.
9.5.6 Absence of Differences - Consider

In determining why no difference in the prevalence of particular types of reasoning on most pretests was observed, it was found that, instead of using different reasoning on the consider variations, students largely used the same reasoning as on the original version. This is likely due to students not needing to change reasoning because the reasoning they use is appropriate for both question versions; however, it may also be due to the consider variation failing to focus students on the suggested option, leading them to answer as if it was the original version.

Some responses are indicative of students responding to consider variations as if they were not simply a yes or no question. Students commonly did not respond to consider variations with a “yes” or a “no” as part of their response and instead gave the response they thought was correct as they would do on the original version. In addition, some students provided the response they thought was correct even if it was not the one they were asked to consider. For example, on a consider up variation, one student responded “its [sic] downward because the force of gravity.” The similarity between the phrasing of these responses and the types of responses usually given on the original version of the question led to an expectation that students might not be engaging with this question differently than they did with the original version.

As will be discussed in the conclusion, this observation—combined with the general lack of differences observed in response to the consider variation—indicate that students are not using alternate ideas when responding to this variation.

9.5.7 Informative Differences - Given Correct

Three instances of informative differences in the prevalence of particular types of reasoning were found on two of the ten given correct variations. These three differences were discussed in Chapters 6 & 7.
Chapter 6 discussed the Work-by-Wall question, where students were asked about the work a wall does on a block-spring system undergoing simple harmonic motion. On the given correct variation of this question, more students used no displacement reasoning to explain that the wall did no work on the block-spring system because the wall did not have a displacement.

Chapter 7 discussed the Change in Kinetic Energy question, where students were asked to compare the kinetic energies of two carts after they were pushed with the same force for the same distance. In this case, two types of reasoning were used more prevalently on the given correct variation of the question. Equal applied force reasoning, which argues that the kinetic energies of the blocks are the same because they had the same force applied to them, was used by more students in response to the given correct variation. Compensation reasoning, which argues that the kinetic energies of the blocks are the same because the differences in mass and velocity compensate for one another, was also used by more students in response to the given correct variation.

These three differences show that the given correct variation is capable of eliciting alternate ideas and often does so for a large portion of the class. Despite providing students with the correct outcome, the given correct variation did not always elicit correct ideas. Only one of the three types of reasoning discussed here is what could be considered correct reasoning.

Although evidence suggests that the given correct variation can elicit alternate ideas, it was rarer to find evidence that providing students with the correct outcome changed the reasoning they used to justify that outcome. In this case, the more prevalent usage of compensation reasoning demonstrated that the types of reasoning students used to justify the correct response changed when the outcome was provided for them, as providing students with the correct response made them statistically more likely to use compensation reasoning.
9.5.8 Expected Differences - Given Correct

Because many types of reasoning used to justify incorrect responses on the original version of the question are no longer applicable when justifying the correct outcome on given correct, there are many expected differences when comparing these questions. In eight of the ten given correct variations, at least one type of reasoning used on the original version to justify why an incorrect response is correct were not used on the given correct variation. Because these differences are expected and common, and examples of expected differences have been discussed previously, an analysis of the given correct expected differences is not presented in this dissertation as it does not provide much value beyond what has already been presented.

9.5.9 Absence of Differences - Given Correct

The questions with no statistically significant difference in the prevalence of any type of reasoning were the Car-Over-Hill and Pivoting Rod questions from Pretests 4 and 11. On the Car-Over-Hill question, many students had smaller, statistically insignificant differences in the prevalence of many types of reasoning, but no larger difference in one particular kind of reasoning. In this case, the diversity of incorrect responses on the original version of the question led to a lack of substantial changes in the usage of any particular type of reasoning. On the Pivoting Rod question, the original version was answered correctly by the majority of students, and both groups of students used the same reasoning to justify the correct response.

9.5.10 Differences in Reasoning Conclusion

Several informative differences in reasoning were identified for all variations, and the results are shown in the list below. Most of these results were discussed in previous chapters and are denoted with an asterisk.

Ignoring students who disagree with the provided response, these are all statistically significant differences.
• Eliminate
  – Work-by-Wall - Force Means Work*
  – Car-Over-Hill - Otherwise
  – Gymnast - Otherwise

• Consider
  – Change in Kinetic Energy - More Velocity*
  – Simple Harmonic Oscillator - Velocity is Zero
  – Simple Harmonic Oscillator - Equilibrium

• Given Correct
  – Work-by-Wall - No Displacement*
  – Change in Kinetic Energy - Equal Applied Force*
  – Change in Kinetic Energy - Compensation*

The informative differences presented in this section were the two examples of otherwise reasoning and the two differences from the consider version of the Simple Harmonic Oscillator question.

Using the eliminate variation allowed for a study of students’ use of otherwise reasoning to show that a significant portion of the class was capable of using this type of reasoning and that it was more common when students were asked to justify why a response was incorrect.

Being aware of which types of reasoning students use to either justify or eliminate responses, as well as the types of reasoning that they use to both justify and eliminate responses, provides more information about these types of reasoning and how students can apply them. This provides a better understanding about the
ideas students have and the way they develop responses to conceptual physics
questions.

Again, the consider variation showed the lowest percentage of results, but the
results it did show reinforced that zero is a response that leads to differences in
reasoning when combined with these variations. The importance of the response
option zero is discussed in the next section.

Several expected differences from each question variation were also identified.
Expected differences are not the most groundbreaking results, but they are
discussed here because they show that students are engaging with the questions in
the desired way most of the time, and are changing their reasoning to be specific to
the responses they have available. Expected differences should be predictable if the
reasoning used on the original version is known. Any reasoning that does not
support the response chosen for the given correct variation or that does not work to
preferentially eliminate one response of the eliminate variation should not be used
as often on those versions.

Expected differences in reasoning were found less often on the consider variation
than on the eliminate and given correct variations. On the consider variation,
students were either not forced to justify or not forced to eliminate a particular
response; therefore, the reasoning used on the original version was more applicable
to the response options on the consider variation than on either of the other two
variations. This is one of the ways that the consider variation is more similar to the
original version than the other variations, making the relatively small number of
expected differences unsurprising.

Finally, there were common absences of differences found for each variation. The
absence of differences seemed to be due to students not needing to use reasoning
different from the original version because it applied to both the original version
and the variation they received or because the reasoning was very scattered and
could not be grouped into large enough bins to identify statistically significant changes. In addition, in the case of the consider variation, students may have been treating the variation as similar to the original version, leading to a lack of differences.

9.6 Zero as a Response Option

This section discusses the combined result that all informative differences on the consider and given correct variations come from variations that target the responses zero or equal to and why targeting these responses may cause the elicitation of alternate ideas more often than other response options.

Zero was a response option on one consider and one given correct variation and, as shown in Table 9.11, both of these variations had an informative difference in reasoning. For the response equal to, the only given correct variation targeting this response option showed an informative difference in reasoning, but only one out of three consider equal to variations showed an informative difference. These results combined account for all of the informative differences found on the consider and given correct variations in this study. In addition, two out of the four conflicted responses found using the eliminate variation targeted these responses, and zero was the most commonly eliminated response both times it was offered as an option. It was eliminated by 71% of students on the Pendulum question and 56% of students on the Work-by-Wall question. These were the two highest percentages of students eliminating a particular response out of all possible responses on all eliminate variations.

These results suggests that zero, equal to, or other similar responses are a special type of response, and students’ ideas may be more influenced by focusing them on these responses than others.
Table 9.11. Features and results of consider and given correct variations with zero or equal to as a response option. The count of the types of reasoning that showed an informative difference is shown in the last column.

<table>
<thead>
<tr>
<th>Question</th>
<th>Zero/Equal To</th>
<th>Variation</th>
<th>Informative Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Harmonic</td>
<td>Zero</td>
<td>Consider Zero</td>
<td>2</td>
</tr>
<tr>
<td>Oscillator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work-by-Wall</td>
<td>Zero</td>
<td>Given Correct</td>
<td>1</td>
</tr>
<tr>
<td>Change in Kinetic Energy</td>
<td>Equal To</td>
<td>Consider Equal To</td>
<td>1</td>
</tr>
<tr>
<td>Change in Momentum</td>
<td>Equal To</td>
<td>Consider Equal To</td>
<td>0</td>
</tr>
<tr>
<td>Pivoting Rod</td>
<td>Equal To</td>
<td>Consider Equal To</td>
<td>0</td>
</tr>
</tbody>
</table>

Because equal to can be expressed as “zero change,” because zero leads to informative differences more consistently than equal to, and because there is prior research about a response bias for zero but not equal to, this section primarily examines how the availability of zero as a response option influenced the results.

Numerous publications in mathematics education research have focused on zero, but such research targets division by zero [83], developing a concept of zero [84], and the history/language of zero [85], which are not applicable here. In addition, several studies have discussed student reasoning about zero [53, 56, 86], but only in response to a single question, not in general.

One previous research result that reinforces the finding that students treat the response zero differently is a paper on student responses to the Force Concept Inventory (FCI) and the Conceptual Survey on Electricity and Magnetism; it showed that students were biased against selecting zero when it was presented as a response option [87]. Although the current research finds different results, this study does reinforce that in some cases students do not treat it the same as other response options.
The three informative differences found on variations with zero as a response option all involved different types of reasoning. One type of reasoning was no displacement, discussed in Section 6.5.1, where students said that a wall did no work on an oscillating block-spring system because there was no displacement of the wall. The two other types of reasoning were velocity is zero and equilibrium reasoning, briefly discussed in Section 9.5.4. Here students explained that the acceleration was not zero, but the velocity was, or that the acceleration was non-zero because it was pointed back toward the point of equilibrium. All of these types of reasoning are correct statements, but are not all complete types of reasoning.

Some similarities emerged between no displacement and velocity is zero reasoning, suggesting that students are being cued to identify quantities that are zero and to use those in their reasoning. In no displacement reasoning, students were provided with an outcome stating that the work done by a wall was zero, and asked to explain. They respond by stating that the work done by the wall was zero because its displacement was zero. In velocity is zero reasoning, students were asked if the acceleration was zero; they explain that it was not, but velocity was zero. Velocity is zero reasoning is particularly odd because it does not involve reasoning about the acceleration, which is what the question asked about. Because this type of reasoning does not address acceleration, we suspect students might be being asked if something is zero, looking for something that is zero—potentially knowing that multiplicatively one zero makes everything zero—and identifying a quantity that is zero, using that as their reasoning. It seems as if ideas related to the response zero can readily be connected to quantities that are zero in order to develop reasoning that supports the response zero. This pattern would lead to both of these types of reasoning, and supports the argument that students were cued to identify quantities that are zero. This result adds to prior research on cuing, discussed in Section 2.5.
Zero may be commonly cued for several reasons. It is an uncommon response option that was only used on three out of 11 questions and may stand out more because of its infrequent use. Furthermore, zero is direction-less and discrete, where the other options students were presented with often had a direction (right, left, upwards) or a range of values (positive, greater than). These features of zero may cause it to be strongly associated with some ideas and also make it an easier response to justify or eliminate because students do not need to justify a specific direction instead of only justifying a lack of direction; moreover, eliminating a discrete value is easier to justify than eliminating a whole range of values. This was previously discussed in Section 6.5.2.

*Equal to* is also non-directional and discrete, is a less common response option than greater than or less than, and could cue students to look for quantities that are equal in a similar way to zero cuing students to look for quantities that are zero. These similarities explain some of the similar results observed across the zero and equal to response options. There may be additional response options that fit into this category as well, such as remains the same, which was used in the pilot study and has these same features.

In designing the across-content study, it was not anticipated that targeting particular responses would have an effect beyond targeting correct or incorrect responses. However, this result indicates that the targeted response is extremely important. Future research using these question variations should definitely take this result into account. In addition, due to the prevalence of these response options in other areas of physics education research, further research to determine why these particular response options have such a strong effect should be performed, in relation to both question variations and otherwise.
9.7 Compensation-like Reasoning

This section discusses three types of reasoning that make a very similar type of argument, but appear as distinct and unique types of reasoning on different question variations. The similarity of these types of reasoning suggests that they are all variations of the same underlying set of ideas, each adapted to the different response options available on a question variation. These types of reasoning are compensation, non-compensation, and unequal elements.

Compensation reasoning was initially presented by Lawson and McDermott [23] and is discussed in this dissertation in Section 7.4.1.1. Unequal elements and non-compensation reasoning are discussed in Sections 7.4.1.6 and 9.4.3, respectively. All three types of reasoning are referred to here as compensation-like reasoning.

All compensation-like reasoning was used in response to the Change in Kinetic Energy and Change in Momentum questions, shown in Figures B.26 and B.32, which asked about the same physical situation of two carts of different mass being pushed with the same force over the same distance.

All types of compensation-like reasoning are based on comparing the mass and velocities of the two carts, with students realizing that one cart has higher mass and lower velocity than the other. All of these types of reasoning are incomplete.

Compensation reasoning argues that the mass and velocity differences compensate, leading to both carts having the same final kinetic energies. The results indicated that this reasoning was used more prevalently on the given correct variation of the Change in Kinetic Energy question than on the original version.

Non-compensation reasoning argues that the mass and velocity differences do not compensate to make the momenta of the carts equal on the given correct variation of the Change in Momentum question. In the results, this reasoning was only seen in response to this variation.
Finally, unequal elements reasoning argues that, because each cart has a different mass and a different velocity, the kinetic energies of the carts cannot be the same. This reasoning was only seen in response to the consider equal to variation of the Change in Kinetic Energy question and was only used to justify the response NOT equal to.

Because all of these types of reasoning focus on how the same mass and velocity differences compare in order to come to a mathematically imprecise conclusion about the final state of the two blocks, they can be considered as highly related types of reasoning. In addition, because these highly related types of reasoning are specific to or used significantly more prevalently on a particular question variation, they can be viewed as variation-specific adaptations of the same underlying set of ideas.

This finding provides further insights into what students are doing when they use compensation reasoning and shows that compensation reasoning is likely not just one fixed type of reasoning students have, but is more likely an idea that is applied to justify a particular response and can take on various forms depending on what response the students are attempting to justify.

This again demonstrates the variability of students’ ideas, as discussed in Section 2.5; in the terms of the resources framework, these ideas could consist of many of the same resources assembled into different knowledge structures to support different responses.

Further study of the connections between these types of reasoning and the responses they are used to justify can help explain more about how students develop reasoning by showing how the same set of ideas is influenced by features of the question variations.

Future research should also determine why students use compensation reasoning to both select a response and justify a provided correct response, but only use non-compensation or unequal elements reasoning to justify a provided correct
response. Such research may provide insights into how to dissuade students from using compensation reasoning, which is often considered a non-productive type of reasoning [23]. Further understanding about why students choose not to use particular types of reasoning with responses they have selected, even though those types of reasoning appear to be accessible, will enhance research investigating how students arrive at answers to physics questions.

9.8 Reasoning Based on the Provided Outcome

This section proposes that a few types of reasoning observed are instances of students using the provided outcome on the given correct variation as the basis for, or as an explicit part of, the reasoning they develop. First, three examples of types of reasoning demonstrating this are presented; then, the discussion examines how the given correct variation enables these types of reasoning.

Two previously discussed examples of types of reasoning in which students use the outcome provided by the given correct variation as part of their reasoning are compensation reasoning observed on the Change in Kinetic Energy question (discussed in Sections 7.4.1.1 and 7.5.1.2) and non-compensation reasoning used on the Change in Momentum question (discussed in Section 9.4.3). Both of these types of reasoning are compensation-like and are presented in the previous section.

In the case of compensation reasoning, our results show that students use this type of reasoning more prevalently on the given correct variation. This may be due to them using the provided outcome, which states that the kinetic energies of the two carts end up being equal, to claim that the mass and velocity must compensate in a way that leads to equal kinetic energies.

In the case of non-compensation reasoning, it appears that many students may be trying to reconcile $p = mv$ with the fact that the momenta are unequal by stating that the mass and velocity must be such that they do not compensate.
In both of these cases students appear to be using the information provided in the outcome to justify a mathematically imprecise argument.

Another example, which has not been previously discussed, is from the Pendulum question on Pretest 3. In the given correct variation of this question, shown in Figure B.8, students were asked to justify why a pendulum bob had an upwards acceleration at the bottom of its swing. Some students reasoned that the acceleration was upwards because the change in velocity was upwards.

Some of these students may be reasoning from the provided outcome—namely, that acceleration is upwards—to determine that the change in velocity is upward, providing that as their reasoning without a correct understanding of the dynamics of the physical situation. Students are being told \( \vec{a} \) is upwards; if they know \( \vec{a} = \frac{\Delta \vec{v}}{\Delta t} \), it is relatively simple to figure out that \( \Delta \vec{v} \) must be upwards as well. Once they have put this together, they can say that the reason the acceleration is upwards is because the change in velocity is upwards, without analyzing the motion of the pendulum. This reasoning is valid for any physical situation in which acceleration is up, and the lack of connection between the physical situation and the reasoning, compared to a direct connection between the outcome and students’ reasoning, suggests that this reasoning may be developed based on the outcome instead of the physical situation.

These types of reasoning again strongly suggest that some students, on some given correct variations, are using the provided outcome as the basis for, or as part of, the reasoning they develop as seen previously in Section 7.5.1.2.

The given correct variation is different from other variations where students must either generate reasoning and pick a response or pick a response and then attempt to justify it without knowing if it is the correct response. When designing the given correct variation, it was intended to filter out incorrect types of reasoning students generate about the physical situation and potentially prompt them to generate other types of reasoning about the physical situation. However, the given
correct variation also clearly allows—and possibly prompts—students to develop reasoning based on the provided outcome. In hindsight, it is not surprising to see signs of reasoning being developed in this way, but it was not anticipated in the design of the given correct variation.

Although no direct evidence demonstrates that students are using the provided outcome as part of their reasoning, this potential was discussed for several reasons.

First, other researchers who use this variation should be aware of the availability of this different way of developing reasoning. If students are able to arrive at correct reasoning based on the outcome, the number of students using correct reasoning on the given correct variation could indicate that students have a better conceptual understanding of the situation than they actually do. Students could use the provided outcome in response to any given correct variation, and it is something that should be taken into account as part of the design and analysis of given correct variations.

Second, being able to identify reasoning based on an outcome is potentially beneficial to future research into how students develop responses to questions. Looking for features of reasoning that indicate they are developed based on a provided outcome can help identify if students use similar methods to develop their reasoning even when a response is not provided. A lack of connection between the physical situation and reasoning may be one indication that reasoning is being developed from a specific outcome.

9.9 Unique Reasoning to Justify Why Something is Incorrect

One of the research questions posed in this dissertation asked if students would develop or use unique types of reasoning to justify why responses were incorrect. Students do in fact do this, and several instances of students using unique types of reasoning to justify why a response is incorrect were observed. However, these types
of reasoning appear to either be adaptations of types of reasoning that are used on other variations or reasoning that is less specific than types of reasoning used on the original version.

This section summarizes all three examples of unique reasoning used to eliminate a response observed in the across-content study.

On the eliminate variation of the Pendulum question, discussed in Section 9.4.1 and shown in Figure B.6, students used velocity must change reasoning to explain that the acceleration is not zero because the velocity of the ball must change as it moves.

This reasoning is correct, but due to the nature of the eliminate variation, it is less thorough than correct reasoning on the original version as students only needed to state that the velocity changes, but did not need to specify the direction of change. Students who responded with correct reasoning on the eliminate variation may not have been able to correctly determine the velocity’s direction of change to provide correct reasoning on the original version.

On the eliminate variation of the Work-by-Wall question, discussed in Chapter 6 and shown in Figure B.30, students used displacement means work reasoning. This reasoning argues that, because there is a displacement, there must be some work being done. This type of reasoning is an incorrect application of the definition of work, which also requires a non-perpendicular force in order for work to be performed on an object.

As discussed in Section 6.5.3, the presence of this unique reasoning and similar force means work reasoning demonstrates that a third of the class was not taking into account that both force and displacement are necessary for work to be done, indicating that these students did not have a correct understanding of the definition of work. As with velocity must change reasoning, these types of reasoning that only discuss either force or displacement are simpler types of reasoning than the correct
reasoning on the *original* version, which requires both force and displacement to be discussed.

In general, *eliminate* variations require less specific correct reasoning than the *original* version. Eliciting these less specific types of correct reasoning does not provide as much information about students’ ideas as eliciting correct types of reasoning on the original version as students may be able to provide correct reasoning on the *eliminate* variation, but unable to provide correct reasoning on the *original* variation.

This finding reinforces that the *eliminate* variation is useful for learning more about students’ ideas when paired with an *original* version, but is likely not an equivalently useful tool when used on its own.

Finally, on the *consider equal to* variation of the Change in Kinetic Energy question, discussed in Chapter 7 and shown in Figure B.27, students used unique *unequal elements* reasoning to explain that the kinetic energies of the two carts are not equal after they have traveled an equal distance because the mass and velocity of the carts are both different. As discussed in Section 9.7, this type of reasoning is an adaptation of *compensation* reasoning being applied to eliminating a response, and both of these types of reasoning have similar complexity.

Overall, these unique forms of reasoning specific to the elimination of a particular response seem to fall into two categories. Adaptations or different expressions of incorrect reasoning similar to types of reasoning are present in other variations, as is the case with *unequal elements* reasoning. Less complete forms of correct reasoning are evident, as with *velocity must change* and *displacement means work* reasoning.

In cases where students are using less complete types of reasoning, their responses are consistent with the *resources* framework. In the terms of this framework, students may only be activating a smaller number of resources when
developing a knowledge structure in response to these variations because they do not demand as thorough of a knowledge structure as the original version.

Learning how students adapt their reasoning when not as much precision is demanded of them allows for further investigation into how students develop reasoning and analyze physical situations. The findings show that some students are capable of expressing correct simplified reasoning when it is appropriate whereas other students express simplified reasoning that is incomplete or incorrect. Further research can use these question formats, or other targeted variations, to investigate how student reasoning changes when less precision is demanded in their responses as well as what leads to correct or incorrect simplification.

These results also help better define the difficulties students are having by eliciting these unique types of reasoning. For example, seeing students use displacement means work reasoning directly indicates that many students think displacement alone is enough to conclude that work is being done, which is not something expressed as clearly or as commonly in response to the original version of the question.

Finally, this result answers one of the research questions by showing that students sometimes use unique reasoning to eliminate responses and that the eliminate question variation is capable of eliciting alternate ideas from students. This result also provides support for theoretical frameworks that describe students’ ideas as assembled on the fly, as similar reasoning is expressed in different ways in response to these question variations.

9.10 Conclusion

In looking at the general results for all variations, not many differences in response prevalence were observed, but unique reasoning and differences in reasoning prevalence were common.
Table 9.12. A summary of the results of all variations from the across-content study. This shows how many results were found on each variation when looking for response differences, unique reasoning, or informative differences in reasoning.

<table>
<thead>
<tr>
<th></th>
<th>Eliminate (6)</th>
<th>Consider (15)</th>
<th>Given Correct (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Difference in Response</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest Showing Result</td>
<td>3 (50%)</td>
<td>1 (7%)</td>
<td>-</td>
</tr>
<tr>
<td>Total Number of Results</td>
<td>4</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td><strong>Unique Reasoning</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest Showing Result</td>
<td>2 (34%)</td>
<td>1 (7%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Total Number of Results</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Informative Differences</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest Showing Result</td>
<td>3 (50%)</td>
<td>2 (13%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Total Number of Results</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

This conclusion summarizes each variation individually as well as some of the major findings in this chapter. The results from all variations are shown again in Table 9.12.

### 9.10.1 Summary of Results from Eliminate

Several cases were presented where asking an `eliminate` variation provided additional information about students’ ideas. In general, it seems that, if used correctly, the `eliminate` variation has the ability to consistently elicit additional information about students’ ideas.

The `eliminate` variation appears to be useful for finding unique types of student reasoning, such as `displacement means work` and `velocity must change` reasoning.

The `eliminate` variation can find differences in the prevalence of particular types of reasoning when compared to the original version. The primary example here is the more prevalent use of `otherwise` reasoning on the `eliminate` variations from the Car-Over-Hill and Gymnast questions.

Being able to observe these differences provides more information about students’ ideas by showing that more students may have access to them than the
original version indicates. It also shows which types of reasoning students use to justify a response, eliminate a response, or do both.

Another useful aspect of the eliminate variation is that it helps identify different profiles of possible responses. Knowing if a classroom of students has reasoning for why a response is correct and incorrect, rather than only having reasoning for why a response is correct, can provide useful information for instructors. For example, if students do not have any ideas about why the correct response is incorrect, it may not require much effort to change students’ ideas to be consistent with a correct understanding; however, if students do have ideas about why the correct response is incorrect, more thorough instruction and examples may be needed. Knowing which responses students have reasoning to support being correct and/or incorrect can provide additional content knowledge for teachers by providing specific examples of the types of reasoning students are using to argue against the correct response as well as offering an overall picture of how students view each response’s correctness and incorrectness. In addition, identifying a response that a class of students has reason to believe is correct and incorrect can provide an opportunity for a rich peer-instruction activity or a whole-class discussion.

9.10.2 Summary of Results from Consider

The overall trend observed when looking at the consider variation was that it very rarely elicited different responses from the students.

Overall, few results from the consider variation were observed: one instance of a difference in responses, one instance of unique reasoning, and three out of fifteen pretests with differences in the prevalence of particular types of reasoning. Furthermore, there was no indication that having a consider variation target a correct or incorrect response had an effect on whether results emerged or not.
This lack of differences is likely due to this question not accessing different students’ ideas than the original version. It was initially predicted that using the consider variation to focus students’ attention on particular responses would lead to them using different ideas to develop their reasoning, but this seems not to have been the case.

One reason the consider variation would not access different ideas than the original version is that the two questions are highly similar. The consider variation was the least different from the original version as it still had one response option in common with the original version that students could eliminate. The similarity of this task to the original version could have caused students to effectively ignore the differences between these two tasks or use their familiarity with the original version to transform the consider variation back into an original version, thereby having these two question versions access the same ideas.

9.10.3 Summary of Results from Given Correct

Overall, differences from the original version were identified in four of the ten given correct variations, with two of these differences being unique types of reasoning and two being informative differences in reasoning.

In the instances where no such results emerged, usually one or two particular kinds of reasoning were used on both versions by a majority of the class alongside expected differences in reasoning.

The differences identified indicate that students have access to other reasoning about these physical situations that they may not use if allowed to choose their own response. By forcing students to justify a particular response, it is possible to elicit alternate ideas about the physical situation. Being able to elicit these alternate ideas, even if they are not correct, provides valuable information about students’ ideas. For example, the given correct variation can show if students are able to
access the correct reasoning more often than the *original* version indicates, as was the case with *no displacement* reasoning; it can also show if incorrect types of reasoning are more common among students than the *original* version indicates, as was the case with *compensation* reasoning. Finally, it can indicate if students have part of the correct reasoning, but not the whole picture, as with *equal applied force* reasoning.

### 9.10.4 Summary of Results Spanning Variations

In addition to learning about the effectiveness of each question variation, four results that spanned multiple variations were presented.

The most prominent of these results was that the responses *zero* and *equal to* were always targeted response options when an informative difference was found on a *consider* or *given correct* variation. This result runs counter to prior research showing that students are biased away from selecting *zero*. It can be proposed that the importance of *zero* and *equal to* are due to ideas related to these responses being readily associated with matching quantities and developed into reasoning or their nature as non-directional, discrete, and uncommon response options.

Clear evidence highlighted that students are able to generate reasoning that is specific to justifying why a response should be eliminated, but these types of reasoning are either adaptations of incorrect reasoning that is similar to reasoning present on the *original* version or less specific forms of correct reasoning. Being able to find these unique types of reasoning provides insights into how students reduce their reasoning to simpler types of reasoning and also provides additional information about the difficulties students are having.

Furthermore, students may be basing their reasoning on the provided outcome when responding to *given correct* variations. This is a way of developing reasoning that the *given correct* variation allows for and possibly prompts. Researchers should
be aware that the amount of correct reasoning used on the *given correct* variation may not correlate to students’ understanding of the physical situation if students are commonly developing their reasoning based on the provided outcome.

Finally, connections were found between multiple types of *compensation*-like reasoning. These types of reasoning all made very similar mathematically imprecise arguments, but were used to justify different responses on different variations. This provides evidence that these types of reasoning are not simply one specific type of difficulty, but are adaptable to multiple responses.
CHAPTER 10
CONCLUSION

This dissertation presented a methodology for investigating alternate ideas, elucidated the alternate ideas students have in many content areas, and studied how students responded to the question variations in our methodology. We found that all of the variations used were able to elicit alternate ideas with varying levels of success and identified patterns in student reasoning that spanned content areas or variations.

In this chapter we present answers to our research questions by summarizing our findings about alternate ideas, and then summarizing our findings about the question variations we developed. We then discuss the utility and applications for each question variation and make suggestions for their use. Finally, we provides suggestions for future research.

10.1 Findings on Alternate Ideas

Through asking question variations over a wide range of content areas we were able to perform a broad study of the alternate ideas students possess. As presented in Chapter 1, our research questions focused on investigating the prevalence, and types of alternate ideas students have. In this section, we summarize how often we found students using different sets of ideas, and what different types of reasoning those sets of ideas lead to.
The research questions we aim to address here are:

- Do these question variations elicit alternate ideas? Is our methodology able to consistently identify alternate ideas, or do they need to be targeted with content-specific question modifications?

- How prevalent are alternate ideas across many physics content areas?

- What alternate ideas do we observe in each content area?

The research in this dissertation has shown that alternate ideas can be elicited by constraining response options through this set of question variations. As summarized in Section 9.10, we found evidence of students using alternate ideas by looking for unique reasoning or informative differences in reasoning, which both indicate that students are using alternate ideas in response to the question variation. In total, we identified five types of unique reasoning and nine informative differences in reasoning across all 31 administrations of a question variation.

A primary reason we did not find unique reasoning or informative differences on most administrations of question variations in this study is that it was an exploratory study with low power that was designed to get a survey of results, rather than find the majority of them, so these variations are likely more capable of identifying alternate ideas than the results of this dissertation indicate. Furthermore, part of what led to most administrations not finding differences in reasoning may have been that the original version of questions used are well-designed research questions that target particular types of reasoning; because they are well designed, making variations may not change the types of reasoning the physical situation and question target.

To show the types of reasoning that provided evidence of alternate ideas we present a summary of the unique types of reasoning we identified, and the informative differences in reasoning we observed.
10.1.1 Unique Reasoning

Unique Reasoning - Consider - Unequal Elements Reasoning  
On the consider equal to variation of the Change in Kinetic Energy question from Pretest 9a (Figure B.27), unequal elements reasoning was used to justify that the kinetic energies of two carts of unequal mass were not equal after being pushed by the same force for the same distance because they ended up with both different masses and different velocities. This reasoning seems to be a form of compensation reasoning that is adapted to arguing for why the results are not equal, as discussed in Section 9.7. This result was previously discussed in Section 7.5.2.

Unique Reasoning - Eliminate - Velocity Must Change Reasoning  
On the eliminate variation of the Pendulum question from Pretest 3 (Figure B.6), velocity must change reasoning was used to argue that the acceleration of a the pendulum bob at the bottom of its swing was not zero. This type of reasoning appears to be a more vague version of the correct non-uniform circular motion reasoning often seen in response to the original version of the question. This result was previously discussed in Section 9.4.1.

Unique Reasoning - Eliminate - Displacement Means Work Reasoning  
On the eliminate variation of the Work-by-Wall question from Pretest 9b (Figure B.30), displacement means work reasoning was used to argue that work was being done on the block because it was undergoing displacement. This type of reasoning and non-unique force means work reasoning together suggests that some students had the view that either force or displacement is all that is necessary for work to be done. This result was previously discussed in Section 6.5.3.

Unique Reasoning - Given Correct - Asymmetry Reasoning  
On the given correct variation of the Gymnast question from Pretest 6 (Figure B.18), asymmetry
reasoning occurred as students used asymmetry arguments to justify why the
tension in one rope supporting a gymnast was greater than the tension in another (see Section 8.6).

**Unique Reasoning - Given Correct - Non-Compensation Reasoning**  
On the given correct variation of the Change in Momentum question from Pretest 10 (Figure B.34), non-compensation reasoning was used to argue that the different masses and velocities of two carts pushed by the same force for the same distance did not compensate to make their momenta equal, as discussed in Section 9.4.3. This reasoning seems to be a form of compensation reasoning that is adapted for arguing why two results are not equal, as discussed in Section 9.7.

**10.1.2 Informative Differences**

**Informative Difference - Consider - More Velocity Reasoning**  
More velocity reasoning was used less prevalently on the consider equal to variation of the Change in Kinetic Energy question from Pretest 9a (Figure B.27). This reasoning explained that the kinetic energy of one cart is greater than that of the other because it has more velocity. This difference was most likely an effect of the presence of unequal elements reasoning, as previously discussed in Section 7.5.2.

**Informative Difference - Consider - Velocity is Zero Reasoning**  
Velocity is zero reasoning was used more prevalently on the consider zero variation of the Simple Harmonic Oscillator question from Pretest 13 (Figure B.42). In velocity is zero reasoning, students stated that it was not the acceleration of an oscillating block at maximum displacement that was zero, but rather the velocity that was zero. The more prevalent usage of this type of reasoning suggests that some students will simply identify a quantity that is zero and then use that identification as their reasoning. This result was previously discussed in Section 9.5.4.
Informative Difference - Consider - Equilibrium Reasoning  

Equilibrium reasoning was used less prevalently on the consider zero variation of the Simple Harmonic Oscillator question from Pretest 13 (Figure B.42). Equilibrium reasoning states that the acceleration of a block undergoing simple harmonic motion will point toward the equilibrium position. The lower prevalence of equilibrium reasoning might be due to a higher prevalence of other types of reasoning, such as velocity is zero reasoning, which are more easily cued by a variation asking about the response zero. This result was previously discussed in Section 9.5.4.

Informative Difference - Eliminate - Otherwise Reasoning  

Otherwise reasoning was used more prevalently on the eliminate variation of the Car-Over-Hill question from Pretest 4 (Figure B.10). Otherwise reasoning is used to argue that the gravitational force exerted by the earth on a car cresting a hill is equal to the normal force exerted by the ground on the car, because otherwise it would sink into the hill or rise into the air. This result suggests that many students are able to use this process of elimination type reasoning to justify why a response is not correct, but prefer to use other arguments on the original version. This result was previously discussed in Section 9.5.1.

Informative Difference - Eliminate - Otherwise Reasoning  

Otherwise reasoning was used more prevalently on the eliminate variation of the Gymnast question from Pretest 6 (Figure B.17). Otherwise reasoning is used to argue that at the bottom of the swing, the magnitude of the upward net-force has to be equal to the magnitude of the weight of a gymnast; otherwise, the gymnast would accelerate upward or downward. Again, this result suggests that many students are able to use this process of elimination type reasoning to justify why a response is not correct, but prefer to use other arguments on the original version. This result was previously discussed in Section 9.5.1.
Informative Difference - Eliminate - Force Means Work Reasoning

*Force means work* reasoning was used more prevalently on the *eliminate* variation of the Work-by-Wall question from Pretest 9b (Figure B.30). In *force means work* reasoning, students argue that a wall does non-zero work on a oscillating block-spring system because there is a force acting on the block. This demonstrates a difficulty students have correctly applying the definition of work and adds to the understanding of students’ difficulties with this topic. This result was previously discussed in Section 6.5.3.

Informative Difference - Given Correct - No Displacement Reasoning

*No displacement* reasoning was used more prevalently on the *given correct* variation of the Work-by-Wall question from Pretest 9b (Figure B.31). This reasoning explains that the wall does not do work on the attached block-spring system because the wall does not undergo a displacement. This is correct reasoning, and this result shows that many students are able to provide correct reasoning when the correct outcome is provided for them. This result was discussed in Section 6.5.1.

Informative Difference - Given Correct - Compensation Reasoning

*Compensation* reasoning was used more prevalently on the *given correct* variation of the Change in Kinetic Energy question from Pretest 9a (Figure B.28). Here, *compensation* reasoning was used to argue that the kinetic energies of the two carts of different mass were the same after being pushed by the same force for the same distance because the mass and velocities of the carts compensated in a way that made the kinetic energies equal. This result suggests that students are using the provided outcome as part of their reasoning. This result was discussed in Section 7.5.1.2.
Informative Difference - Given Correct - Equal Applied Force Reasoning

*Equal applied force* reasoning was used more prevalently on the *given correct* variation of the Change in Kinetic Energy question from Pretest 9a (Figure B.28). *Equal applied force* reasoning argues that, because the two carts are pushed with the same force, they must have equal kinetic energies. We suggest that the higher prevalence of this reasoning stems from the information that the two carts are pushed by the same force being more salient than the information that the carts are pushed for the same distance, which allows students to identify the forces being the same more readily and leads to a more prevalent usage of this type of reasoning on the *given correct* variation. This result was discussed in Section 7.5.1.1.

10.2 Findings About Question Variations

In addition to investigating students ideas, the work in this dissertation also investigated how students responded to the question variations we developed. In this section, we discuss the findings that address our research questions about question variations.

Do certain content areas or question features more commonly elicit alternate ideas in combination with these question variations?

We found two question features that more often elicited alternate ideas. In summarizing the results of all question variations it was found that informative differences in reasoning were only found on *consider* or *given correct* variations when the responses *zero* and *equal to* were targeted. The effects of *zero* were discussed in more detail in Section 9.6. We posit that this difference is due to the fact that *zero* is not commonly used or due to the nature of zero as a discrete and direction-less quantity.

This finding shows that students react differently to these responses and that these question features do more commonly lead to the elicitation of alternate ideas.
How do students respond to a question that is not asking them to provide a correct response? Do they typically try to give the right response anyway? To what extent can students explain why an incorrect response is incorrect?

The results of this dissertation showed that students can develop unique reasoning when they are justifying why a response is incorrect. This demonstrates that students are able to come up with types of reasoning specific to eliminating responses. However, these types of reasoning commonly appeared to be adaptations of or less specific types of reasoning observed on the original version. Additionally, we did see some students provide the response they viewed as correct, and often used that response in place of reasoning. These results were discussed in more detail in Section 9.9.

Another result specific to students eliminating responses were conflicted responses discussed in Section 9.3.1. There were four instances of conflicted responses, where a large portion of the class selected as correct and eliminated the same response. Both negative and zero were conflicted responses on the Work-by-Wall question shown in Figure B.30 and discussed in Section 6.5.2. Students were also conflicted about the response equal to from the Newton’s 3rd Law question on Pretest 5, shown in Figure B.14, and the response greater than from the Gymnast question on Pretest 6, shown in Figure B.17. Identifying conflicted responses indicates that some students may have ideas to support or argue against a particular response and potentially provide good opportunities for class discussion.

How do students respond to a question that is providing them with a correct response? Do they often have or develop reasoning to support it? Students can sometimes come up with the correct reasoning when provided with the correct outcome. This was discussed in Section 6.5.1 we saw more students use correct no displacement reasoning on the Work-by-Wall question to explain why the
work done by the wall attached to a block-spring system undergoing simple harmonic motion does no work. However, students more commonly use incomplete reasoning more prevalently. This was seen in Chapter 7 where the Change in Kinetic Energy question asked students to compare the kinetic energies of two carts after they were pushed with the same force for the same distance. Here we saw students more commonly using *Equal applied force* reasoning to argue that the two blocks had the same kinetic energy because they were pushed with the same force, or using *compensation* reasoning to argue that the kinetic energies were the same because the mass and velocity differences of the carts balanced out. Both of these types of reasoning are incomplete. We suspect, based research by Heckler and Bogdan [45], that these differences may be due to the *given correct* variation causing changes in the accessibility or availability of students’ ideas.

Additionally, we saw students respond to the *given correct* variation by disagreeing with the provided response and treating it as anomalous data, as discussed in Chapter 8. Students disagreed with many different provided outcomes across content areas, and the rate of disagreement appears to be correlated with question difficulty. This result was unexpected, but provides further insights into how students make sense of anomalous data as well as evidence to support the framework developed by Chinn and Brewer [17].

Finally, we suspect that students sometimes used the provided outcome on the *given correct* variation as a basis for their reasoning. This was discussed in Section 9.8, which described two instances of students using *compensation*-like reasoning to argue for the provided outcome and students equating $\vec{a} = \frac{\Delta \vec{v}}{\Delta t}$ to state that $\Delta \vec{v}$ is in the same direction as $\vec{a}$ without connecting their reasoning to the physical situation. This was not an anticipated result, leading to questions about whether students reason from instinct or intuition in similar ways when responding to the *original* version.
Do all the variations lead to similar types of responses, or does the amount the responses are constrained matter?

The variations had varying levels of success. One overall trait identified affecting how prevalently a variation elicits alternate ideas is how confronting it is, which is tied to its similarity to the original version. The consider variation was most similar to the original version of the question because it gave students a choice of responses to justify and was the least effective variation. The given correct variation was the most confronting and most distinct from the original variation because it did not give students an option of response and showed the highest prevalence of alternate ideas. Therefore, the more confronting or different from the original version a variation is, the better it will be at eliciting alternate ideas.

10.3 Variation Utility, Applications, and Suggestions for Use

We developed these question variations to be a tool that other researchers and instructors could also use to investigate student ideas. All of the question variations were able to elicit unique types of reasoning and different distributions of students’ reasoning than the original version, but—as discussed in Chapter 9—they did so with varying levels of success. In this section we summarize the utility and potential applications of these research questions, along with suggestions for how to use them. We present this information in order to help other researchers or teachers use these questions more effectively in their work. This discussion is based on both the results and design of the variations.

10.3.1 Consider Utility and Applications

The consider variation found results the least often of all variations. A primary explanation for why the consider variation is not as useful for accessing alternate

1The lack of results on the consider variation and its similarity to the original version are discussed in Section 9.5.6.
ideas is that it is too similar to the original version. In this variation, students can still choose a response as correct and justify why it is correct, similar to the original version. In addition, unlike the eliminate variation, there is only one response that can be eliminated. This similarity may prevent students from engaging with this question in a way that is different from how they engage with the original version, leading to the same information the original version would, as discussed in Section 9.5.6. This is consistent with research on the question format, which says that well-designed multiple-choice and free-response questions get similar results [88].

Furthermore, the consider variation only showed interesting differences when it targeted zero or equal to, so its lack of effectiveness may be tied to the low number of variations that targeted those responses. However, it did not show results as prevalently as the given correct variation, even when targeting zero or equal to; thus, the targeted response is likely not the only factor contributing to the lack of results seen on this variation.

The one advantage to using the consider variation is that it is capable of gathering a lot of student reasoning about one particular response—a feature inherent to its design. The ability of the consider variation to target one response makes it a potentially useful tool if a researcher or instructor is interested in students’ ideas about a particular response rather than the physical situation in general, although this variation is only occasionally able to identify differences in reasoning when compared to the original version.

Finally, the consider variation was able to identify instances where students were using unique reasoning to justify why responses were incorrect.
10.3.2 Eliminate Utility and Applications

The eliminate variation was often able to elicit different information than the original version of the question by asking students to explain why an outcome does not occur. Sometimes student explanations on the eliminate variation are unique to that variation, and other times the reasoning is also present in the original version, but with a significantly different prevalence.

Eliciting these types of reasoning can be an instructional tool to stimulate discussion and can help researchers develop curricula to improve student learning in many content areas.

If students are using incorrect reasoning to eliminate the correct response, an instructor can try to address that difficulty or ask students in the course to address why that type of reasoning is incorrect. If students are providing correct reasoning to eliminate an incorrect response, the instructor can share that reasoning with the class, essentially using peers’ ideas to explain to the class why something does not occur. Although these are not new practices, the use of an eliminate variation to elicit ideas about why responses are incorrect can make this a more student-centered or interactive approach.

An instructor can also make sure that students know not only the correct response, but also why other responses are incorrect, which may represent a more thorough understanding of the content. Examples of students using incorrect reasoning to eliminate the correct response and using correct reasoning to eliminate an incorrect response are discussed in Section 6.5.3.

Having some students justify the elimination of a particular response and others justify their choice of a correct response can identify responses a class is conflicted about. Knowing what responses a class has ideas to support and ideas to reject as well as what those ideas are is useful instructional information. This could be applied in an instruction environment via clickers or other methods in order to start
a peer discussion about a conflicted response, knowing that students in the class have ideas for both supporting and eliminating that response. This would help the class reinforce a correct physics understanding and address incorrect ideas at the same time.

Asking both original and eliminate versions also allows researchers to see what types of reasoning students can use to justify why a response is correct or incorrect. Learning more about the types of reasoning students use to answer both of these versions can help researchers measure the entrenchment of students’ ideas and learn more about how students develop responses to conceptual physics questions. For example, otherwise reasoning was used on both the original and eliminate versions, as discussed in Section 9.5.1.

Overall, multiple types of unique reasoning, and several differences in the usage of non-unique reasoning were found when administering the eliminate variation. The information found when asking an eliminate variation is particularly useful, especially when combined with the original version.

10.3.3 Suggestions for Using Eliminate

There are a couple of things to be aware of when administering an eliminate variation.

One consideration for using an eliminate variation is that students respond with a wider variety of reasoning than they do on the original version, as evident on the Pendulum question from Pretest 3; and this large variety of reasoning can potentially make it harder to find statistically significant differences in reasoning between variations. This large variety of reasoning may be due to the fact that the original version of the question was designed with the goal of investigating particular students’ ideas. When changed to an eliminate variation, it may no
longer target the same ideas or may target them less effectively, leading to the elicitation of many different ideas and no significant differences in reasoning.

In addition, unless the goal is to identify obviously wrong distractors, it is important to not have any in the multiple-choice question being modified. Having a obviously wrong distractor will cause the majority of students to explain why that one particular response is incorrect and will not provide information about what they think of the other responses. This is analogous to asking an original version of a question where it is too easy, resulting in the majority of students discussing only the correct response.

Another thing to be cautious of is creating eliminate variations of questions where the most common reasoning to justify why something is correct is also good reasoning for eliminating a particular response. If this is the case, as discussed in Section 9.5.3, the majority of students will likely eliminate that one response and use that same reasoning, which will not provide information that is different from the original version.

10.3.4 Given Correct Utility and Applications

The given correct variation was the most effective at eliciting alternate ideas from students. These results occurred when the responses zero and equal to were targeted, as was seen with the consider variation, but additional results were seen where students treated the provided outcome as anomalous data. The given correct variation was the most confronting, providing students with no choice of response to discuss, which is likely the reason it was the variation that most successfully elicited alternate ideas.

When eliciting more common correct reasoning, the given correct variation can demonstrate that more students have access to the correct reasoning than the original version indicates.
The single example of the given correct variation eliciting correct reasoning more prevalently was the use of no displacement reasoning discussed in Section 6.5.1. Demonstrating that correct reasoning can be elicited when the correct outcome is provided is useful information for instructors because it allows them to identify correct ideas their students possess that can be built upon to help the students develop a correct understanding of the content. In addition, knowing that students have correct reasoning, but are not using it, may indicate that students will be able to learn the material more easily or that the incorrect types of reasoning students are using instead may need to be addressed directly. These results do not indicate whether building upon correct ideas or addressing incorrect ideas is likely to be more effective; this would be a fruitful area for future research.

However, the given correct variation more commonly elicits incorrect reasoning, as seen in Sections 7.5.1.2 and 7.5.1. These results suggest that students will sometimes use inappropriate reasoning to justify a correct response, which further supports the notion that understanding the reasoning behind students’ choice of response is important for assessing their understanding. More specifically, results from given correct variations that show more common usage of incorrect reasoning indicate to an instructor to focus on helping students develop a correct set of ideas that they appear to be lacking, explain why the incorrect reasoning used is incorrect in this case, or transition students from the incorrect type of reasoning to a correct one. In these ways, knowing more about the incorrect ideas students are using to justify the correct response is valuable knowledge for an instructor.

The given correct variation can also be used as a tool to investigate how students develop responses to conceptual physics questions. Administrations of the given correct variation found several results that provided insights into how students develop reasoning to conceptual physics questions.
Chapter 8 discussed how students treat the data provided to them in a \textit{given correct} variation as anomalous data, often disagreeing with or reinterpreting the data provided by the question. This shows that students’ reactions to anomalous data are an important factor in the reasoning they develop to answer \textit{given correct} variations. Knowing this can help find instances where students’ incorrect ideas appear to be entrenched and may be causing them greater difficulty in developing a correct understanding of the content.

The similarity between the \textit{given correct} variation and the \textit{original} version also enables researchers to draw conclusions about how providing students with a response affects the reasoning they use. As discussed in Section 7.5.1.2, students used \textit{compensation} reasoning more prevalently when justifying the provided response. The fact that simply providing the response for students changes the reasoning they use to justify the response suggests that students have different approaches for developing reasoning in these cases and shows that the process of selecting a response may be an important part of developing an argument to justify it. This result is similar to the effect of providing students with a solution method, as discussed in Section 2.5.3, which discusses the work done by Heckler [16]. This result is likely to have occurred because, as discussed in Section 9.8, students may be using the outcome as part of the reasoning they develop. Further research using these question versions along with larger sample sizes and interviews may be able to identify how students are affected by the provided response and what effects that provided response has on their reasoning.

These two examples of findings about how students develop responses to conceptual physics questions are just what was observed in this analysis. The \textit{given correct} variation may be able to produce other findings of this nature through additional administrations or analysis methods.
10.3.4.1 Suggestions for Using Given Correct

There are a few challenging aspects of administering *given correct* variations. Questions of appropriate difficulty must be used to enable the *given correct* variation to elicit alternate ideas. If the question is too easy, the reasoning students provide will likely be identical to that provided in response to the *original* version because the students are not being provided with any new information. If the question is too difficult, students may disagree with the provided response and not provide reasoning for why they think that response is correct. The level of medium difficulty where *given correct* is effective seems to occur when between 33% and 66% of a class is able to select the correct response on the *original* version.

Students may also reinterpret the *given correct* variation, as discussed in Section 8.6.2. Therefore, the physical situation should be described as clearly as possible to avoid reinterpretation.

Finally, students occasionally used the provided outcome as part of their reasoning, as discussed in Section 9.8. An analysis of the results from *given correct* variations should be performed with an awareness that these types of responses may exist.

10.4 Future Research

This section presents several follow-up studies that the results of this dissertation indicate would be fruitful areas of research with the potential to improve instruction, provide more insights into how students develop responses to questions, and test theoretical frameworks.

**Applications to Instruction** One important follow-up study to perform is to apply these question variations or the results of this dissertation to improve student learning. It was postulated that the alternate ideas these variations discovered
would enhance the understanding of the ideas students have in many content areas, but follow-up work should be done to confirm that the ideas these variations elicit are useful when developing improvements to instructional methods or curriculum. Showing this would establish these variations as effective tools for improving learning.

Specifically, it would be interesting to answer the following questions:

- Can alternate ideas be used to improve instruction or curriculum?
- Do class discussions about conflicted responses lead to good conversations? Do these conversations improve learning?
- Does utilizing alternate ideas in instruction improve scores on the original version of the task or just the variations?

**Investigating Given Incorrect**  As each variation used in this dissertation had some level of success eliciting alternate ideas, future research could investigate the use of the *given information* variation to verify that it also elicits alternate ideas as expected. The *given information* variation, presented in Section 3.3, asks students to justify why a provided outcome is incorrect. It would also be interesting to see if students treat the provided outcome as anomalous data on this variation as well as the *given correct* variation. If students do treat the provided outcome as anomalous data, it could potentially create a way to measure the entrenchment of incorrect reasoning that supports particular incorrect responses. This would allow for a better understanding of the incorrect and correct ideas students have and provide a way to gather information about how well students are able to justify why incorrect responses are incorrect.
Investigate Zero and Equal To One of the major findings of the research in this dissertation was that alternate ideas were found mostly when the responses zero and equal to were targeted. This implies that these are responses that students think about differently from other responses. Although this research does not clearly show what makes these responses different, several reasons students may be treating them differently can be postulated. They are discrete, are direction-less, and have the multiplicative properties of zero (or no change). Future research should be performed to investigate if students treat these responses differently from more common responses on the original version of research tasks or for other question formats, such as free-response. Research should also explore which features of these response options cause students to treat them differently, even if they only do so when responding to question variations. This area of research is particularly important given the prevalent use of research tasks with zero and equal to as response options in the PER literature [23, 24, 26, 59].

Follow-up Interviews on Conflicted Responses In a couple of instances, large portions of a classroom of students had ideas that justified or eliminated the same response, which suggests that individual students may have reasoning that both supports and counters that response [56]. Interviews could be conducted to ask students directly if they can provide reasoning to support one of these responses and then to ask if they can provide reasoning to counter that response. This methodology would probably need to have distractor questions in the middle and to alternate the order of the two primary questions. Such a study could potentially show that on short-term time scales students have ideas to both support and counter a single outcome. This research should provide insights into how students form responses to physics questions and possibly offer examples of students assembling related ideas differently to justify different outcomes.
Investigate Variations from a Accessibility/Availability Perspective

An exciting direction to pursue would be to further investigate how the accessibility and availability of students’ ideas are influenced by these question variations. Section 7.5.1.3 we find evidence that suggests the given correct variation may influence the accessibility and availability of students’ ideas.

We suspect that the constraints to response options that the variations implement causes changes to the set of ideas or explanatory factors that are accessible to students. Further research could determine if this is the case by expanding the methods used by Heckler and Bogdan [45] to measure the availability and accessibility of students’ ideas while providing a correct outcome.

For example, the Change in Kinetic Energy question, show in Figure 7.2 and discussed in Section 7.5.1.3 could be made to have a few different versions. A version that measures the availability of compensation reasoning consisting of a description of the physical situation followed by “Is the following statement certain ‘The kinetic energies of the two carts are equal because their mass and velocity differences compensate?’”’ This could be compared to: a similar version that asks the same question but also provides the correct outcome, an original version, and a given correct variation.

When comparing across these questions it would be possible to determine if providing the correct outcome changes the availability of compensation reasoning, and if changes in availability may affect accessibility.

It may also be useful to measure the availability of correct reasoning for common conceptual physics question that students have been seen to have difficulty with using this method. Is the correct reasoning always available? Is it less available on questions where students disagreed with the provided outcome?
Investigate Associative Reasoning In several instances, students’ reasoning appeared to consist of them simply identifying a quantity that was the same as the response they were attempting to justify and providing that as their reasoning. For example, *no displacement* reasoning was used to justify *zero*, *same force* reasoning was used to justify *equal to*, and *velocity is zero* reasoning was used to justify *zero* from Sections 6.5.1, and 7.5.1, and 9.5.4.

This associative reasoning is very simple and often does not explain how the identified variable and response are connected through physics principals. This is therefore an incomplete form of reasoning that occurs independent of content area. Further research should be done to determine if this way of developing reasoning is common outside of question variations and the extent to which students may be using these types of reasoning to respond to research tasks. This research could potentially provide insights into how students respond to physics questions and how to interpret the results of research tasks focused on understanding student reasoning.

Investigate Effects of Small Changes to Question Statements A few cases offered evidence that small changes to the phrasing of particular research tasks might have a large impact on the way students respond to them.

One example is from the Gymnast question, where students responding to the *given correct* variation sometimes assumed that the situation was asymmetrical, as discussed in Section 8.6. Further research could investigate how student reasoning changes when the problem statement is changed to explicitly state that the situation is symmetrical. If students no longer assume asymmetry, what do they do instead? Do they still reinterpret or respond to the anomalous data in a different way? Does changing the question statement in this way affect the reasoning used on the *original* version?
Another example is from the Change in Kinetic Energy question, as discussed in Section 7.5.1.1, where a higher prevalence of reasoning stated that the forces on the two carts were the same, but a matching higher prevalence of reasoning stating that the distances the carts were pushed for were the same did not emerge. This could be due to the problem statement explicitly stating that the forces exerted on the carts are the same, but not explicitly stating that the displacements the forces are exerted over are the same. Further investigation should be conducted to determine if increasing the salience of the carts traveling the same distance would lead to more correct reasoning on the given correct variation as students will more readily mention both the force and displacements being the same. Furthermore, this may lead to more correct reasoning on the original variation as well.

Such studies would examine the effects of small changes to research tasks and add to the existing research on cuing [14, 15, 16, 42], as discussed in Section 2.5.

**Study to Predict Previously Undiscovered Reasoning**  
One way to test the accuracy of theoretical frameworks is to use them to make predictions and then test those predictions. This dissertation has identified one way to make testable predictions based on theoretical frameworks.

Section 9.7 demonstrated that students appear to be using the same elements of information in different configurations to justify different responses. It is possible that other common types of reasoning could be changed in similar ways if students adapt them to different response options targeted by question variations.

Identifying common types of reasoning and how they could be adapted to different response options using theoretical frameworks would lead to predictions of previously undocumented types of reasoning students would use, which could be tested by administering variations that target those other response options. This
would enable researchers to further refine theoretical frameworks, and possibly
develop further experiments to test frameworks based on their predictions.

10.5 Summary

The goal of the research in this dissertation was to develop question variations
based on the same physical situation that could elicit alternate ideas across multiple
content areas, implement them to demonstrate they could elicit alternate ideas,
understand and report the ideas found, and look for patterns in the ideas they
elicited. Several instances revealed informative differences in reasoning between the
original version of a question and one of the variations. In administering these
variations, five unique types of reasoning were identified, differences in the usage of
many types of reasoning between question versions were observed, and other results
specific to particular questions or variations were defined. These results are
summarized in Chapter 9.

All of the variations tested elicited alternate ideas, although with varying
degrees of success and most often when focused on the responses zero or equal to.
The consider variation, which was administered most, seems to be the least effective
at eliciting information that is different from the original version, most likely due to
the similarity of these question versions. The eliminate and given correct variations
show that they are capable of finding informative results more frequently, with
results being found in 20% or more of variations.

Although only seeing results on 20% to 40% of the variations may seem to imply
that the variations are not capable of eliciting different information more than half
of the time, this claim is not made here. This study was an exploratory study and
was designed to ask many question variations at once with lower higher statistical
power. Therefore, finding results in a few different content areas achieved the goal
of showing that these variations could elicit more information about students’ ideas,
even without the statistical power necessary to catch most of the differences these variations could cause.

Some variations included types of results that span content areas. Students responding to the *given correct* variation often consider it anomalous data and sometimes use the provided outcome as the basis for their reasoning. In other instances, larger portions of a classroom of students responding to the *eliminate* variation selected the same response as being correct and incorrect.

Results also spanned question variations. Students were able to develop unique reasoning to justify why an outcome was not what occurs and treated *zero* and *equal to* responses differently; in addition, *compensation* reasoning can take many forms, depending on the response students are attempting to justify.

Based on these findings, many areas of research were proposed in which these variations would be useful for future studies designed to understand students’ ideas and develop theoretical frameworks.

In conclusion, using this set of variations often allows for a more detailed view of students’ ideas, demonstrating the value of using these question versions for investigating students’ ideas. The additional information gleaned by asking question variations has the potential to impact instruction in physics courses by better informing instructors and researchers about the ideas students have. Moreover, using these question versions in pairs or individually can be a productive tool for instruction and research.
REFERENCES


[59] B. A. Lindsey, P. R. L. Heron, and P. S. Shaffer, American Journal of Physics 77, 999 (2009), ISSN 00029505.


[84] H. M. Wellman and K. F. Miller, British Journal of Developmental Psychology 4, 31 (1986), ISSN 0261510X.
[86] B. A. Lindsey, P. R. L. Heron, and P. S. Shaffer, American Journal of Physics 80, 154 (2012).
APPENDIX A
ELECTRIC CIRCUITS HANDOUTS

The four versions of the Electric Circuits question that were administered during the pilot study are shown below. The full page handout is shown. The course number and date where changed to match the class and date for each administration.
1. The circuit at right contains an ideal battery, three identical light bulbs, and a switch. Initially the switch is open.

After the switch closes:

Does the brightness of bulb A increase, decrease, or remain the same? Explain.

Figure A.1. The original version of the Electric Circuits question [3].
The circuit at right contains an ideal battery, three identical light bulbs, and a switch. Initially the switch is open.

After the switch closes, does the brightness of bulb A increase? Explain.

Figure A.2. The consider increase variation of the Electric Circuits question.
Imagine you are taking an exam with the question shown in the box below.

You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate?

1. The circuit at right contains an ideal battery, three identical light bulbs, and a switch. Initially the switch is open.

After the switch closes:

Does the brightness of bulb A increase, decrease, or remain the same? Explain.

Figure A.3. The eliminate variation of the Electric Circuits question.
1. The circuit at right contains an ideal battery, three identical light bulbs, and a switch. Initially the switch is open.

After the switch closes, the brightness of bulb A increases. Explain.

Figure A.4. The given correct variation of the Electric Circuits question.
APPENDIX B
ACROSS-CONTENT STUDY PRETEST QUESTIONS

B.1 Pretest 2 - Oval Track Question

Original, Consider: Greater Than, Less Than, and Equal To

B.1.1 Original Version

A car moves clockwise at constant speed around an oval track as shown in the top-view diagram below. (Note: This is not a strobe photograph.)

Is the magnitude of the acceleration of the small car at Point E greater than, less than, or equal to the acceleration of the car at Point G? Explain.

Figure B.1. The original version of the Oval Track question from Pretest 2 [3].
B.1.2 Consider Greater Than Variation

A car moves clockwise at constant speed around an oval track as shown in the top-view diagram below. (Note: This is not a strobe photograph.)

Is the magnitude of the acceleration of the small car at Point E greater than the acceleration of the car at Point G? Explain.

Figure B.2. The consider greater than variation of the Oval Track question from Pretest 2.

B.1.3 Consider Less Than Variation

A car moves clockwise at constant speed around an oval track as shown in the top-view diagram below. (Note: This is not a strobe photograph.)

Is the magnitude of the acceleration of the small car at Point E less than the acceleration of the car at Point G? Explain.

Figure B.3. The consider less than variation of the Oval Track question from Pretest 2.
A car moves clockwise at constant speed around an oval track as shown in the top-view diagram below. (Note: This is not a strobe photograph.)

Is the magnitude of the acceleration of the small car at Point E equal to the acceleration of the car at Point G? Explain.

Figure B.4. The consider equal to variation of the Oval Track question from Pretest 2.
B.2 Pretest 3 - Pendulum Question

Original, Eliminate, Consider Upward, Given Correct

B.2.1 Original Version

A pendulum is released from rest at the position shown below. When it reaches
the bottom of its swing is the direction of its acceleration upward, upward and to
the right, right, downward, or zero? Explain.

Figure B.5. The original version of the Pendulum question from Pretest 3 [89].
Imagine you are taking an exam with the question shown in the box below. You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate? Explain your reasoning.

Figure B.6. The eliminate variation of the Pendulum question from Pretest 3.
B.2.3 Consider Upward Variation

A pendulum is released from rest at the position shown below. When it reaches the bottom of its swing is the direction of its acceleration upward? Explain.

Figure B.7. The consider upward variation of the Pendulum question from Pretest 3.

B.2.4 Given Correct Variation

A pendulum is released from rest at the position shown below. When it reaches the bottom of its swing the direction of its acceleration is upward. Explain.

Figure B.8. The given correct variation of the Pendulum question from Pretest 3.
B.3 Pretest 4 - Car-Over-Hill Question

Original, Eliminate, Consider Greater Than, Given Correct

B.3.1 Original Version

A car drives over a hill as shown above.

Is the magnitude of the gravitational force exerted by the earth on the car greater than, less than, or equal to the magnitude of the force exerted by the hill on the car? Explain.

Figure B.9. The original version of the Car-Over-Hill question from Pretest 4 [54].
B.3.2 Eliminate Variation

Imagine you are taking an exam with the question shown in the box below. You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate? Explain your reasoning.

A car drives over a hill as shown above.

Is the magnitude of the gravitational force exerted by the earth on the car greater than, less than, or equal to the magnitude of the force exerted by the hill on the car? Explain.

Figure B.10. The *eliminate* variation of the Car-Over-Hill question from Pretest 4.

B.3.3 Consider Greater Than Variation

A car drives over a hill as shown above.

Is the magnitude of the gravitational force exerted by the earth on the car greater than the magnitude of the force exerted by the hill on the car? Explain.

Figure B.11. The *consider greater than* variation of the Car-Over-Hill question from Pretest 4.
B.3.4 Given Correct Variation

A car drives over a hill as shown above.

The magnitude of the gravitational force exerted by the earth on the car is greater than the magnitude of the force exerted by the hill on the car. Explain.

Figure B.12. The *given correct* variation of the Car-Over-Hill question from Pretest 4.
B.4 Pretest 5 - Newton’s 3rd Law Question

Original, Eliminate, Consider Greater Than

B.4.1 Original Version

Part I: Block A is on a platform as shown. Consider the three cases described below.

Case 3: The platform is now moving upward and *speeding up*.
In Case 3, is the magnitude of the force exerted by the platform on Block A *greater than*, *less than*, or *equal to* the magnitude of the force exerted by Block A on the platform? Explain.

Figure B.13. The *original* version of the Newton’s 3rd Law question from Pretest 5. Only the third case students were asked to consider is shown, as that was the only case that the question variations asked about [3].
B.4.2 Eliminate Variation

Part I: Block A is on a platform as shown. Consider the three cases described below.

Case 3: The platform is now moving upward and *speeding up*.

Imagine you are taking an exam with the question shown in the box below. You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate? Explain your reasoning.

Figure B.14. The *eliminate* variation of the Newton’s 3rd Law question from Pretest 5.
B.4.3 Consider Greater Than Variation

*Part I:* Block A is on a platform as shown. Consider the three cases described below.

Case 3: The platform is now moving upward and *speeding up.* In Case 3, is the magnitude of the force exerted by the platform on Block A greater than the magnitude of the force exerted by Block A on the platform? Explain.

Figure B.15. The *consider greater than* variation of the Newton’s 3rd Law question from Pretest 5.
B.5 Pretest 6 - Gymnast Question

Original, Eliminate, Given Correct

B.5.1 Original Version

A gymnast who weighs 500N is suspended by two ropes as shown. Is the magnitude of the tension in the left rope greater than, less than, or equal to 250N? Explain your reasoning.

Figure B.16. The original version of the Gymnast question from Pretest 6 [54].
B.5.2 Eliminate Variation

Imagine you are taking an exam with the question shown in the box below. You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate? Explain your reasoning.

A gymnast who weighs 500N is suspended by two ropes as shown. Is the magnitude of the tension in the left rope greater than, less than, or equal to 250N? Explain your reasoning.

Figure B.17. The eliminate variation of the Gymnast question from Pretest 6.
B.5.3 Given Correct Variation

A gymnast who weighs 500N is suspended by two ropes as shown. The magnitude of the tension in the left rope is greater than 250N. Explain your reasoning.

Figure B.18. The given correct variation of the Gymnast question from Pretest 6.
B.6 Pretest 7a - 1-D Conservation of Momentum A

Original, Eliminate, Consider Greater Than

B.6.1 Original Version

Part II
Two experiments are conducted with gliders on a level, frictionless track:

Experiment 1: Glider A is launched toward a stationary target, Glider X. After the collision, Glider A is at rest.

Experiment 2: Glider B is launched toward a stationary target, Glider Y. Glider B has the same mass and initial velocity as Glider A in Experiment 1. After the collision, Glider B has reversed direction.

The final speed of Glider X is greater than the final speed of Glider Y (i.e., $v_{Xf} > v_{Yf}$).

The mass of Glider X is less than the mass of Glider Y (i.e., $m_X < m_Y$).

In Experiment 2, is the magnitude of the change in momentum of Glider B greater than, less than, or equal to the magnitude of the change in momentum of Glider Y? Explain your reasoning.

Figure B.19. The original version of the 1-D Conservation of Momentum A question from Pretest 7a [3].
B.6.2 Eliminate Variation

Part II

Two experiments are conducted with gliders on a level, frictionless track:

Experiment 1: Glider A is launched toward a stationary target, Glider X. After the collision, Glider A is at rest.

Experiment 2: Glider B is launched toward a stationary target, Glider Y. Glider B has the same mass and initial velocity as Glider A in Experiment 1. After the collision, Glider B has reversed direction.

The final speed of Glider X is greater than the final speed of Glider Y (i.e., \( v_{Xf} > v_{Yf} \)).

The mass of Glider X is less than the mass of Glider Y (i.e., \( m_X < m_Y \)).

Imagine you are taking an exam with the question shown in the box below. You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate? Explain your reasoning.

Figure B.20. The eliminate variation of the 1-D Conservation of Momentum A question from Pretest 7a.
B.6.3 Consider Greater Than Variation

Part II

Two experiments are conducted with gliders on a level, frictionless track:

Experiment 1: Glider A is launched toward a stationary target, Glider X. After the collision, Glider A is at rest.

Experiment 2: Glider B is launched toward a stationary target, Glider Y. Glider B has the same mass and initial velocity as Glider A in Experiment 1. After the collision, Glider B has reversed direction.

The final speed of Glider X is greater than the final speed of Glider Y (i.e., $v_{Xf} > v_{Yf}$).

The mass of Glider X is less than the mass of Glider Y (i.e., $m_X < m_Y$).

In Experiment 2, is the magnitude of the change in momentum of Glider B greater than the magnitude of the change in momentum of Glider Y? Explain your reasoning.

Figure B.21. The consider greater than variation of the 1-D Conservation of Momentum A question from Pretest 7a.
B.7 Pretest 7b - 1-D Conservation of Momentum B

Original, Consider Less Than

B.7.1 Original Version

**Part II**
Two experiments are conducted with gliders on a level, frictionless track:

Experiment 1: Glider A is launched toward a stationary target, Glider X. After the collision, Glider A is at rest.

Experiment 2: Glider B is launched toward a stationary target, Glider Y. Glider B has the same mass and initial velocity as Glider A in Experiment 1. After the collision, Glider B has reversed direction.

The final speed of Glider X is *greater than* the final speed of Glider Y (i.e., \( v_{Xf} > v_{Yf} \)).

The mass of Glider X is *less than* the mass of Glider Y (i.e., \( m_X < m_Y \)).

Is the magnitude of the final momentum of **Glider X** greater than, less than, or equal to the magnitude of the final momentum of **Glider Y**?

Figure B.22. The *original* version of the 1-D Conservation of Momentum B question from Pretest 7b [3].

280
B.7.2 Consider Less Than Variation

**Part II**
Two experiments are conducted with gliders on a level, frictionless track:

Experiment 1: Glider A is launched toward a stationary target, Glider X. After the collision, Glider A is at rest.

Experiment 2: Glider B is launched toward a stationary target, Glider Y. Glider B has the same mass and initial velocity as Glider A in Experiment 1. After the collision, Glider B has reversed direction.

The final speed of Glider X is greater than the final speed of Glider Y (i.e., $v_{XF} > v_{YF}$).

The mass of Glider X is less than the mass of Glider Y (i.e., $m_X < m_Y$).

Is the magnitude of the final momentum of Glider X less than the magnitude of the final momentum of Glider Y?

Figure B.23. The consider less than variation of the 1-D Conservation of Momentum B question from Pretest 7b.
B.8 Pretest 8 - 2-D Collision

Original, Given Correct

B.8.1 Original Version

Two objects are arranged on a level, frictionless table as shown. An experiment is conducted in which Object A is launched toward the stationary Block B. The initial and final velocities of Object A are shown.

The mass of Block B is six times that of Object A. \( m_B = 6m_A \)

Is the magnitude of the change in velocity of Object A greater than, less than, or equal to that of Block B? Explain.

Figure B.24. The original version of the 2-D Collision question from Pretest 8 [3].

B.8.2 Given Correct Variation

Two objects are arranged on a level, frictionless table as shown. An experiment is conducted in which Object A is launched toward the stationary Block B. The initial and final velocities of Object A are shown.

The magnitude of the change in velocity of Object A is greater than that of Block B. Explain.

Figure B.25. The given correct variation of the 2-D Collision question from Pretest 8.
B.9 Pretest 9a - Change in Kinetic Energy Question

Original, Consider Equal To, Given Correct

B.9.1 Original Version

Two carts, A and B, are initially at rest on a frictionless, horizontal table. The mass of Cart A is less than that of Cart B. The same constant force, $F_0$, is exerted on each cart, in turn, as it travels between the two marks on the table. After they pass the second mark, the carts are allowed to glide freely.

After the carts have passed the second mark, is the kinetic energy of Cart A greater than, less than, or equal to the kinetic energy of Cart B? Explain your reasoning.

Figure B.26. The original version of the Change in Kinetic Energy question from Pretest 9 [3].
B.9.2 Consider Equal To Variation

Two carts, A and B, are initially at rest on a frictionless, horizontal table. The mass of Cart A is less than that of Cart B. The same constant force, \( F_0 \), is exerted on each cart, in turn, as it travels between the two marks on the table. After they pass the second mark, the carts are allowed to glide freely.

After the carts have passed the second mark, is the kinetic energy of Cart A equal to the kinetic energy of Cart B? Explain your reasoning.

Figure B.27. The consider equal to variation of the Change in Kinetic Energy question from Pretest 9.
B.9.3 Given Correct Variation

Two carts, A and B, are initially at rest on a frictionless, horizontal table. The mass of Cart A is less than that of Cart B. The same constant force, $F_0$, is exerted on each cart, in turn, as it travels between the two marks on the table. After they pass the second mark, the carts are allowed to glide freely.

After the carts have passed the second mark, the kinetic energy of Cart A is equal to the kinetic energy of Cart B. Explain.

Figure B.28. The *given correct* variation of the Change in Kinetic Energy question from Pretest 9.
B.10 Pretest 9b - Work-by-Wall Question

Original, Eliminate, Given Correct

B.10.1 Original Version

A block of mass $m$ is attached to an ideal (massless) spring with coefficient $k$, as shown below. The friction between the block and the surface is negligible. Initially, the block is pulled to Point $P$, stretching the spring. At time $t = t_0$, the block is released from rest. When the block is released, it moves back toward Point $Q$, where the spring is at its equilibrium length. Consider System A, consisting of the block and the spring.

Does the wall do positive, negative, or zero work on System A? Explain your reasoning.

Figure B.29. The original version of the Work-by-Wall question from Pretest 9b [60].
B.10.2 Eliminate Variation

A block of mass $m$ is attached to an ideal (massless) spring with coefficient $k$, as shown below. The friction between the block and the surface is negligible. Initially, the block is pulled to Point $P$, stretching the spring. At time $t = t_0$, the block is released from rest. When the block is released, it moves back toward Point $Q$, where the spring is at its equilibrium length. Consider System A, consisting of the block and the spring.

Imagine you are taking an exam with the question shown in the box below. You want to first eliminate one response you are pretty sure is incorrect. Which response would you eliminate? Why is that response the best one to eliminate? Explain your reasoning.

Figure B.30. The eliminate variation of the Work-by-Wall question from Pretest 9b.
B.10.3 Given Correct Variation

A block of mass $m$ is attached to an ideal (massless) spring with coefficient $k$, as shown below. The friction between the block and the surface is negligible. Initially, the block is pulled to Point $P$, stretching the spring. At time $t = t_0$, the block is released from rest. When the block is released, it moves back toward Point $Q$, where the spring is at its equilibrium length. Consider System A, consisting of the block and the spring.

![Diagram of a block attached to a spring](image)

The wall does zero work on System A. Explain.

Figure B.31. The *given correct* variation of the Work-by-Wall question from Pretest 9b.
Two carts, A and B, are initially at rest on a frictionless, horizontal table. The mass of Cart A is less than that of Cart B. The same constant force, F, is exerted on each cart, in turn, as it travels between the two marks on the table. The carts are then allowed to glide freely.

After the carts have passed the second mark, is the magnitude of the momentum of Cart A greater than, less than, or equal to the magnitude of the momentum of Cart B? Explain your reasoning.

Figure B.32. The original version of the Change in Momentum question from Pretest 10 [3].
B.11.2 Consider Equal To Variation

Two carts, A and B, are initially at rest on a frictionless, horizontal table. The mass of Cart A is less than that of Cart B. The same constant force, F, is exerted on each cart, in turn, as it travels between the two marks on the table. The carts are then allowed to glide freely.

After the carts have passed the second mark, is the magnitude of the momentum of Cart A equal to the magnitude of the momentum of Cart B? Explain your reasoning.

Figure B.33. The consider equal to variation of the Change in Momentum question from Pretest 10.
B.11.3 Given Correct Variation

Two carts, A and B, are initially at rest on a frictionless, horizontal table. The mass of Cart A is less than that of Cart B. The same constant force, $F$, is exerted on each cart, in turn, as it travels between the two marks on the table. The carts are then allowed to glide freely.

After the carts have passed the second mark the magnitude of the momentum of Cart A is less than the magnitude of the momentum of Cart B. Explain.

Figure B.34. The *given correct* variation of the Change in Momentum question from Pretest 10.
B.12 Pretest 11 - Pivoting Rod Question

Original, Consider Equal To, Given Correct

B.12.1 Original Version

A rod is attached at its center to a fixed, frictionless pivot. Points A and B are marked for reference. The rod spins counterclockwise (in the plane of the paper) at a constant rate.

Is the speed of Point A greater than, less than, or equal to the speed of Point B? Explain your reasoning.

Figure B.35. The original version of the Pivoting Rod question from Pretest 11 [3].
B.12.2 Consider Equal To Variation

A rod is attached at its center to a fixed, frictionless pivot. Points A and B are marked for reference. The rod spins counterclockwise (in the plane of the paper) at a constant rate.

Is the speed of Point A equal to the speed of Point B? Explain your reasoning.

Figure B.36. The consider equal to variation of the Pivoting Rod question from Pretest 11.

B.12.3 Given Correct Variation

A rod is attached at its center to a fixed, frictionless pivot. Points A and B are marked for reference. The rod spins counterclockwise (in the plane of the paper) at a constant rate.

The speed of Point A is greater than the speed of Point B. Explain.

Figure B.37. The given correct variation of the Pivoting Rod question from Pretest 11.
An aluminum disk and an iron wheel (with spokes of negligible mass) have the same radius $R$ and mass $M$ as shown below. Each is free to rotate about its own fixed horizontal frictionless axle. Both objects are initially at rest. Identical small lumps of clay are attached to their rims as shown in the figure.

Is the angular acceleration of the disk+clay system greater than, less than, or equal to the angular acceleration of the wheel+clay system? Explain your reasoning.

Figure B.38. The original version of the Rotating Disk/Wheel question from Pretest 12 [63].
B.13.2 Consider Greater Than Variation

An aluminum disk and an iron wheel (with spokes of negligible mass) have the same radius $R$ and mass $M$ as shown below. Each is free to rotate about its own fixed horizontal frictionless axle. Both objects are initially at rest. Identical small lumps of clay are attached to their rims as shown in the figure.

Is the angular acceleration of the disk+clay system greater than the angular acceleration of the wheel+clay system? Explain your reasoning.

Figure B.39. The consider greater than variation of the Rotating Disk/Wheel question from Pretest 12.
B.13.3  Given Correct Variation

An aluminum disk and an iron wheel (with spokes of negligible mass) have the same radius $R$ and mass $M$ as shown below. Each is free to rotate about its own fixed horizontal frictionless axle. Both objects are initially at rest. Identical small lumps of clay are attached to their rims as shown in the figure.

The angular acceleration of the disk+clay system is greater than the angular acceleration of the wheel+clay system. Explain.

Figure B.40. The given correct variation of the Rotating Disk/Wheel question from Pretest 12.
B.14  Pretest 13 - Simple Harmonic Oscillator (SHO) Question

Original, Consider Zero, Given Correct

B.14.1  Original Version

A spring is used to connect a block to a wall (neglect the mass of the spring, and assume the surface is frictionless). In the situation shown at below, the block remains at rest.

![Diagram of the original version of the SHO question](image)

A student moves the block 0.5 m to the right of its original location and at a certain instant (instant 1) releases it from rest. The subsequent motion of the block is shown in the image below. (The diagrams show the position of the block at equal time intervals.)

![Diagram of the block motion](image)

For each of the following instants, state whether the acceleration of the block is to the left, to the right, or zero at that instant. Explain your reasoning in each case.

**Instant 5:**
Explain your reasoning.

---

Figure B.41. The original version of the Simple Harmonic Oscillator question from Pretest 13 [3].
B.14.2  Consider Zero Variation

A spring is used to connect a block to a wall (neglect the mass of the spring, and assume the surface is frictionless). In the situation shown at below, the block remains at rest.

A student moves the block 0.5 m to the right of its original location and at a certain instant (instant 1) releases it from rest. The subsequent motion of the block is shown in the image below. (The diagrams show the position of the block at equal time intervals.)

For each of the following instants, state whether the acceleration of the block is to the left, to the right, or zero at that instant. Explain your reasoning in each case.

At instant 5 is the acceleration of the block zero? Explain your reasoning.

Figure B.42. The consider zero variation of the Simple Harmonic Oscillator question from Pretest 13.
A spring is used to connect a block to a wall (neglect the mass of the spring, and assume the surface is frictionless). In the situation shown at below, the block remains at rest.

A student moves the block 0.5 m to the right of its original location and at a certain instant (instant 1) releases it from rest. The subsequent motion of the block is shown in the image below. (The diagrams show the position of the block at equal time intervals.)

For each of the following instants, state whether the acceleration of the block is to the left, to the right, or zero at that instant. Explain your reasoning in each case.

The acceleration at instant 5 is to the right. Explain.

Figure B.43. The given correct variation of the Simple Harmonic Oscillator question from Pretest 13.
This appendix describes how students were broken into random groups using the Catalyst system.

Because the Catalyst system does not have a way to assign students random questions, or random versions of questions, a “question” was created to end up direct students randomly down different “tracks”.

The Catalyst system could do two things to help randomly direct students down different tracks: It could send students to different questions based on their response to a previous question, and it could randomize the order of responses to multiple-choice questions. These two tools were used in conjunction to direct students down tracks in a random manner.

The randomization method consisted of a single question that students were always asked first. This randomization “question” told them which pretest this was and how long this pretest should take them. There was then a drop-down box with three identical responses that all said “Click Next to Proceed.” These three identical responses would each direct students to a different track and were set to be in random order. This ensured that the response selected by default was always randomly selected. If a student simply clicked next they would be taken to the randomly selected track. If a student decided to change the response to a different “Click Next to Proceed,” they would be selecting a random track to be directed to. Either way, the students were sent to a random track. An example of a randomization “question” is shown in Figure C.1, and a general depiction of this method is given in Figure 5.1.
Phy 121 Pretest #6

This is the sixth pretest for PHY 121 at The University of Maine. This pretest has 5 questions and should take you about 13 minutes to finish.

Next >>

Phy 121 Pretest #6

This is the sixth pretest for PHY 121 at The University of Maine. This pretest has 5 questions and should take you about 13 minutes to finish.

Click Next to Proceed

Next >>

Figure C.1. The randomization question used to assign the students to random groups. Shown with the menu of options collapsed (above) and expanded (below).
APPENDIX D

GROUP SIMILARITY IN THE ACROSS-CONTENT STUDY

Although the groups were randomly generated every week in the across-content study, the goal was to have a secondary way to ensure that different responses or reasoning across versions were not due to group inequality. No questions were asked without variations in the pilot study; therefore, there was no method for checking for group similarity there.

D.1 Testing the Similarity of a Set of Groups

To test how similar the randomly generated groups were, a statistical test was run on students’ responses to questions where they all answered the same version. The questions were referred to as “base questions.” Because base questions could have three or more options and there were always three or more groups, a $\chi^2$ test was used to check for differences in students’ responses to base questions. In order to avoid having low cell counts (less than 5) in any cell, responses that were not commonly chosen were combines into an “other” category. In cases where this would have resulted in all the responses being compressed into a single group, two groups were kept and the $\chi^2$ test result was accepted as perhaps being somewhat inaccurate. If the correct response was chosen by at least 5 people in each track, it was not combined with a different group [90]. An example of an analysis of a single base question is shown in Table D.1.

D.2 Similarity of Pretest Groups Overall

Because differences in the groups were expected in some weeks, which may be statistical anomalies and not meaningful, the similarity of groups in individual
Table D.1. Number of students selecting each response or some other response for each group of students on a base question. Using a $\chi^2$ test of the whole 3x4 table resulted in $p = 0.5317$ for this base question, which indicates the groups are similar.

<table>
<thead>
<tr>
<th>Response</th>
<th>Track 1</th>
<th>Track 2</th>
<th>Track 3</th>
<th>Track 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>19</td>
<td>11</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>22</td>
<td>26</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>Other</td>
<td>14</td>
<td>16</td>
<td>16</td>
<td>12</td>
</tr>
</tbody>
</table>

weeks was not examined. Instead, the group similarities were averaged over all pretests to ensure that similar groups were achieved in general.

To check for group similarity over all pretests, tests were run to determine if the groups’ responses to many base questions were similar enough to have an average $p > .2$. Although an average $p$ value greater than .2 on its own is not a justification for group similarity, when combined with attempts to ensure that the groups were selected randomly, it indicates that a statistically significant difference in the groups was far enough away that there was no need to worry about the groups being different. Because of the effort involved in coding reasoning, only students’ responses to questions were considered when looking for differences in groups.

The number of base questions varied from week to week, so a few rules were followed when determining which base questions to use in the analysis of group similarity. No more than 2 base questions per pretest were used to avoid giving too much weight to any one set of random groups. When there were more than two base questions from which to choose, the two most closely related in context to the varied question that followed were chosen. If there were no appropriate questions, the two base questions directly before the varied question were used. There was one exception to these rules where a base question was not used because it had many low cell count errors.
In total, 18 base questions were examined. There were was one week with no base question analyzed, six weeks with one base question analyzed, and six weeks with two base questions analyzed. The average $p$ value for the 18 base questions was .542, which is much greater than the $p > .2$ requirement. Therefore, all groups were assumed to be similar initially in each of the 13 weeks.
APPENDIX E

DEMOGRAPHICS OF PHYSICS COURSE

The students used as research subjects in all studies in this dissertation were from introductory physics courses at the University of Maine. The across-content study used students from only the calculus-based PHY121 sequence described below.

The University of Maine (UMaine) is a land-grant institution in central Maine. It is a public university offering bachelor's, master’s, post-master's certificates and doctoral degrees of research/scholarship. The university has approximately 8,500 undergraduates and 2,000 graduate students. The college population is 51.5% female, 80% in state, and 78% white (with 10% unspecified), with an average age of 21.5. Most students commute to campus, with 38% of students living on campus. The university accepts 81% of applicants; accepted students have a mean SAT Verbal score of 537 and a mean SAT Quantitative score of 546 [91, 92].

The school has a large Engineering and Engineering Tech population, accounting for 16.5% of degrees conferred. Life sciences and health-related programs make up 12.3% of the population based on degrees conferred. Physical sciences make up only 1.3% of the student population based on degrees conferred [92].

The Physics 121 (PHY121) sequence is a calculus-based introductory sequence. It is intended for engineering and physical science majors and was the course in which the across-content study took place.

The course has a one-hour lecture twice a week, a one-hour recitation twice a week, and one two-hour lab per week. The lecture sizes range from 100 to 240 students, and the recitations and labs usually contain 24 students. Usual enrollment is around 240 students.
This courses use *Physics for Engineers & Scientists: A Strategic Approach* by Knight and follow a traditional sequence of introductory material [93].

Recitation activities involve having the TA go over homework assignments on the board, engaging in group problem solving, or completing tutorials. In one recitation session per week, students worked on a tutorial, usually from *Tutorials in Introductory Physics* [3].

The laboratories are traditional labs written at UMaine.
BIOGRAPHY OF THE AUTHOR

Jeffrey Mathis Hawkins grew up in Irvine, California. He graduated from the Kline School in 1998, and from St. Margret’s Episcopal School in 2002. He completed his two Bachelor of Science degrees in Physics and astronomy at the University of Washington in Seattle in 2007.

He has served as a Graduate Research Assistant and a Graduate Teaching Assistant at the University of Maine. He is a member of the American Association of Physics Teachers (AAPT) and the American Physical Society (APS). His latest publication is “Hawkins, J. M., Frank, B. W., Thompson, J. R., Wittmann, M. C., & Wemyss, T. M. (2012). Probing Student Understanding With Alternative Questioning Strategies. AIP Conference Proceedings, 1413(1), 207-210.”

Jeffrey Mathis Hawkins is a candidate for the Doctor of Philosophy degree in Physics from the University of Maine in May 2019.