ABSTRACT

Title of Dissertation: BEYOND PROBLEM SOLVING:
EVALUATING INTRODUCTORY PHYSICS COURSES THROUGH THE HIDDEN CURRICULUM
Jeffery M. Saul, Doctor of Philosophy, 1998

Dissertation directed by: Professor Edward F. Redish
Department of Physics

A large number of innovative approaches have been developed based on Physics
Education Research (PER) to address student difficulties introductory physics
instruction. Yet, there are currently few widely accepted assessment methods for
determining the effectiveness of these methods. This dissertation compares the
effectiveness of traditional calculus-based instruction with University of Washington’s
Tutorials, University of Minnesota’s Group Problem Solving & Problem Solving Labs,
and Dickinson College’s Workshop Physics. Implementation of these curricula were
studied at ten undergraduate institutions. The research methods used include the Force
Concept Inventory (FCI), the Maryland Physics Expectation (MPEX) survey, specially
designed exam problems, and interviews with student volunteers. The MPEX survey is a
new diagnostic instrument developed specifically for this study.

Instructors often have learning goals for their students that go beyond having
them demonstrate mastery of physics through typical end-of-chapter problems on exams
and homeworks. Because these goals are often not stated explicitly nor adequately
reinforced through grading and testing, we refer to this kind of learning goal as part of
the course’s “hidden curriculum.” In this study, we evaluate two aspects of student
learning from this hidden curriculum in the introductory physics sequence: conceptual
understanding and expectations (cognitive beliefs that affect how students think about
and learn physics).

We find two main results. First, the exam problems and the pre/post FCI results
on students conceptual understanding showed that the three research-based curricula
were more effective than traditional instruction for helping students learn velocity
graphs, Newtonian concepts of force and motion, harmonic oscillator motion, and
interference. Second, although the distribution of students’ expectations vary for
different student populations, the overall distributions differ considerably from what
expert physics instructors would like them to have and differ even more by the end of the
first year. Only students from two of the research-based sequences showed any
improvement in their expectations.
BEYOND PROBLEM SOLVING:
EVALUATING INTRODUCTORY PHYSICS COURSES
THROUGH THE HIDDEN CURRICULUM

by

Jeffery M. Saul

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1998

Advisory Committee:

Professor Edward F. Redish, Chair/Advisor
Professor Thomas D. Cohen
Associate Professor Richard F. Ellis
Professor James T. Fey
Professor Jordan A. Goodman
Professor John L. Layman
Dedication

This work is dedicated to my family, friends, and colleagues who always gave me support and encouragement, to my students who made me think about how to teach them more effectively, to Joy Watnik, my significant other, for her understanding, patience, assistance, and moral support and to my advisor, Edward Redish, without whose help this dissertation would never have been completed.
Acknowledgements

I would like to thank the members of the University of Maryland Physics Education Research Group, Richard Steinberg, Lei Bao, Dan Campbell, John Layman, John Lello, Chris Allen and Mel Sabella, who contributed substantially to the collection and analysis of the data presented in this dissertation. I would also like to thank Maria Crosson, Sam Berner, and Carlee Boettger for their help processing the data. Visitors to the group, including John Christopher, Alvin Sapirstein, and Pratibha Jolly, contributed valuable comments and insights. I am grateful to Pat Cooney, Ibrahim Halloun, Richard Hake, Curt Heiggelke, Alan van Heuvelen, Ed Adelson, Chris Cooksey and Tom Foster as well Lillian C. McDermott and the members of the University of Washington Physics Education Research Group for the discussions we shared on assessment issues and for their encouragement of this research. I am particularly grateful to Priscilla Laws for her many years of encouragement and support for this project and to me personally.

A project like this can not be completed without the help and cooperation of many people at many institutions. I would like to thank the very many faculty members at Carroll College, Dickinson College, Drury College, Ohio State University, University of Maryland, University of Minnesota, Moorhead State University, Nebraska Wesleyan University, Prince Georges Community College, and Skidmore College who cooperated with us and gave us class time for their students to fill out the concept tests and our survey. I am especially grateful to Bill Welch at Carroll College, Maurinda Wingard and Phil Thompson at Dickinson College, Bruce Callen at Drury College, Ken Heller, Pat Heller, Tom Foster and Laura McCullough at the University of Minnesota, Gerald Hart at Moorhead State University, Bob Fairchild and Bill Wehrbein at Nebraska Wesleyan
University, Chris Cooksey and Leith Dyer at the Ohio State University, Scott Sinex and Barbara Gage at Prince Georges Community College, and William Standish at Skidmore College for agreeing to participate in this study and coordinating the collection of data at their institutions.

I am also very grateful to Priscilla Laws and her colleagues at the Dickinson College Summer Workshops for agreeing to let us give our survey to workshop participants and to Larry Kirkpatrick and his colleagues for permitting us to survey the US Physics Olympics Team students.

I would also like to thank several people for their help while I was writing this dissertation. I would like to thank Richard Steinberg and Michael Wittmann for the many hours spent proofreading the manuscript. I would also like to thank Mel Sabella for his help with various aspects of the project including coming in on a Sunday to rescue my advisor and me when we were trapped in the basement. I am also grateful to my committee for the opportunity to write the first physics education dissertation at the University of Maryland.

Last but not least, I would like to thank my advisor, Edward Redish, who has been a good friend, mentor, and colleague. Thank you, Joe, for your time and effort, for helping me develop as an instructor and a researcher, and most of all for believing in me.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiv</td>
</tr>
<tr>
<td><strong>PART I. INTRODUCTION AND BACKGROUND: PHYSICS EDUCATION</strong></td>
<td></td>
</tr>
<tr>
<td>RESEARCH, STUDENTS LEARNING, AND ASSESSMENT</td>
<td></td>
</tr>
<tr>
<td>Chapter 1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Motivation</td>
<td></td>
</tr>
<tr>
<td>Course Goals and the Hidden Curriculum</td>
<td>2</td>
</tr>
<tr>
<td>Problem Statement/Research Questions</td>
<td>4</td>
</tr>
<tr>
<td>Experimental Design</td>
<td>5</td>
</tr>
<tr>
<td>Dissertation Overview</td>
<td>7</td>
</tr>
<tr>
<td>Dissertation Summary</td>
<td>8</td>
</tr>
<tr>
<td>Chapter 2. BACKGROUND: AN OVERVIEW OF RELEVANT EDUCATION RESEARCH</td>
<td>16</td>
</tr>
<tr>
<td>Warning Signs</td>
<td></td>
</tr>
<tr>
<td>Mazur’s Example: Student Difficulties with Conceptual Understanding</td>
<td>16</td>
</tr>
<tr>
<td>Hammer’s Example: Two approaches to Learning Physics</td>
<td>21</td>
</tr>
<tr>
<td>Two approaches to learning physics</td>
<td>22</td>
</tr>
<tr>
<td>Two approaches to problem solving</td>
<td>25</td>
</tr>
<tr>
<td>Implications</td>
<td>26</td>
</tr>
<tr>
<td>Tobias’ Example: Observations on the Traditional Lecture</td>
<td>28</td>
</tr>
<tr>
<td>Teaching Method</td>
<td></td>
</tr>
<tr>
<td>Lack of a narrative or story line</td>
<td>32</td>
</tr>
<tr>
<td>Overemphasis on quantitative problem solving</td>
<td>33</td>
</tr>
<tr>
<td>A classroom culture that discourages discussion and cooperation</td>
<td>36</td>
</tr>
<tr>
<td>Implications</td>
<td>39</td>
</tr>
<tr>
<td>Summary</td>
<td>40</td>
</tr>
<tr>
<td>Problem Solving: Experts vs. Novices</td>
<td>41</td>
</tr>
<tr>
<td>Characterizing Expert and Novice Problem Solvers</td>
<td>41</td>
</tr>
<tr>
<td>Knowledge structure</td>
<td>42</td>
</tr>
<tr>
<td>Problem solving approaches</td>
<td>42</td>
</tr>
<tr>
<td>Use of example problems and reading</td>
<td>43</td>
</tr>
<tr>
<td>How to Help Novices become Experts</td>
<td>44</td>
</tr>
<tr>
<td>Instruction to improve problem solving</td>
<td>44</td>
</tr>
<tr>
<td>WISE</td>
<td>45</td>
</tr>
<tr>
<td>Overview, Case Study</td>
<td>46</td>
</tr>
</tbody>
</table>
Group Problem Solving with context-rich problems. 46
Studies on improving students’ depth and
organization of knowledge………………..50
Appropriate problems…………………………………………………53
Cues, triggers, and models used by students………………61
Conceptual Understanding…………………………………………63
Kinematics and Newton’s Laws…………………………………………64
The language of physics………………………………………………..65
Common sense beliefs on force and motion………………..65
Representations………………………………………………………68
Mechanisms for Changing Students’ Beliefs………………………71
Expectations and Epistemology……………………………………….73
Previous Research on Cognitive Expectations
in the Pre-College Classroom……………………………………75
Studies of Young Adults’ Attitudes Towards
Knowledge and Learning…………………………………….77
Student Expectations in Undergraduate Physics……………..79

PART II. RESEARCH METHODS AND ASSESSMENT: HOW DO WE
DETERMINE WHAT STUDENTS ARE LEARNING?………………93

Chapter 3. OVERVIEW OF METHODS IN PHYSICS EDUCATION
RESEARCH………………………………………………………………93
Physics Education Research…………………………………………..93
Physics Education Research Methods………………………………….95
Overview of Research Methods Used in this Study……………….98

Chapter 4. MULTIPLE CHOICE TESTS:
THE FORCE CONCEPT INVENTORY…………………………….100
Chapter Overview……………………………………………………101
Development of the FCI……………………………………………102
Evaluation of FCI Results…………………………………………106
Validation and Reliability of the MDT & the FCI………………106
The MDT…………………………………………………………..106
The FCI…………………………………………………………….108
Results form FCI/MDT studies……………………………………111
Halloun, Hestenes, and the MDT (1985)…………………………111
The Hestenes, Wells, and Swackhamer Study (1992)…………..113
Hakes 6000 Student Study and the h-factor (1993-97)………….114
Discussion of What is Measured by the FCI……………………120
The 4-H Controversy (1995)…………………………………………120
Steinberg & Sabella’s Comparison of FCI Responses
& Exam Problems (1997)…………………………………………124
The Force and Motion Concept Evaluation (FMCE)……………129
Chapter 7: UNDERSTANDING STUDENT THINKING THROUGH INTERVIEWS

Overview .................................................................................................................. 231
Expectations ............................................................................................................ 235
MPEX Survey Protocol ......................................................................................... 235
Open MPEX Protocol ............................................................................................ 239
Concepts .................................................................................................................. 240
Waves-Math Demonstration Interview ................................................................. 240
Two Rock Problem Interview ............................................................................... 247
Limitations .............................................................................................................. 250

PART III. EVALUATION OF RESEARCH-BASED TEACHING METHODS

Chapter 8. COURSES, TEACHING METHODS, AND SCHOOLS ....................... 253
Overview ................................................................................................................. 253
Schools and Student Populations ......................................................................... 254
Teaching Methods .................................................................................................. 257
Traditional Instruction .......................................................................................... 258
Motivation ............................................................................................................... 258
Implementation ....................................................................................................... 258
Difficulties ............................................................................................................... 259
Evaluation ............................................................................................................... 260
Tutorials ................................................................................................................... 261
Motivation ............................................................................................................... 261
Implementation ....................................................................................................... 261
Difficulties ............................................................................................................... 263
Evaluation ............................................................................................................... 263
Group Problem Solving ......................................................................................... 264
Motivation ............................................................................................................... 264
Implementation ....................................................................................................... 265
Difficulties ............................................................................................................... 269
Evaluation ............................................................................................................... 270
Workshop Physics .................................................................................................. 272
Motivation ............................................................................................................... 272
Implementation ....................................................................................................... 273
Difficulties ............................................................................................................... 274
Evaluation ............................................................................................................... 275
Courses and Implementations ............................................................................... 277
Traditional .............................................................................................................. 277
Chapter 9. CONCEPTUAL UNDERSTANDING
Overview............................................................................................................. 301
Students’ Understanding of Basic Concepts .................................................. 304
   Overall Concept of Force and Motion ......................................................... 304
   University of Maryland Tutorials ................................................................. 304
   FCI results at other schools ......................................................................... 308
Velocity Graphs................................................................................................. 313
Newton's Third Law......................................................................................... 319
   FCI Newton 3 cluster at University of Maryland ........................................ 319
   FCI Newton 3 cluster for other research-based and traditional lecture curricula ........................................................................................................ 324
Representations and Application of Concepts in Complex Problems .............. 328
   Mechanics...................................................................................................... 329
      Velocity..................................................................................................... 329
      Newton's third law .................................................................................... 333
Beyond Mechanics .......................................................................................... 324
   Harmonic Oscillator .................................................................................... 334
   Two-Slit Interference .................................................................................... 341
Problem Solving Interview ............................................................................. 343
Summary........................................................................................................... 345

Chapter 10. STUDENT EXPECTATIONS
Overview............................................................................................................. 351
What are expectations? .................................................................................... 351
Why study expectations? .................................................................................. 351
Description of Study ....................................................................................... 353
Research Questions .......................................................................................... 354
Evaluation of the MPEX survey results ........................................................... 355
| Site visits and interviews                       | 356 |
| Chapter layout                                    | 357 |
| Student Expectations: Distribution and Evolution | 358 |
| Overall results from all schools                  | 364 |
| The Independence Cluster                          | 367 |
| The Coherence Cluster                             | 370 |
| The Concepts Cluster                              | 372 |
| The Reality Link Cluster                          | 375 |
| The Math Link Cluster                             | 376 |
| The Effort Cluster                                | 380 |
| Workshop Physics: Site Visits and Interviews       | 382 |
| Dickinson College: Pre-course and mid-year        | 382 |
| Nebraska Wesleyan University: Post sequence       | 385 |
| The view from the top: Charlie, John, and Amy      | 387 |
| The view from the bottom: Hanna and Kim           | 396 |
| Drury College: Post sequence                      | 408 |
| Discussion                                        | 409 |
| MPEX Survey                                       | 409 |
| MPEX Interviews                                   | 410 |
| Expectation issues                                | 410 |
| Implementation issues                             | 419 |

### PART IV. CONCLUSION

Chapter 11. CONCLUSION

Summary................................................................................................................. 423

Why are research-based curricula necessary?.................................................. 423
How do we evaluate research-based curricula?.............................................. 426
Evaluation of PER-based curricula............................................................... 428
Implications........................................................................................................ 436

### APPENDICES

APPENDIX A. Force Concept Inventory.............................................................. 442
APPENDIX B. Force and Motion Conceptual Evaluation..................................... 449
APPENDIX C. MPEX Survey.................................................................................. 456
APPENDIX D. Tutorial Materials......................................................................... 465
Velocity Tutorial................................................................................................. 466
Force and Motion Tutorial.................................................................................. 477
Newton’s Third Law Tutorial.............................................................................. 487
Harmonic Oscillator Tutorial............................................................................ 497
APPENDIX E. Student Volunteer Release Form for taping interview............... 506
APPENDIX F. Student Interview Responses to selected MPEX Survey Items...... 508
Item 2................................................................................................................... 509
APPENDIX G. Student Interview Transcript Summaries

Amy……………………………………………………………………... 519
John………………………………………………………………………… 527
Charlie……………………………………………………………………... 537
Ramsey………………………………………………………………………… 546
Krystal
Leb
Kim
Hannah
Roger

APPENDIX H. Factor Analysis of MPEX Survey Results... 580

APPENDIX I. Pre/Post MPEX Survey Results by Item... 60x

REFERENCES………………………………………………………………........ 6xx
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2-1</td>
<td>Detailed five-step problem solving strategy used in the University of Minnesota’s group problem solving curriculum.</td>
<td>48</td>
</tr>
<tr>
<td>Table 2-2</td>
<td>Comparison of (A) a typical textbook problem with (B) a context rich problem for an object on an inclined plane</td>
<td>55</td>
</tr>
<tr>
<td>Table 2-3</td>
<td>Hammer’s expectation dimensions of students learning and student learning types</td>
<td>80</td>
</tr>
<tr>
<td>Table 4-1</td>
<td>Newtonian Concept Taxonomy from the Force Concept Inventory</td>
<td>104</td>
</tr>
<tr>
<td>Table 4-2</td>
<td>A taxonomy of common-sense student beliefs about force and motion probed by the Force Concept Inventory</td>
<td>105</td>
</tr>
<tr>
<td>Table 4-3</td>
<td>Comparison of concepts needed to solve exam problems and FCI items</td>
<td>126</td>
</tr>
<tr>
<td>Table 5-1</td>
<td>MPEX dimensions of student expectations</td>
<td>144</td>
</tr>
<tr>
<td>Table 5-2</td>
<td>Prevalent responses of our expert group</td>
<td>165</td>
</tr>
<tr>
<td>Table 5-3</td>
<td>Percentages of the calibration groups giving favorable/unfavorable responses on Overall and Cluster MPEX survey</td>
<td>166</td>
</tr>
<tr>
<td>Table 5-4</td>
<td>Percentage of variance associated with the extracted factors</td>
<td>183</td>
</tr>
<tr>
<td>Table 5-5</td>
<td>Results of factor analysis for three, four, and five factor extraction</td>
<td>184</td>
</tr>
<tr>
<td>Table 5-6</td>
<td>Comparison of factors and clusters</td>
<td>186</td>
</tr>
<tr>
<td>Table 5-7</td>
<td>Cronbach Alpha for the overall, cluster and factor MPEX results</td>
<td>191</td>
</tr>
<tr>
<td>Table 5-8</td>
<td>Comparison of measured and calculated σ for overall and cluster MPEX results for the schools where data was collected from at least 6 classes</td>
<td>196</td>
</tr>
<tr>
<td>Table 7-1</td>
<td>Think aloud hints given to students at the beginning of demonstration or problem solving interviews</td>
<td>234</td>
</tr>
</tbody>
</table>
demonstration or problem solving interviews

Table 7-2  MPEX Survey protocol (Spring 1997 version)  236
Table 7-3  Open MPEX protocol  237
Table 7-4  Waves-Math Interview protocol  243
Table 7-5  Two-Rock Problem protocol  249

Table 8-1  A description of schools participating in this study  256
Table 8-2  Description of Introductory Calculus-Based Classes Studied  293
Table 8-3  Introductory course content coverage at the participating schools  294
Table 8-4  Summary of the data collected for each course  295

Table 9-1  Overall FCI results for University of Maryland Traditional Lecture and Tutorial Classes.  305
Table 9-2  Overall FCI Scores for all curricula  309
Table 9-3  Overall FMCE results for Workshop Physics schools  309
Table 9-4  Percentage error on the VQ with and without MBL  316
Table 9-5  Newton 3 FCI results for University of Maryland Classes  322
Table 9-6  Newton 3 FCI results for all four curricula  325
Table 9-7  Results on student constructions of the velocity graphs in the long qualitative exam problem from two classes at the University of Maryland  331
Table 9-8  Results on students’ use of Newton’s third law in the long exam problem from two classes at the University of Maryland  331
Table 9-9  Results on Newton 3 FCI questions for University of Maryland classes C2 and G2  331
Table 9-10  Summary of student responses from a University of Maryland Tutorial class to the two harmonic oscillators problem shown in Figure 9-9 before the harmonic oscillator tutorial was modified  336
Table 9-11  Summary of student responses from University of Maryland  340
Tutorial class to the two harmonic oscillators problems in Figure 9-10 after the harmonic oscillator tutorial was modified

Table 10-1. Institutions and classes from which pre, mid, and post MPEX survey data were collected. 359

Table 10-2. Pre-sequence percentages of students giving favorable/unfavorable responses overall and on the clusters of the MPEX survey. 360

Table 10-3. Pre-sequence and mid-sequence percentages of students giving favorable/unfavorable responses overall and on the clusters of the MPEX survey. 361

Table 10-4. Pre-sequence and post-sequence percentages of students giving favorable/unfavorable responses overall and on the clusters of the MPEX survey. 362

Table 10-5. Interview responses for items 1, 14, and 34 from students in the top third of their class at NWU. 389

Table 10-6. Interview responses for items 1, 14, and 34 from students in the lower third of their class at NWU. 397

Table 10-7. Responses from all NWU interviews for MPEX Survey items 17 and 33. 414
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2-1.</td>
<td>Conceptual and traditional exam problems on the subject of DC circuits</td>
<td>18</td>
</tr>
<tr>
<td>Figure 2-2</td>
<td>Test scores for the problems shown in Figure 2-1.</td>
<td>19</td>
</tr>
<tr>
<td>Figure 2-3.</td>
<td>Student group solution to the traditional textbook problem in Table 2-2A</td>
<td>56</td>
</tr>
<tr>
<td>Figure 2-4.</td>
<td>Student group solution to the context-rich problem shown in Table 2-2b</td>
<td>59</td>
</tr>
<tr>
<td>Figure 3-1.</td>
<td>McDermott’s iterative cycle of PER, curriculum development, and instruction with the Redish axle.</td>
<td>94</td>
</tr>
<tr>
<td>Figure 4-1.</td>
<td>Schematic of the Hake plot.</td>
<td>115</td>
</tr>
<tr>
<td>Figure 4-2.</td>
<td>Hake plot from 6000 student study of MD and FCI results</td>
<td>117</td>
</tr>
<tr>
<td>Figure 4-3</td>
<td>Exam problems used in Steinberg and Sabella study</td>
<td>125</td>
</tr>
<tr>
<td>Figure 5-1</td>
<td>A-D plot for the calibration groups, average of all survey items</td>
<td>165</td>
</tr>
<tr>
<td>Figure 5-2</td>
<td>Two dimensional view of a hexagonal piece of crystal</td>
<td>179</td>
</tr>
<tr>
<td>Figure 5-3</td>
<td>Scree plot: the # of Eigenvectors vs the # of factors</td>
<td>182</td>
</tr>
<tr>
<td>Figure 6-1</td>
<td>Student solution to a traditional textbook style problem</td>
<td>207</td>
</tr>
<tr>
<td>Figure 6-2</td>
<td>Student solution to a qualitative exam problem</td>
<td>208</td>
</tr>
<tr>
<td>Figure 6-3</td>
<td>A correct solution to the qualitative problem in Figure 6-2a</td>
<td>209</td>
</tr>
<tr>
<td>Figure 6-4</td>
<td>Student response to an essay exam question on the nature of waves</td>
<td>214</td>
</tr>
<tr>
<td>Figure 6-5</td>
<td>Newton’s third law pretest used in several first semester classes of the introductory physics sequence for engineering majors at the University of Maryland from 1994-1995</td>
<td>221</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6-6</td>
<td>Quick tally of the 86 student responses from the fall 1995 semester to the Newton’s third law tutorial pretest in Figure 6-5</td>
<td>222</td>
</tr>
<tr>
<td>6-7</td>
<td>Mathematical reasoning with waves pretest</td>
<td>225</td>
</tr>
<tr>
<td>7-1</td>
<td>Waves-Math Pretest from the 1996 Spring Semester</td>
<td>241</td>
</tr>
<tr>
<td>8-1</td>
<td>The arrangement for the Newton 3 tutorial</td>
<td>283</td>
</tr>
<tr>
<td>9-1</td>
<td>Overall FCI figure of merit histogram for classes at the University of Maryland</td>
<td>306</td>
</tr>
<tr>
<td>9-2</td>
<td>FCI and FMCE figure of merit histogram for classes from all ten schools participating in the study</td>
<td>310</td>
</tr>
<tr>
<td>9-3</td>
<td>Thorton-Sokoloff Velocity Graph Question (VQ)</td>
<td>314</td>
</tr>
<tr>
<td>9-4</td>
<td>Error rate on velocity questions (VQ)</td>
<td>317</td>
</tr>
<tr>
<td>9-5</td>
<td>Newton’s third law FCI questions (N3 FCI)</td>
<td>320</td>
</tr>
<tr>
<td>9-6</td>
<td>Histogram of average figures of merit for the Newton 3 FCI cluster for University of Maryland traditional and tutorial classes</td>
<td>323</td>
</tr>
<tr>
<td>9-7</td>
<td>Histogram of average figures of merit for the Newton 3 FCI cluster for all four curricula</td>
<td>326</td>
</tr>
<tr>
<td>9-8</td>
<td>Long qualitative exam problem requiring both the construction of a velocity graph and an application of Newton’s third law</td>
<td>330</td>
</tr>
<tr>
<td>9-9</td>
<td>Vertical harmonic oscillator problem for first tutorial class</td>
<td>335</td>
</tr>
<tr>
<td>9-10</td>
<td>Horizontal harmonic oscillator problem for second tutorial class</td>
<td>339</td>
</tr>
<tr>
<td>9-11</td>
<td>2-slit interference problem</td>
<td>342</td>
</tr>
<tr>
<td>9-12</td>
<td>Graph of student responses to 2 slit diffraction problem shown in Figure 9-11</td>
<td>342</td>
</tr>
<tr>
<td>10-1</td>
<td>Pre/Mid Redish Plots for all schools, average of all items</td>
<td>365</td>
</tr>
<tr>
<td>10-2</td>
<td>Pre/Post Redish Plots for all schools, average of all items</td>
<td>366</td>
</tr>
</tbody>
</table>
Figure 10-3.  Pre/Post Redish Plots for all schools, independence cluster  369
Figure 10-4.  Pre/Post Redish Plots for all schools, coherence cluster  371
Figure 10-5.  Pre/Post Redish Plots for all schools, concepts cluster  374
Figure 10-6.  Pre/Post Redish Plots for all schools, reality-link cluster  377
Figure 10-7.  Pre/Post Redish Plots for all schools, math-link cluster  379
Figure 10-8.  Pre/Post Redish Plots for all schools, effort cluster  381
PART I. INTRODUCTION AND BACKGROUND:
PHYSICS EDUCATION RESEARCH,
STUDENT LEARNING AND ASSESSMENT

Chapter 1. Introduction

Physics Education Research (PER) has led to the development of a large number of innovative instructional approaches to address student difficulties with traditional instruction. Yet, as a community, we are only beginning to determine how to evaluate these curricula appropriately. There are currently few widely accepted methods for determining if these improved curricula are more effective than traditional instruction in helping students reach a broad range of course goals. This dissertation examines the effectiveness of traditional instruction and three research-based curricula. It will also examine the methods used to evaluate effectiveness. The three research-based curricula are University of Washington Tutorials, University of Minnesota Group Problem Solving & Problem Solving Labs, and Dickinson College Workshop Physics.

I. MOTIVATION

Over the last twenty years, PER has changed our view of student learning in the traditional introductory course. The main findings of PER can be summarized in the following three points:

1. Traditional instruction is not working for many students in the introductory physics course.
2. Students are not “blank slates.” Students’ experiences and cognitive attitudes can affect what they learn in an introductory course. Many student experiences and beliefs are not compatible with what we want them to learn, hinder students’ learning of physics, and outlast instruction.
3. Research-based curricula can be developed to improve student learning by helping students change their common sense conceptual beliefs through active engagement. These new research-based curricula often require additional resources and changes in teaching style in order to be implemented effectively. The cost of change has caused many instructors to ponder if it is worth the effort. This question is difficult to answer unless we can determine if these new curricula help students learn more effectively than traditional instruction, particularly in institutions adopting curricula developed at other institutions. Thus, there is a great need to carefully examine how to evaluate the effectiveness of a new curriculum in terms of student learning. Before this issue can be discussed further, we need to consider what are appropriate course goals.

II. COURSE GOALS AND THE HIDDEN CURRICULUM

The first step in evaluating the success of any curriculum is to examine the goals of the course. The main goal in most traditional physics classes is for the students to demonstrate mastery of the course material through typical end-of-chapter problems on course assignments and exams. However, most physics instructors have goals for their students that go beyond this. These additional goals are often neither stated explicitly to the students in class nor reinforced through grading and testing. We refer to this kind of learning goal – a goal not listed in the course syllabus or in the textbook – as part of the courses’ “hidden curriculum.” Many students who are considered successful in traditional lecture classes are often unsuccessful with these additional learning goals. For example, many students who have mastered the main goal of solving end of chapter problems:

- have a weak grasp of basic physics concepts,
are unable to apply what they know to new situations,
believe that physics is just a collection of equations and procedures that deal
with very specific situations,
do not believe that physics has anything to do with their everyday life, and
do not see physics as a process of trying to make sense out of the physical
world.

Examples of these difficulties will be cited in chapter 2 and/or documented from detailed
observations throughout this dissertation.

The model of student learning used in this dissertation is a growth model rather
than a knowledge-transfer model. This model focuses on what is happening to the
student trying to learn rather than on what the teacher is doing. Based on this model and
the results stated above, the main goal of an introductory physics course, even for non-
majors, should be to help students build a good functional understanding of physics that
they can use to solve problems in new contexts, i.e. to become more like expert problem
solvers. This requires students to develop multiple skills including the following:

- to be able to understand and use the underlying physics concepts,
- to know when and where specific concepts apply,
- to be able to express their functional understanding in multiple
  representations including graphs, equations, and words, and
- to understand the nature of physics and how to use it effectively in and out of
  class.

Traditional exam problems and multiple-choice concept tests can help assess what is
happening in some of these goals, but not all. As we see in chapter 2 in the examples of
Mazur, Hammer, Tobias, and others, what traditional assessments tell us about student
learning is often unclear. What is needed is to develop an assessment strategy that
determines a more complete picture of these learning outcomes and what they tell us
about teaching effectiveness.
In developing such a strategy, we need to keep in mind that we do not expect students to become expert problem solvers as the result of one introductory physics sequence. However, we can determine elements, i.e. skills and qualities, that are on the path to the development of expert problem solving, scientific reasoning, and a scientific viewpoint. Once we identify and select the elements we want to study, then we can determine how to assess these elements to see if progress is being made.

III. PROBLEM STATEMENT/RESEARCH QUESTIONS

This dissertation discusses assessment issues, evaluates assessment techniques, and applies these techniques to evaluate the effects on student learning of three PER-based curricula compared with traditional instruction. Because PER-based curricula are fairly new and because so little has been done with evaluation in terms of the hidden curriculum, there are many issues to consider concerning course evaluation. To limit the scope of this dissertation, I focus on two aspects of student learning from the hidden curriculum: (1) conceptual understanding of physics and (2) student “expectations.”

1. Conceptual understanding of physics: Successful students should develop both a good understanding of physics concepts and the ability to use that understanding to solve physics problems. This includes the ability to use conceptual understanding in solving new and complex problems. This aspect is discussed in more detail in chapters 2, 4, 6, and 9.

2. Expectations: By “expectations” I mean the attitudes and beliefs students have about the nature of learning, physics, and mathematics. Studies by Hammer and Schoenfeld (discussed in Chapter 2) have shown that expectations can have a significant effect on how students learn and what they get out of a course. In particular, if the students’ expectations are different from the instructors, it can distort what the students get out of the class. This is discussed in more detail in chapters 2, 5, and 10.
The dissertation focuses on the following three questions:

1) What are the characteristics of different student populations coming into the calculus-based introductory physics class?
2) How do we determine if students are improving their knowledge of physics concepts and expectations?
3) Are the research-based curricula more effective for teaching students to improve their conceptual understanding and their expectations of physics?

IV. EXPERIMENTAL DESIGN

This dissertation evaluates both the methods used to determine what students learn and what these methods tell us about students’ learning. These methods include observations, survey instruments, interviews, and exams. The strengths and weaknesses of each method are discussed with examples. The two main survey instruments used are the Force Concept Inventory (FCI)\(^{10}\) and the Maryland Physics Expectation (MPEX) survey.\(^{11}\) The FCI is used to measure gains in students’ understanding of basic physics concepts while the MPEX survey is used to study student expectations. The development of the FCI by Hestenes et al. is discussed in chapter 4. The development of the MPEX survey by Redish, Steinberg, and the author is discussed in depth in chapter 5.

To determine the effect of instruction, the students were evaluated with concept tests and the MPEX survey at the beginning and end of the first quarter or semester of the introductory physics sequence. The MPEX survey was also given at the end of the first year of instruction.

Most students in service courses only see introductory physics topics from a physicist’s viewpoint once. Because of this and the increasing importance for physics departments to document learning outcomes for these classes, we have limited our study to introductory sequences that were designed primarily for non-physics majors. For this
dissertation, I selected the calculus-based introductory sequence. This sequence is particularly interesting for studying how students view the role of mathematics in physics since students in calculus-based physics courses should be expected to have a reasonably strong math background and to develop a mathematical understanding of introductory physics.

The students participating in this study were taught with one of four curricula which vary in the amount of active-learning activities used in class:

1. The traditional lecture method uses almost no active learning activities in lecture or recitation and its associated laboratories tend to be over structured (cookbook).
2. The University of Washington’s tutorial curriculum is a modification of the traditional lecture course that substitutes cooperative-group concept-building activities for the recitations but does not change the lecture part of the course.
3. In the University of Minnesota’s Group Problem Solving & Problem Solving Laboratory approach, the structure of the course is the same as the traditional lecture method, but the lecture, recitation, and laboratory are all modified into a more coherent course. The lecture emphasizes major themes and models a prescribed problem-solving strategy. Working in cooperative groups, the students use the prescribed problem-solving strategy to solve story problems and laboratory problems in recitation and laboratory, respectively.
4. *Workshop Physics* is a no lecture/all guided-discovery laboratory course developed at Dickinson College. This curriculum consists almost entirely of cooperative group active-learning activities and makes heavy use of microcomputer-based laboratory (MBL) tools for data acquisition and modeling.

The four curricula were studied at implementations at ten colleges and university across the United States. The ten schools either were asked to participate in this project because of the teaching methods used in the introductory course or asked Redish or the author to participate in the project to increase their awareness of student learning in the hidden curriculum. The implementations include both primary implementations, implementing curricula developed by that institution, and secondary implementations, in which institutions adopt curricula developed at other institutions. A detailed description
of both the curricula and the implementations is given in chapter 8. I conducted
interviews with students during site visits to five of the ten schools. Note that not all
types of data were collected at all ten schools.

V. DISSERTATION OVERVIEW

This dissertation is divided into the four parts:

Part I. Introduction and Background:
   Physics Education Research, Student Learning, and Assessment

Part II. Research Methods and Assessments:
   How Do We Determine What Students are Learning?

Part III. Evaluation of Research-based Teaching Methods

Part IV. Conclusion

Part I is an introduction to the dissertation and provides the reader with
background information on what is known from PER on student learning in physics
including the following: problem solving, conceptual understanding, students’ cognitive
beliefs or expectations and their implications for physics instruction (chapter 2), and an
overview of the research methods and models used in PER (chapter 3). In Part II, I
describe in detail the research methods used in this dissertation and I discuss the
reliability, validity, limitations, and what is learned from each method. The research
methods include: multiple choice concept tests (chapter 4), the Maryland Physics
Expectations (MPEX) survey (chapter 5), specially constructed conceptual quizzes and
exam problems (chapter 6), and interviews with students at five of the ten schools
(chapter 7). Part III contains a description of the three types of research-based
instruction and traditional instruction including details on the implementations at each of
the ten schools (chapter 8), assessment of students’ conceptual understanding (chapter
9), and assessment of students’ expectations (chapter 10). Part IV contains the conclusion which summarizes the dissertation results, discusses the implications for instruction and curricula improvement, and suggests directions for future study.

VI. DISSERTATION SUMMARY

In this dissertation, I am evaluating student learning with respect to conceptual understanding and expectations (cognitive or epistemological beliefs) in classes using either one of three PER-based curricula or traditional lecture instruction. The classes participating in this study were calculus-based introductory physics classes at ten undergraduate institutions including a community college, five small liberal arts colleges, and three large state universities. For this study, I collected four types of assessment data to study students’ expectations and students’ conceptual understanding: multiple choice concept test (mainly FCI) data, MPEX survey data, student interviews, and qualitative exam problems. The FCI is a nationally recognized tests of students’ conceptual understanding of mechanics.\textsuperscript{12} The MPEX survey is a new instrument to study student expectations developed specifically for this study in collaboration with Richard N. Steinberg and Edward F. Redish. The survey results are evaluated in terms of the overall result and six sub-scores or dimensions of expectation including:

- the role of the student in learning physics – \textit{independence},
- the structure of physics knowledge – \textit{coherence},
- understanding physics equations vs. just using them – \textit{concepts},
- the connection between physics and everyday life – \textit{reality link}, and
- the connection between mathematics and physical situations – \textit{math link}.  


Over one hundred hours of interviews were conducted with students at five of the participating institutions. Two specially designed exam problems were used in the introductory sequence at University of Maryland to look at students’ application of conceptual knowledge on final exams.

The results that I present in this dissertation will show the following five points:

1. There are significant differences among the different student populations participating in the study as they come into the introductory course with regards to both students’ conceptual understanding and expectations.

   Concepts: The three large state universities and Drury College had initial average FCI scores of approximately 50%. The three other private liberal arts schools had significantly lower FCI scores at the beginning of the semester. Their initial FCI average scores range from an average of 37.6% at NWU to 43.6% at Dickinson College.

   Expectations: The University of Minnesota and Dickinson College classes’ initial responses to the survey were significantly more favorable for learning physics than the classes at University of Maryland for three of the six MPEX dimensions measured: independence, reality, and effort. (In this dissertation, significant changes refer to statistically significant differences.) Nebraska Wesleyan University and Drury College classes’ initial expectations were statistically significantly more favorable overall and in two or more MPEX dimensions than classes at Maryland, Ohio State University, Moorhead State University, and Prince Georges Community College. In addition, students at Skidmore College and Ohio State had significantly more favorable reality link expectations initially than students at Maryland.

2. The three research-based curricula are more effective than traditional instruction in helping students learn key concepts.13
The figure of merit of basic conceptual understanding of key concepts in this dissertation is the fraction of the possible gain$^{14}$ achieved by the class from matched$^{15}$ pre- to post-test scores on the FCI.$^{16}$ In this study, the traditional lecture classes at University of Maryland had an average FCI fractional gain of 0.19±0.03. Courses using one of the three research based curriculum had average FCI fractional gains of 0.35±0.01 to 0.46 ±0.03. These results are consistent with the results of a previous 6000 student study by Hake of pre and post FCI scores from traditional and research-based instruction (see chapter 5).$^{17}$ In addition, the classes using research-based curricula had significantly better gains on Newton’s third law questions on the FCI, often found to be one of the hardest concepts in introductory mechanics. At Maryland, the classes using tutorials also did significantly better than traditional lecture classes in responding to multiple choice velocity graph questions.

3. For all classes participating in this study, overall student expectations as measured by the MPEX survey deteriorated at least slightly after one year of instruction, regardless of the curriculum. However, the four Workshop Physics courses and the Maryland sequence with two semesters of Tutorials did not have significant unfavorable shifts in overall expectations.

For the MPEX survey, shifts in student expectations are measured by comparing pre- and post-sequence matched student responses. A shift in student responses for a given sequence ≥ 2σ is considered to be significant.$^{18}$ The overall MPEX responses from students in introductory sequences after one year of traditional instruction and or one year of the Group Problem Solving curriculum became significantly more unfavorable for learning physics over the year. Recall that unlike the other research-based curricula, the emphasis of the Group Problem Solving curriculum is on problem solving, not conceptual understanding.
4. The MPEX responses from Dickinson College, where the Workshop Physics curriculum was developed and the Tutorial sequences at Maryland both improved significantly in one dimension, student expectations towards conceptual understanding.

The Workshop Physics sequences at Dickinson College where it was developed and Tutorial sequences with two semesters of Tutorials both had significant increases in the number of favorable responses to the survey on the items in the concept dimension. It should be noted that the other Workshop Physics classes were either in their first or second year of implementation. Dickinson College had been using Workshop Physics for several years before their participation in this study. Surprisingly, only the Group Problem Solving sequence at Minnesota showed a significant decrease in this dimension.

5. For all four curricula, student expectations deteriorated over the year in terms of connecting physics to the real world. Two of the Workshop Physics sequences, both Group Problem solving sequences, and the traditional sequence at Minnesota deteriorated in student expectations on linking mathematics to physical situations.

The only sequences where student responses did not become more unfavorable with regards to the link between physics and the students’ everyday experiences was the Workshop Physics class at Drury College. The decrease in favorable responses was significant for the sequences taught with traditional instruction and Group Problem Solving as well as the Workshop Physics sequence at Moorhead State University. In addition, responses from half the sequences studied including the Workshop Physics sequences at Nebraska Wesleyan and at Moorhead State and the sequences at Ohio State and Minnesota regarding the connection between math and the system under study became significantly more unfavorable.

6. The mechanics exam problem and the interviews indicate that many students are having difficulty applying mechanics concepts to open-ended qualitative problems. The two harmonic oscillator problems indicate that many
Maryland students have trouble connecting concepts and graphs to physical situations in problems. After completing a modified tutorial, the students showed significant improvement on a similar problem.

The Tutorial students did significantly better than the students who had traditional recitations on the qualitative mechanics exam problem shown in Figure 6-2 in drawing a velocity graph and ranking forces for the accelerating two-cart system. However, the number of correct student responses was less than might be expected from the post FCI results. This discrepancy suggests that while the multiple choice questions provide an indication of students’ conceptual knowledge, a good score on a test like the FCI may overestimate students’ ability to use their conceptual knowledge on more open-ended qualitative problems.

The two harmonic oscillator problems (Figures 9-9 & 9-10) require students to qualitatively graph the displacement vs. time curves for two spring-mass systems released at rest at t = 0 on the same axis. On the first harmonic oscillator final exam problem, approximately one-quarter of the students drew graphs that showed the correct basic features of the two curves including starting at maximum displacement. After changes were made to the tutorial, approximately three quarters of the tutorial students drew curves that had the correct basic features.

Using the combined research methods of surveys, non-traditional problems, and interviews to study student learning in classes using research-based and traditional lecture instruction, this dissertation provides new insights into students’ conceptual understanding and the role of expectations in learning physics. This is the first wide-scale study of the implementation of Tutorial, Group Problem Solving, and Workshop Physics curricula that includes secondary implementations, i.e. schools that adopt a
curriculum instead of developing their own. The results of this study will help give
researchers and instructors a better understanding of the issues in the hidden curriculum
for consideration in curriculum design and/or implementation in the calculus-based
introductory physics sequence.

1 L.C. McDermott and P.S. Shaffer, *Tutorials in Introductory Physics* (Preliminary

2 P. Heller, R. Keith, and S. Anderson, “Teaching problem solving through cooperative
grouping. Part 1: Group versus individual problem solving,” *Am. J. Phys.* 60, 627-
636 (1992); P. Heller and M. Hollabaugh, “Teaching problem solving through
coopeative grouping. Part 2: Designing problems and structuring groups,” *Am. J.
Phys.* 60, 637-644 (1992); P. Heller, T. Foster, and K. Heller, “Cooperative group
problem solving laboratories for introductory courses,” in AIP Conference Proceeding No. 399 *The Changing Role of Physics Departments in Modern
Universities: Proceedings of the International Conference on Undergraduate Physics
Education*, edited by E.F. Redish and J.S. Rigden (AIP Press, Woodbury NY, 1997),
913-934.


ccepts using real-time microcomputer-based laboratory tools,” *Am. J. Phys.* 58 (9),
is learned — Closing the gap,” *Am. J. Phys.* 59 (4), 301-315 (1991); A. van
Heuvelen, “Learning to think like a physicist: A review of research based
and G. Swackhammer, “Force concept inventory,” *Phys. Teach.* 30, 141-158 (1992);
P.S. Shaffer, *Research as a Guide for Improving Instruction in Introductory Physics*,

5 L. Viennot, “Spontaneous reasoning in elementary dynamics,” *Eur. J. Sci. Educ.* 1,
understanding of the concept of velocity in one dimension,” *Am. J. Phys.* 48, 1020-
understanding of the concept of acceleration in one dimension,” *Am. J. Phys.* 49, 242-
253 (1981); A. Caramaza, M. McCloskey and B. Green, “Naive beliefs in
‘sophisticated’ subjects: Misconceptions about trajectories of objects,” *Cognition* 9,
117-123 (1981); J. Clement, “Students’ preconceptions in introductory mechanics,”


8 See Ref. 4.


12 See Ref. 10.

13 Preliminary data suggests that with research-based instruction, student performance on traditional measures is as good as with traditional instruction and may be better.

14 The fraction of possible gain h is given by the following equation:

\[ h = \frac{\text{class post-test average} - \text{class pre-test average}}{100 - \text{class pre-test average}} \]

This quantity is discussed in more detail in chapter 4.

15 Here “matched” means that only students who took both the pre-test and the post-test FCI are included in the analysis.

17 See Ref. 10.

18 The standard deviation $\sigma$ for an MPEX measurement is discussed and defined in chapter 5.
Chapter 2. Background:

An Overview of Relevant Education Research

In the last twenty years physics education research (PER) has shown that the traditional lecture course is not working for many of the students in introductory physics courses.\textsuperscript{1, 2, & 3} PER has succeeded in defining some of the fundamental problems with the introductory course and how to solve them. PER and other areas of educational research including math education and cognitive science have contributed greatly to our understanding of student learning in the introductory course. This chapter is an overview of those education research results relevant to this investigation.

I. WARNING SIGNS

I will discuss three examples that illustrate the need for assessment that takes into account conceptual understanding and student expectations. The first example is from Eric Mazur,\textsuperscript{4} a physics professor at Harvard, the second from David Hammer\textsuperscript{5} while he was a graduate student at Berkeley, and the third comes from Sheila Tobias,\textsuperscript{6} a non-traditional researcher of student difficulties with math and science. These three studies represent a small fraction of the available literature; they were selected primarily for their value as specific illustrations of what happens in traditional introductory physics lecture courses.

A. Mazur’s Example: Student Difficulties With Conceptual Understanding

Eric Mazur is the Gordon McKay Professor of Applied Physics and a professor of physics at Harvard University. He currently leads a research program in applied physics and maintains an active interest in innovative instruction. He has developed his
own strategy for teaching physics to lecture classes which is described in his book, *Peer Instruction: A User’s Manual*. Mazur taught the introductory course for science and engineering majors for six years in the traditional lecture format before he became aware of a series of PER articles by Halloun and Hestenes concerning the persistence of students’ common sense beliefs about mechanics. He was at the time quite satisfied with his teaching. In Mazur’s own words,

...I taught a fairly conventional course consisting of lectures enlivened by classroom demonstrations. I was generally satisfied with my teaching - my students did well on what I considered difficult problems and the evaluations I received from them were very positive. As far as I knew there were not many problems in my class.

Like most instructors when they first become aware of the poor performance of students on multiple choice concept tests like the Force Concept Inventory (FCI) following instruction, Mazur thought, “Not my students.” When he gave his students the test, his first warning that there was a problem was when one of his students asked, “Professor Mazur, how should I answer these questions? According to what you taught us, or by the way I think about these things?” Despite the concept tests’ simplicity, the students scored little better than they had on their last midterm examination which Mazur thought was of far greater difficulty.

To better understand these results, Mazur gave his students the two problems shown in Figure 2-1 as the first and last problems on a midterm examination in a traditional lecture class. Note that the first problem is purely conceptual and requires only a basic understanding of simple DC circuits. The second problem is a traditional circuit problem that might be found in any introductory physics text. Problem 2 deals
1. A series circuit consists of three identical light bulbs connected to a battery as shown here. When the switch S is closed, do the following increase, decrease, or stay the same?
   (a) The intensities of bulbs $A$ and $B$
   (b) The intensity of bulb $C$
   (c) The current drawn from the battery
   (d) The voltage drop across each bulb
   (e) The power dissipated in the circuit

2. For the circuit shown, calculate (a) the current in the 2-Ω resistor and (b) the potential difference between points P and Q.
Figure 2-2: Test scores for the problems shown in Figure 2-1. For the conceptual problem, each part was worth a maximum of 2 points. (a) Test score distribution for the conceptual problem. (b) Test score distribution for the traditional problem. (c) Correlation between the scores on each problem. The radius of each data point is a measure of the number of students represented by that point. (Figures from E. Mazur, *Peer Instruction: A User's Manual*, Prentice Hall 1997.)
with the same concepts as problem 1 but also requires setting up and solving equations using Kirchhoff’s laws. Most physics instructors consider the first problem to be easier than the second. The results shown in Figure 2-2 show that the students disagree. Both problems were graded out of ten points. The average on the conventional problem is significantly higher than on the conceptual problem (6.9 vs. 4.9). Figures 2-2 a and b are histograms of the student scores for the two problems. Note that two-thirds of the students scored at least 7 out of 10 on the conventional problem while only one third of the students did as well on the conceptual problem. In fact, almost one third of the students were unable to answer more than one of the five parts correctly.

The reason for the large peak at two for the conceptual problem was that 40% of the students only answered 1b correctly. Mazur suggests that these students believe that closing the switch doesn’t change the current through the battery but that the current splits into two at the top junction and rejoins at the bottom. Despite this many of them still managed to correctly solve the traditional problem. Also note that the result on problem 1b is an overestimate since the responses of students who believe that bulb C will dim rather than go out will be scored correct.

In Figure 2-2c the student scores for each problem are plotted against each other. Notice the apparent lack of correlation between student scores on the two problems. The broad diagonal band represents the boundary for roughly equal scores (± 3 points) on the two problems. Although 52% of the students lie on the band, 39% of the students did substantially worse on the conceptual problem. A large number of these students managed to score 2 points or less on the conceptual problem while getting 10 points on the traditional problem.
According to Mazur, this result was repeated many times on pairs of problems during the remainder of the semester. The students performed significantly better on standard end-of-chapter textbook problems than on simple conceptual problems covering the same material. Mazur suggests that this example has three implications for physics instructors:

1) It is possible for students to do well on conventional problems by memorizing algorithms (recipes) without understanding the underlying physics.
2) It is possible even for an experienced instructor to be misled by traditional assessments into thinking that students have been taught effectively.
3) Students who use this algorithmic approach may believe they have mastered the material and then are severely frustrated when they learn that their plug-and-chug strategies don’t work on all problems.

B. Hammer’s Example, Two Approaches to Learning Physics

One of the few studies of student epistemological beliefs on learning and physics in introductory undergraduate physics courses was done by David Hammer while he was a graduate student at University of California, Berkeley. In his early work on expectations, Hammer reported results from his case studies on students in the introductory “pre-med” physics course at UC Berkeley. In this study, Hammer investigated three questions:

1. What general conceptions of physics do students have, if any?
2. How do these conceptions affect their understanding and performance?
3. How are these conceptions affected by the instruction in the class?

Students were selected at random from the roster and asked to participate in the study until five students had volunteered. Each student was interviewed individually five times during the semester. Each interview took about an hour and involved a variety of tasks including open-ended discussions of the student’s impressions of the course and of
physics, semi-directed tasks such as going through a midterm exam, and specific discussions of physics concepts as well as quantitative and quantitative problem solving.

Hammer chose to report initially on two of the five students, Liza and Ellen (pseudonyms) who on paper appeared very similar. They were both planning to go to medical school: both had math SAT scores around 700 and had A’s in mathematics courses through calculus. Liza’s record was stronger, including a 5 on the BC Calculus Advanced Placement Exam, an A- in her first semester chemistry course, and an A in her high school physics class. Ellen received a C in the same chemistry course and had not taken a physics course beforehand.

1. Two approaches to learning physics

Hammer chose to report on these two students because, unlike the other subjects in the study, Liza and Ellen both put a great deal of time and effort into the course and, while their grades were similar (Liza B+, Ellen B), initially their approaches to the course were not. Liza relied heavily on memorization and pattern matching without trying to understand the material. From Hammer’s paper,

Throughout her transcripts there are both explicit and implicit indications that Liza’s approach to the course was to learn formulas and facts based on the authority of the instructor and the text. To Liza, the formulas were the physics and that the professor said it in lecture or that it was in the book constituted sufficient justification.

She almost always said she understood the lecture or the reading; it was only when I pressed for explanations beyond citations that she ‘didn’t think about it’ or was ‘not sure’.” [Excerpt from Hammer’s transcripts follows]

Liza: …he [the instructor] defines the acceleration of circular motion in this formula, and then later on he proves it.

Hammer: OK, how does he prove it?
Liza: *Um, (flips pages) this acceleration is referred to, um, inward acceleration.*

Hammer: Why is the acceleration inward?

Liza: *He said there is an inward acceleration for circular motion.*

Hammer: Does that make sense?

Liza: (pause) *I didn’t think about it* (laughs).

She solved problems by ‘figuring out which formula to use:’

Liza: *I look at all those formulas, say I have velocity, time, and acceleration, and I need to find distance, so maybe I would use a formula that would have those four things.*

But having an explanation for the formula itself did not seem relevant.

Liza: *This if v f equals v oh plus a t ( v f = v oh + at ).*

Hammer: OK, and where did that come from?

Liza: *It came from here* (laughs and indicates book).

Hammer: Why is it true?

Liza: *Well, there is an initial velocity, and there is a final velocity, we know what the a is which is the acceleration, and I just use this to find what time it takes.*

Liza felt no need to check presented facts for consistency … .

On the other hand, Ellen tried to build her own understanding of the course material. She worked very hard to reconcile what she was learning in class with her experiences and her intuition. In Hammer’s words,

Ellen’s attitude was more independent, and, to her, the formalism was only one way of looking at the physics. She did want to make sense of the material, to integrate it with her own intuitions … Ellen criticized lectures for failing to connect ‘theory’ with ‘reality,’ for emphasizing the formalism over simple explanations. She felt, for example, that the professor’s treatment of finding the range of a projectile was needlessly complicated: [excerpt from transcript follows]

Ellen: *…obviously when it stops being in the air it stops going horizontally. … it seems like we spent a couple of lectures just*
trying to get that through people’s heads, and I’m sure if you just say that to someone they’ll (say) well obviously. … I guess that’s what it is, we get theory with theory, and then we get application with application.

Hammer: What do you mean by theory?

Ellen: It means formulas, … let’s use this formula because it has the right variable, … instead of saying OK, we want to know how fast the ball goes in this directions, because if we know that, all we have to do is find out how long it goes in that direction before it hits the ground and we can find out how far. [Note that this quote seems to indicate she believes that the theory is the mathematical formalism.]

After the first few weeks she became frustrated, finding it difficult to reconcile different parts of the formalism with each other and with her intuition … Eventually she compromised her standards:

Ellen: I’d rather know why for real.

Hammer: What do you mean by ‘know why for real?’

Ellen: Well, sometimes I rationalize things, and I know that it’s probably not the soundest rationalization, or something, it’s just that I can kind of make sense for the time being, if I don’t have time to really figure everything about it…even though I don’t really know how it was derived.

Like Liza, Ellen was able to apply prescribed methods to solve problems, but for her this was not sufficient:

Hammer: so part b, can you do it?

Ellen: No. (laughs) I can pretend to. I can do it the way I know it’s supposed to be done, but I pretty much already did that.

Both students put in a great deal of time and effort to learn physics in this course. Liza worked extra problems and attended every lecture, recitation, and lab. While Liza felt she was succeeding in learning physics, Ellen did not feel she was getting much out of the course. While at least initially she became comfortable with the material, even then she was aware of large gaps in her understanding.
2. Two approaches to problem solving

The two approaches to learning physics had a significant effect on how the two students approached physics problems. As Hammer relates,

An assigned problem asked: “One ball is thrown horizontally with a velocity \( v_0 \) from a height \( h \), and another is thrown straight down with the same initial speed. Which ball will land first?” Liza wrote out \( x = x_0 + v_0 t + \left( \frac{1}{2} \right) a t^2 \) for the horizontal and vertical components of each ball, substituting appropriate initial positions, speeds, and accelerations. [Author’s note: this problem is very difficult for introductory students to solve symbolically using only the equations for position and velocity as a function of time.] She then used the vertical components to show, not without some difficulty, that the second ball travels a greater vertical distance after time \( t \) than the first, so it hits first. Ellen, in contrast, said immediately that the answer was ‘obvious,’ and went on to explain qualitatively that the second ball hits first because it has a greater speed downward to cover the same vertical distance.

… When Ellen was able to make sense of the material, her intuition helped her understand the formalism and guided her solutions. After the first few weeks, however, Ellen’s attitude interfered with her ability to answer the course questions. For a while she continued to try to reason based on her own sense of things, rather than accepting what she was told or answering by rote, but her intuitions were not generally in accord with Newtonian physics.

Liza, following the algorithms provided, often failed to apply her own knowledge from everyday life. She almost certainly had the common sense to know that the ball thrown down would hit first, but she did not think to make use of it in this context. At no time did she acknowledge that this was the answer one would expect. However, with her methodical adherence to the procedures, she was more reliable than Ellen in answering correctly.

3. Implications

For Liza, understanding physics meant knowing the facts and formulas and being able to apply them to problems. She did not deliberately or consciously use her intuition or her experience. According to Hammer, the effect of Liza’s approach was
experience. Her understanding remained incoherent and fragmented …, but it served her comparatively well on problem sets and examinations. Ellen, when her approach was successful, was able to bring fragments of her understanding together, to integrate different pieces from the course with each other and with her intuitive knowledge. Unfortunately, it was not successful after the first few weeks.

Most physics instructors would prefer their students to approach an introductory physics more like Ellen than Liza, to think and reflect on the material while trying to reconcile intuition based on experience and the course material. Yet, as the course proceeded, Ellen had more and more trouble trying to build a comprehensive understanding of the course material while Liza consistently did well in the course. By the second midterm exam, Ellen was struggling and scored below the mean. Eventually she was only able to get a B in the course by abandoning her approach and doing “what everyone else does” by adopting an approach more like Liza’s.

It is also worth noting Liza’s response in her last interview when she was asked if she had liked the course. She answered, “Not really…it was kind of boring…all those formulas and derivations.” When asked if she was describing the course or physics in general, she replied, “Both.”

The course did not support Ellen’s thinking, reflective approach but seemed to reward and encourage Liza’s memorization approach. While Liza’s approach allowed her to be ‘successful in the class’ as measured by grades, the course left her with a very naive and boring view of what physics is and what it means to do physics. Is this course an aberration or a typical course? Again from Hammer’s paper,

This was a standard introductory physics course. The method of instruction was to present material, demonstrate its validity, provide examples of its application, and assign further problems for students to do on their own. Lectures paralleled the text at a pace of about a chapter a
week. … Similarly, two of the ten labs involved determination of an empirical rule … All others concerned verification of known results or the use of known results to measure some quantity. In every case, the procedures were specified.

This description would fit many traditional lecture format classes. Hammer indicates that while he only reported on the results for two students, the other students in the study made similar comments implying that these two students are not atypical. What is happening in the typical traditional lecture course to cause results like these?

The previous two examples illustrate some of the difficulties with the introductory physics course taught by the traditional lecture method uncovered by PER. Note that in both cases, traditional assessment offered no indications of problems with the course. Both Mazur and Hammer offered explanations of what was wrong with traditional lectures. Mazur uses an analogy to explain.

If I were lecturing not on physics but, say, on Shakespeare, I would certainly not spend the lectures reading plays to the student. Instead, I would ask the students to read the plays before coming to the lecture and I would use the lecture periods to discuss the plays and deepen the students’ understanding of and appreciation of Shakespeare.

Yet in physics (and many other introductory courses) lecturing is often little more than reading the book to the students with demonstrations as illustrations. But the problem with traditional lectures is more complicated than that. To explain his observations described in the previous section, Hammer offers a more analytical view.

The flow of reasoning was always from the theory to the phenomena, the flow of information always from the professor and text to the students. The students applied the theoretical concepts of the course, but they were not involved in or even witness to the formation of those concepts. The laws were simply provided, and students were to become familiar with them through practice in solving problems. …

The emphasis on formalism and the goal of problem solving facility seem consonant with conceptions of physics as a collection of facts and
formulas. The style of instruction, with knowledge passing from instructor to student, procedures and results to experimentation specified in advance, seems consistent with a reliance on authority. Furthermore, the pace of the course, with a chapter of reading and ten problems due every week, would not allow much independent exploration, except for those students who find the material easy.

Here, Hammer is making some serious accusations concerning why the course seemed to encourage student views of learning and physics that are contrary to what most instructors want. He obviously believes that the problem is caused by the nature and structure of the traditional lecture course, specifically the emphasis on problem solving and the over-reliance on authority. However, he also states that these results must be considered tentative both for the small sample size and the inexactness of the process of deriving information from student comments. But he is not alone in his finding that the nature and structure of the traditional lecture course is interfering with student learning. Similar results were found in a study conducted by Sheila Tobias.

C. Tobias’ Example: Observations on the Traditional Lecture Teaching Method

In her study, Tobias was trying to learn why many qualified undergraduate students with good math and science backgrounds decide to turn away from majoring in science after exposure to introductory courses. Of college freshman who switch out of science and engineering majors, only about a third switch because they found the course work too difficult. For the rest, 40% find other fields more interesting and 25% believe they would have better job prospects elsewhere. She refers to these students who have the ability to pursue science but choose not to as “the second tier.” These second tier students are some of the same students who are not being reached by traditional lecture
instruction. In this study, she wanted to understand the second tier students’ view of the process and problems of learning science in these introductory classes.

Because she wanted her observers to record their thoughts on the learning process as it was happening, she could not use students who had already taken introductory science classes. Instead she chose to use as her observers mature postgraduate or senior undergraduate students with good high school math and science background who chose not to take science in college. All the observers had demonstrated ability in their own fields. Tobias believed (correctly as we see below) that the responses made by the observers would be typical of students in the introductory science courses.

Three humanities graduate students (Eric, Jackie, & Michel) and one fifth-year social science undergraduate (Vicki) acted as student observers in three different calculus-based introductory physics classes at the University of Arizona and the University of Nebraska, Lincoln. To make sure they had the proper background, the study candidates were required to have taken at least one semester of college calculus in addition to four years each of high school mathematics and science. Two of the student observers with rusty but solid calculus skills, Eric & Jackie, did very well in their respective courses. Another student observer, Michel, whose math skills were weaker, struggled in the course but was able to learn the concepts and solve the assigned problems. The other student observer, Vicki, found that what had worked for her in the social sciences did not work for her in physics. She had to struggle and would have dropped the course if she were not involved in the project. Tobias considers Vicki’s viewpoint to be closest to that of an average “C” student.
Each student observer was paid to seriously audit one semester of different calculus-based introductory physics courses as observers. In return for the stipend, they were expected to perform as well as they could in the class including attending all classes, submitting all work, and taking all examinations up to the final. In addition, they were expected to keep a journal where they would monitor the instructor’s style of presentation, the material in the book, and the assignments, as well as to record their personal experiences with the course and where possible those of their classmates. At the end of the semester, they were asked to address the following two questions in a final essay:

1. How and in what ways was this course different from other college courses you have taken in other fields?
2. What were the specific knowledge, skills, and experience deficits you noticed in yourself and in your fellow students that got in the way of your mastery of the material?

Since her study also uses a very small sample set, Tobias engaged Abigail Lipson, a psychologist and senior member of the Harvard University Bureau of Study Counsel, to compare the student observers’ results with an analysis of interviews of eighty science and non-science Harvard-Radcliffe students at the beginning of each of the students’ four years of study. These interviews were part of a much larger study of Harvard-Radcliffe students who were traced, tested, and interviewed throughout their undergraduate careers to look at predictors of success in science and non-science degrees as well as gender differences. Although the Tobias’ student observers were not typical of the students who participated in Lipson’s study, the student observers’ reactions to the courses and the subject matter were similar to those of Lipson’s interviewed Harvard-Radcliffe students who started as science majors and then switched out.
While Tobias’ student observers’ comments on the their own progress and difficulties are very illuminating, it is their observations on how the class was taught that help us understand what is happening in traditional lecture classes. For example:

Eric: *To some extent science is hard because it simply is hard. That is to say, the material to be learned involves a great many concepts, some of which are counterintuitive. The process of mastering these concepts and being able to demonstrate a computational understanding of actual or theoretical situations required a great deal of time and devotion. In my experience, this fact is well understood by the students, the professor, and the general public. What is not as well understood are the various ways in which this already hard subject matter is made even harder and more frustrating by the pedagogy itself.*

All four student observers commented on three particular aspects of the class that they felt made learning physics more difficult:

1. The lack of a narrative or story line,
2. An overemphasis on quantitative problem solving, and
3. A classroom culture that discourages discussion and cooperation.

To understand the nature of these difficulties, it is instructive to review some of their comments on the standard lecture course. These comments include excerpts from their journals and the final essay as well as comments from Tobias. Despite being in different classes, the comments are very similar.

**1. Lack of a narrative or story line**

In Hammer’s study, Liza, the student who focused on developing recipes for problem solving from the start, indicated that to her physics was a collection of equations and formulas that applied to many different situations. She did not see physics as the process of learning and applying a few fundamental ideas to understand many different situations. While she could solve the assigned problems, she missed the big picture.
This may have contributed significantly to why she found both the course and the subject boring.

In Tobias’ study, the professors’ lectures apparently failed to help the student observers see common themes and links in the material. In fact, Eric, Jackie, and Michel all found it patronizing not to be told in advance where they were headed or what they needed to understand as indicated by their comments below.

Eric: We had marched through the chapters, doing the required work but never digging deeper … I was able to keep myself on track by concentrating on one chapter at a time. But I never got the idea that the professor had any understanding of how the concepts were related, as he rarely tied together information from more than one chapter. His lectures did not seem to build upon each other, and he gave no indication of a linear movement through a group of concepts … The final then asked the most primary basic questions about only the most important laws of physics. We were not required, at any time, to interrelate concepts or to try to understand the “bigger picture.”

Jackie: Why, I wanted to know, did we begin by studying only the idealized motion of particles in straight lines? What about other kinds of motion? If he [the professor] could tell us what’s coming next, why we moved from projectile motion to circular motion, for example, I would find it easier to concentrate; I’d know what to focus on. In college, I always wanted to know how to connect the small parts of a large subject. In humanities classes, I searched for themes in novels, connections in history, and organizing principles in poetry.

Each of the four student observers were accomplished undergraduates and had at least four years of experience in undergraduate courses in other fields. Yet, not one of them was able to see the connections or the context in what they were learning in the introductory physics course.

2. Overemphasis on quantitative problem solving

The exams, homework assignments, and lecture emphasized algorithmic problem solving, especially ‘how much’ problems. All four student observers were more
interested in building an understanding of the physics concepts, in particular, the ‘why’ questions. As Michele noted above and again below, she felt her curiosity was not supported by the course.

Michele:  *My curiosity simply did not extend to the quantitative solution. I just didn’t care how much. I was more interested in the why and the how. I wanted verbal explanations with formulae and computations only as a secondary aid. Becoming capable at problem solving was not a major goal of mine. But it was a major goal of the course.*

Eric made a similar observation.

[Eric] “Eric was learning that, for the most part, ‘why’ questions are neither asked nor answered. The preference is for ‘how’ questions.”

Initially this caused Eric to have a very negative view of the course. Later he found,

Eric:  *As I am able to ask more knowledgeable questions, class becomes more interesting. I am finding that while the professor is happy to do example problems for the entire period, he will discuss the real world ramifications of a theory if asked.*

Note that this professor was only willing to discuss the ramifications in response to Eric’s questions. What would the professor do if no questions were asked? In addition, his fellow classmates did not appreciate Eric’s questions. They lost patience with his silly ‘why’ questions when they felt that they got in the way of finding the right solution to their assigned problems. According to Eric, this was what physics was all about for them. The other student observers made similar observations about their classmates.

Vicki:  *When [John, a classmate she was studying with,] brought up an equation, I would try to relate it to another equation to help me learn them both all the better. This confused the effort. When John works a problem he uses only what is necessary and brings nothing else into the process. If I am going to learn a subject I need to know what is similar and dissimilar about an item. Why, when you are pushing down on a moving block, is there no “work” done? Isn’t there a force downward on the block? I know that force and displacement have to be in the same direction; I needed to relate this to the concepts in the problem…. John*
seemed disturbed by this and thought I had not yet mastered the fundamentals.

Jackie: [In her study group, the other students’] concerns focused on the kinds of problems they would encounter on the exams, not at all on a general understanding of the concepts … They ignored all the fun parts, seeing the whole picture, laying out the equations and solving these. Instead, they wanted to know what equaled what and solve for an answer. The elegance of problem solving was lost …

Jackie later speculated that this might be less a difference of mind and more a matter of efficiency.

Jackie: I think the students around me are having the same sort of thought-provoking questions about the material that I put into my journal, but under time pressure they don’t pursue them, eventually they learn to disregard “extraneous” thoughts and to stick only to the details of what they’ll need to know for the exam. Since the only feedback we get is on the homework assignments, the students cannot help but conclude that their ability to solve problems is the only important goal of this class.

Part of the reason for this attitude in the students may have been caused by the exams. The student observers noted that the exam problems were usually more like the easier homework problems rather than (in their opinion) the more interesting harder problems.

Michele: Too easy exams in contrast to too hard homework. On philosophy exams, instructors expect their students to do more than what they’ve done before, not less.

Jackie: It [the exam] was nothing like the homework problems. It was simple. It didn’t really test understanding. There were so many things I thought I was going to have to know that weren’t on the exam: that normal forces are what a scale reads, the direction and nature of frictional force, ..., the difference between kinetic and static friction, what an inclined plane does to free fall, ..., etc. I don’t completely get all of this. But these are the questions I was thinking about when I prepared for the test.

Eric: The problems [on exams] seldom required the use of more than one concept or physical principle. Only once were we asked to explain or comment on something rather than complete a calculation.
Eric found the four class exams in his class biased towards computation and away from conceptual understanding. While he understood that some level of conceptual understanding was required to complete the computations, he found that the level was not particularly high.

One consequence of this emphasis on quantitative problems is that physics becomes less of a creative process and more of a craft. This, also, is reflected in the student observer comments below:

Eric: *I do not feel that what this professor is doing can be considered teaching in any complex or complete sense. My understanding is that we are to learn primarily by reading the text, secondarily by doing problems on our own and comparing our solutions to those on sale in the physics office, and thirdly by mimicking the professor’s problem solving examples. Simply by intuition I know physics, and more generally science to involve creativity and finesse; but this man makes it more into a craft, like cooking, where if someone follows the recipe, he or she will do well.*

Jackie: *Learning to solve physics problems is a process, not a matter of insight...Understanding the free body diagram means knowing how to do them.*

[Tobias comments:] The problems were of limited interest because they had all been solved before. Only occasionally did these exercises provide intellectual satisfaction; rarely were they a source of new insight. [The student observers] looked upon the effort to be training at the expense of education in science, too many scales, not enough music.

As Tobias notes, in addition to their view of creativity in introductory physics, the emphasis on quantitative problem solving also influenced their opinion of the design and goal of the course. In essence, to the student observers the course values mastery over understanding.

Michele: *A course design that assumes that everyone in the class has already decided to be a physicist and wants to be trained, not educated, in the subject.*
Vicki: … so the physics instructor could not imagine mastery being demonstrated by anything but increasing skill at problem solving.

3. A classroom culture that discourages discussion and cooperation

The way material is presented in the traditional lecture format makes it appear that the book and the instructor are the acknowledged authorities and they are passing what they know to students who learn it and repeat it back in homework and exams. The student is trying to learn what the instructor knows. In this transmissionist view (knowledge is transmitted from the authority to the student) of learning, the emphasis is on learning facts, not understanding. It is interesting to note that the student observers perceived their courses to be ‘low’ on concepts & theory and mired in facts (dry formulas and dull reality).

As Eric’s comment below indicates, this model of instruction can cause students to approach the course in ways that hinder student learning.

Eric: I still get the feeling that unlike a humanities course, here the professor is the keeper of the information, the one who knows all the answers. This does little to propagate discussion or dissent. The professor does examples the ‘right way’ and we are to mimic this as accurately as possible. Our opinions are not valued, especially since there is only one right answer, and at this level, usually only one right way to get it.

Part of this perception may be due to the course emphasis on quantitative problems discussed above, but part of it is due to a lack of student discussion in and out of class on the ideas and concepts being learned. The students do not feel they have ownership of the material; they have not made it an intrinsic part of what they know and understand. All of the observers commented and speculated on the passive nature of their classmates in lecture. For example, Vicki and Eric noted,
Vicki: No one ever asks questions in class except for that rare “Will you clarify?” question. I feel like it is grab and run.

[And in a later journal entry] Some people seemed to go to class only to hand in their homework. … Others would attend the lectures purposefully to get information in order to digest it later, and in private. I wanted to digest it there, in class, through questions and discussion. I learn verbally. I like being put on the spot. I am not as passive as they seemed, to me, to be.

Eric: [after learning of the low class average on quizzes and exams] What this means is that there are a good many people sitting quietly and not asking questions. This is always the case to some extent in college, but physics seems harder on these people than the humanities.

Even outside of class, when Eric asked his classmates about what they were studying, they weren’t able to articulate an answer.

Eric: I wonder if this is because they lack communication skills or because they haven’t had the time to reflect on what they have learned, or because they don’t really know much about their subject — if knowledge is defined to mean a deep thoughtful understanding rather than a superficial ability to regurgitate formulas.

Some of their observations speculate as to why students are so passive in lecture

Jackie: When he goes through these problems, the work seems so obvious, the equations so inevitable, that I tend not to question what he’s doing…Lectures in physics can be incredibly passive experiences for students, particularly dangerous for those who believe that if they can follow the professor, they’ve mastered the material.

This makes an interesting contrast to situations which the student observers thought were helpful to them in learning the course material.

[Tobias comments:] Vicky eventually found that her best studying came while working with at least one other student, teaching the material to one another. Here she was able do what she could not do in class: question and try out what she thought she understood, and then question again. … The best class in Eric’s view was one where the professor brought in five or six demonstrations, the results of which were counter-intuitive, and then asked the class to speculate as to why the particular results occurred. In this class, there was substantial interchange.
The comments shown below suggest that part of the problem lies in the course format and culture.

Eric: *The lack of community, together with the lack of interchange between the professor and the students combines to produce a totally passive classroom experience...The best classes I had were classes in which I was constantly engaged, constantly questioning and pushing the limits of myself and the subject and myself. The way this course is organized accounts for the lack of student involvement...The students are given pre-masticated information simply to mimic and apply to problems. Let them rather be exposed to conceptual problems, try to find solutions to them on their own, and then help them to understand the mistakes they make along the way.*

Both Eric and Vicki believed that part of the problem was due to classroom competition.

Vicki noted,

Vicki: *[In my other classes] learning is done through discussion with other students and with the professor. [She found it demoralizing to be working alone, in isolation, in a culture she characterized as destructively competitive.] I have the answer. Do you? If you don’t, I’m not going to share it with you.*

In addition to hindering their learning of physics, this competition could also have unhealthy results later in the students’ careers.

Eric: *[The sense of competition is in no way beneficial.] It automatically precludes any desire to work with or to help other people. Suddenly your classmates are your enemies. ... My class is full of intellectual warriors who will some day hold jobs in technologically-based companies where they will be assigned to teams or groups in order to collectively work on projects. [But] these people will have had no training in working collectively. In fact, their experience will have taught them to fear cooperation, and that another person’s intellectual achievement will be detrimental to their own.

Another factor causing the lack of community in the physics classroom may be due to the models of doing physics as portrayed in class and in the textbook. In both cases, physics is often portrayed as being done by the scientist working alone. This can
be compounded if, as was true in Vicki’s case, the students have no experiences of group work in recitation or lecture to contradict the model of the lone scientist.

4. Implications

The four student observers, like Hammer’s student who tried to build her own understanding, exhibited beliefs and attitudes toward learning and physics that most instructors would greatly desire to see in their students. Particularly, their drive for a deep, useful understanding of physics beyond traditional problem solving, their need for debate and discussion, and their need for a coherent structure in what they were learning. Yet, as is the case of Hammer’s student, the class did not support these attitudes. In fact the classes seemed to encourage students to memorize algorithmic solutions to physics problems rather than using an understanding of the physics concepts and discourage discussion of the course material in or out of class. Even when the student observers met with study groups outside of class, the focus of the discussion was on the mechanics of the solution, not a discussion of the physics.

In both examples expounded by Hammer and Mazur, non-traditional research-based assessment methods showed indications of inconsistencies in what the students were learning that were not apparent using traditional assessment. Here, each of the four student observers felt that the traditional exams were narrowly focused on quantitative problem solving and were not reflective of what they were learning. If we wish to improve student learning in the introductory course, both the traditional lecture format and the assessment methods must change.
D. Summary

In the three examples discussed in this section, we have seen research results that suggest that even when students do well in lecture-format classes with traditional textbook-style problems, some of them have a very superficial understanding of physics. We have also seen that this superficial understanding seems to be due to a lack of coherent themes, an overemphasis on quantitative problems, and a lack of meaningful student discussion on the principles, ideas, and concepts of introductory physics in and out of class. Traditional assessment and student evaluations showed no indication of any of these problems.

Another interesting point is that the studies by Mazur, Hammer, and Tobias make use of many of the research methods currently used in physics education research. Mazur used the FCI, a multiple choice concept evaluation, and specially designed exam problems. Hammer used extensive interviews or case studies. Tobias’ study used student interviews, classroom observations, and reflective journals. These are all forms of alternative assessment.

Before we can discuss how to use these methods to evaluate traditional and innovative instruction, we need a better understanding of the student difficulties that will need to be addressed. The remainder of this chapter is an overview of what is known regarding the issues of complex problem solving, conceptual understanding, multiple representations, and cognitive development for students in introductory physics classes.
II. PROBLEM SOLVING: EXPERTS VS. NOVICES

Over the last twenty-five years, cognitive scientists and artificial intelligence researchers as well as math and physics education researchers have studied how people solve physics problems. The key point relevant to this investigation is that there are significant differences between novice and expert problem solvers. This section will summarize the nature and results of some of the more pertinent studies. For further information, the reader is referred to David Maloney’s excellent review article on problem solving in physics, which is paraphrased extensively in this section.20

A. Characterizing Expert and Novice Problem Solvers

A number of problem-solving studies have focused on identifying differences between expert and novice problem solvers as well as good and bad problem solvers. According to Maloney, interview studies of this type found differences in knowledge structure, approach to problem solving, and the use of examples and readings. These studies are generally done by interviewing individuals as they perform a problem-solving task or by studying computer models designed to simulate human problem-solving behavior. Most of the interview studies classify graduate students and professors as expert problem solvers while students from introductory physics sequences are considered novices. Novice students are further classified as either good problem solvers or poor problem solvers on the basis of how well they do on traditional textbook-style physics problems on class exams.
1. Knowledge structure

Although physics content knowledge is an integral part of physics problem solving, de Jong and Ferguson-Hessler argue “that just having the knowledge is not sufficient, it must be organized in a useful manner.” The following results tend to support this view.

- Experts problem solvers have extensive domain knowledge that is organized hierarchically, with general physics principles at the top. Novices have significantly less domain knowledge that, when organized, tends to be organized around the surface features of physical situations.

- Experts tend to classify problems by the physical principles involved in the solution of the problem. Novices tend to classify problems based on physical objects, configurations, properties, and concepts explicitly described in the problem statement.

- Good problem solvers had better knowledge organization than poor problem solvers.

2. Problem solving approaches

One major emphasis of the studies on problem solving from the late 1970s through the early 1980s was to explore the differences between the problem solving approaches of novices and experts. The results can be summarized as follows:

- Expert problem solvers make use of and often require a detailed qualitative representation to plan their solution before working forward to a solution. Novice problem solvers tend to start with an equation that contains the unknown quantity asked for by the problem and work backward to the given information (means-end analysis).

- Expert problem solvers plan their solutions more carefully and in greater detail before carrying them out.

- Expert problem solvers make greater use of physical reasoning.

- Expert problem solvers also conduct an exploratory analysis, evaluate their thinking, and check their solutions to make sure they are reasonable when solving problems.

Two of the studies found “strong evidence” that one of the main causes of poor problem solvers’ difficulties is their failure and/or inability to construct an appropriate qualitative
representation. They suggest that the inability to use qualitative representations is due to poor understanding of the physics concepts.

3. Use of example problems and reading

In 1989, Chi et al. reported on a study of how students used worked examples when they are trying to learn how to solve problems. This study used “think aloud” interviews (described in more detail in chapter 7). They observed what students did when studying the examples and how they used the examples when they were solving problems. Chi et al. found the following differences between good and poor students:

- The good students were significantly better at working out the missing steps and identifying the points they did not understand in an example. They would continue to study an example until they did understood it. The poor students tended to walk through the examples without checking to see if they understood them.
- Good students worked out the procedural aspects of the problems that are usually left implicit in textbook presentations, but poor students needed assistance in identifying and understanding these aspects.
- Good students tended to refer to a short segment of an example when they needed a specific relation or when they wanted to compare what they had done on a problem. Poor students tend to go back and read major segments of examples in search of procedures to use on the problem that they were trying to solve.

Fergusson-Hessler and de Jong conducted a similar study investigating how students read segments of text. At selected points in the text, the students would be asked to think aloud about what they were doing. The students were judged as “good” or “poor” performers on the basis of test scores from three exams. The think alouds from five students who did well on all the exams and from five students who did poorly on all three exams were analyzed. Ferguson-Hessler and de Jong found the following result:

- Good performers used more deep processing than the poor performers when reading the text. Deep processing here means imposing structure not given in the text and making procedures and assumptions explicit.
• Poor performers tended to take more for granted and focused more on declarative knowledge (principles, formulas, and concepts) when reading the text. Good performers focused more on procedural (when a particular relation is used) and situational (characteristics of problem situations) knowledge.

B. How to Help Novices Become Experts

The research summarized in the previous section helps us describe the characteristics of expert problem solvers. But the real question from an instructional standpoint is, “How do we help students learn to become good problem solvers?” Although this is a current area of research, few studies have been done so far and not much is known. However, the results that do exist are encouraging. In this section, I will describe several studies of methods of instruction designed to encourage students to become better problem solvers.

1. Instruction to improve problem solving

As part of their research identifying the factors that contribute to good problem solving, Larkin and Reif tested methods of improving problem solving ability. They found that students who were taught with either an explicit problem solving strategy or taught in a way that emphasized a hierarchical physics knowledge structure with the main principles at the top performed significantly better at problem solving than control groups. Although these studies were successful, they were conducted on a small scale over a relatively short time period and not incorporated into a regular introductory physics course.

Three classroom-based studies on improving physics problem solving have been reported since the characteristics of good and poor problem solvers discussed above appeared in the PER literature. All three studies were conducted in modified
undergraduate introductory physics classes with typical class-size samples. Two additional studies that report on changes in the student’s knowledge use are also described.

a. WISE

In the first study, conducted by Wright and Williams at a community college in 1986, the students were taught and encouraged to use an explicit problem solving strategy as part of the Explicitly Structured Physics Instruction (ESPI) system. The WISE problem solving strategy consists of four steps described below:

1. **What’s happening**
   - Identify givens and unknowns
   - Draw a diagram
   - Identify the relevant physical principle

2. **Isolate the unknown**
   - Select an equation
   - Solve algebraically
   - Look for additional equations if necessary

3. **Substitute**
   - Plug in both numbers and units

4. **Evaluate**
   - Check the reasonableness of the answer

Wright and Williams found that students who used the WISE strategy on homeworks and exams performed significantly better (as determined by course grade) than students who did not use the strategy. In addition, student comments were strongly favorable to the WISE strategy and class retention was better for the experimental class. However, the authors found that the students were reluctant, even actively opposed in some cases to adopting the WISE problem strategy because of the extra work.
b. Overview, Case Study

In 1991, Van Heuvelen reported on the Overview, Case Study (OCS) approach to physics instruction.⁴⁹ Although this approach restructures the entire format of the class, it does specifically address problem solving. An integral feature of the OCS approach is the explicit development of multiple representations, especially qualitative physics representations, in problem solving. The approach also includes explicit discussion of knowledge hierarchy, an emphasis on active reasoning, and the use of cooperative group activities. Students who were taught with the OCS approach scored significantly better on problems from the College Board Advanced Placement physics test than students who were taught with traditional instruction at the same school.

c. Group Problem Solving with context-rich problems

In 1992, Heller and the Physics Education Group at University of Minnesota reported on a research-based lecture-recitation-laboratory curriculum they developed that emphasizes group problem solving.⁵⁰ In their curriculum, the lecture component stresses underlying themes, i.e. the main principles of physics, and an explicit problem solving strategy that is presented and modeled for the students. In recitation, the students do group problem solving using the strategy, and in lab they work in groups on laboratory problems where they must decide on what measurements to make and how to analyze the data they collect to answer the lab problem. This is one of the three research-based curricula being evaluated in Part III of this dissertation. A more detailed description of the curriculum and their own evaluation can be found in chapter 8.

In lecture the students were taught a five-step problem-solving strategy.⁵¹ A detailed outline of the strategy is shown in Table 2-1. The five steps are as follows:
1. Visualize the problem (make a physical representation of the situation)
2. Describe the problem in physics terms
3. Plan a solution
4. Execute the plan
5. Check and evaluate

Notice that this five-step plan is very similar to the WISE plan discussed above but with less of an emphasis on the mathematical calculation. To get the students to use this strategy productively, they found it necessary to use specially constructed problems called “context-rich problems” for the groups to use in recitation and lab. These
Table 2-1: Detailed five-step problem solving strategy used in the University of Minnesota’s group problem solving curriculum. [Reference Heller Part 1]

1. **Visualize the problem**
   Translate the words of the problem statement into a visual representation:
   • draw a sketch (or series of sketches) of the situation;
   • identify the known and unknown quantities and constraints;
   • restate the question;
   • identify a general approach to the problem—what physics concepts and principles are appropriate to the situation.

2. **Describe the problem in physics terms (physics description)**
   Translate the sketch into a physical representation of the problem:
   • use identified principles to construct idealized diagram(s) with a coordinate system (e.g., vector component diagrams) for each object at each time of interest;
   • symbolically specify the relevant known and unknown variables;
   • symbolically specify the target variable (e.g., find $v_0$ such that $h_{\text{max}} \geq 10 \text{ m}$).

3. **Plan a solution**
   Translate the physics description into a mathematical representation of the problem:
   • start with the identified physics concepts and principles in equation form (e.g., $\ddot{a}_s = \Delta v_s$, $\Sigma F = ma$);
   • apply the principles systematically to each object and type of interaction in the physics description (e.g., $N_i - W_i \cos \theta = m_i a_{ix}$ and $W_i = m_i g$);
   • add equations of constraint that specify the special conditions that restrict some aspect of the problem (e.g., two objects have the same acceleration, $a_1 = a_2$);
   • work backward (from target variable) until you have determined that there is enough information to solve the problem (the same number of independent equations as unknowns);
   • specify the mathematical steps to solve the problem (e.g., solve equation #2 for $N_i$, then substitute into equation #1, etc.).

4. **Execute the plan**
   Translate the plan into a series of appropriate mathematical actions:
   • use the rules of algebra to obtain an expression with the desired unknown variable on one side of the equation and all the known variables on the other side;
   • substitute specific values into the expression to obtain an arithmetic solution.

5. **Check and evaluate**
   Determine if the answer makes sense:
   • check—is the solution complete?
   • check—is the sign of the answer correct, and does it have the correct units?
   • evaluate—is the magnitude of the answer reasonable?
context-rich problems are more complex than typical textbook problems. They are described in more detail in the next section.

Heller et al. designed this curriculum so that three out of six hours of class time each week are spent on group problem solving. They hypothesized that,

…in well functioning groups, students share conceptual and procedural knowledge as they solve a problem together. [That is, the students discuss their ideas, their strategies, and their understanding of the physics of the problem.] During this joint construction of a solution, individual group members can request explanations and justifications from one another. This mutual critique would clarify all the members’ thinking about the concepts and principles to be used, and how those concepts and principles should be applied to the particular problem. Moreover, each member can observe others perform the varied thinking strategies that he or she must perform independently and silently on individual problem assignments.

To evaluate this curriculum, Heller et al. made extensive observations, studied the copied solutions, surveyed the students, and interviewed them. Although they did not report on how this type of activity affects the students’ thinking and reasoning processes, they did show that groups that have more elaborate discussions produce better qualitative descriptions\textsuperscript{53} and that students who were taught with this curriculum produced better solutions.

Class exams included both an individual and a group component. By comparing group and individual solutions to exam problems rated equally difficult by a rating scheme of their own design, Heller et al. showed that the group solutions were consistently better than the solutions of the best individuals. The greatest difference between the individual and group solutions was in the qualitative analysis in steps two and three.
The students also did better on traditional textbook problems. One exam composed entirely of typical textbook problems was given to one class taught with the modified curriculum and one class taught with the traditional lecture, recitation, and laboratory curriculum. The solutions were judged on the expertness of the approach to the solution, not on the final answer.

The results from these three studies indicate that research-based instruction methods can help introductory physics students develop expert problem solving skills. Students can be taught an expert-like problem solving strategy and they can learn to use qualitative representations. In each case, learning these skills had a positive effect on the students’ problem solving abilities. But is there evidence of improving students’ depth and organization of knowledge?

d. Studies on improving students’ depth and organization of knowledge

In their studies of the use of computer aids to help students solve problems, Dufresne et al. found that students who used a computer tool they created, the Hierarchical Analysis Tool (HAT), showed significant improvement in problem solving in terms of reaching the correct solution and applying the correct principle. The HAT program asks the student questions about the primary principle and the ancillary concepts that apply to the problem in question. It helps the student to work from the primary principle to the specific case of the problem. Students who initially based their reasoning on surface features or a mixture of physical principles and surface features showed the most improvement. This result is even more remarkable than it seems at first because the students were not given any feedback about their performance. The change is a direct result of using the HAT.
A related study by Volet on modified instruction in computer science found that students taught a programming strategy that involved reflection and evaluation of a computer program did not know significantly more content than the control group, but they were better able to apply their computing knowledge to an unfamiliar complex problem on the final exam. This study is particularly interesting for three reasons.

First, the typical introductory computer science class in this study is very much like a traditional introductory physics course, taught with a lecture and a recitation but no laboratory. New programming concepts and ideas are presented to the students in lecture. The students are assigned programming problems each week to apply what they learn in lecture. In the typical recitation, the students work on their homework problems while waiting for the recitation instructor to come around and answer their questions individually. Except for the individual instruction in recitation this sounds very much like most physics classes.

Second, the curriculum modifications are similar to the group problem solving approach used by Heller et al. (described above) with some interesting exceptions. There is no laboratory component and the only part of the course that was modified was the recitation section. The students were taught an explicit 5-step programming strategy. The five steps are as follows:

1. problem definition
2. algorithm development (list the procedure step by step in plain English)
3. conversion of the algorithms into a flowchart or pseudocode representation
4. coding from the flowchart or pseudocode into a specific programming language
5. execution of the code, debugging of errors, and improvement of the program
Note that this strategy is similar to the WISE strategy and the one used by Heller et al. The strategy was modeled extensively for the students by the recitation instructor, but the majority of the time in recitation the students were working in groups modeling and coaching each other on exercises similar to the homework problems. The students were required to verbalize not only their results but to go through the programming process as a think aloud with opportunity for group discussion at every decision point. Very little, if any, time was spent by the instructor working with the students individually. Special efforts were made to create a cooperative atmosphere that would be more conducive to learning. A partners system was set up at the start of the semester to demonstrate that collaboration was a normal expectation of the instructor. The students were encouraged to work with their partners and/or their groups in and out of recitation.

Third, the results of the modified instruction were extremely positive. Volet compared students in the two experimental sections to an equal number of students in the same class but in different recitations with 4 different instructors. The experimental and control students were paired based on the students’ background in computing, overall program of study, gender, interest in computing, and initial study goals for the computing course. The final exam was graded by a professor who was blind as to the type of instruction experienced by the students. Parts one and two of the final exam asked students questions on their factual and procedural knowledge, for example, asking them to describe programming concepts, functions, and procedures and asking them why certain techniques are used. Part three required the students to solve an unfamiliar, fairly complex programming problem. While the experimental group did not do significantly better than the control group on parts one and two, they did do significantly better on
part three. Volet claims this result indicates “that experimental and control students did not differ in the amount of computer programming knowledge they had acquired (as assessed in parts one and two [on the final]),” but the result of part three “indicates that the experimental students’ computing knowledge was more accessible and more usable than control students’ knowledge.” In addition, the experimental group did significantly better in more advanced computer science classes.

We see from these studies that it is possible for introductory physics instruction to help students acquire at least some expert problem-solving skills. But there are additional issues to consider.

2. Appropriate problems

As we saw in the examples of Mazur, Hammer, and Tobias earlier in this chapter, the fact that students can solve traditional textbook problems is not always a good indication that students have a good understanding of the physics of a situation. Are traditional textbook problems useful for learning physics? Are they useful for teaching problem solving?

For the purposes of teaching an explicit problem solving strategy, there is another aspect to consider. While developing the group problem solving curriculum described above, Heller et al. studied what type of problems would be useful for teaching students to use the five-step strategy. They found that this is a difficult proposition for individual problem solvers because in their words,\textsuperscript{56}

\begin{quote}
if the problems are simple enough to be solved moderately well using their novice strategy, then students see no reason to abandon this strategy — even if the prescribed strategy works as well or better. If the problems are complex enough so the novice approach clearly fails, then students are
\end{quote}
initially unsuccessful at using the prescribed strategy, so they revert back to their novice strategy.

In researching what types of problems are effective for promoting the use of the prescribed “expert” problem solving strategy in group problem solving, Heller and Hollabaugh began by studying student problem solutions and group interactions for standard textbook end of chapter problems to characterize the typical novice strategy for solving this type of problem. An example problem for motion on an inclined plane is shown in Table 2-2 Problem A. When solving problems of this type, the group discussions tended to revolve around “what formulas should we use?” rather than “what physics concepts and principles should be applied to this problem?” One group, whose solution was typical, used the following solution process:

1. Rather than begin with a discussion and analysis of the forces acting on the block, this group began by attempting to recall the force diagram and formulas from their text. The example from the text was for a book sliding down the ramp. As a result, their frictional force was in the wrong direction and there was a sign error in their force equation.

2. At no point did the students plan a solution. They plugged numbers into formulas and manipulated equations until they had a numeric solution.

3. The group discussion was concerned with finding additional formulas that contained the same symbols as the unknown variables. They did not discuss the meaning of either the symbols or the formulas. They incorrectly tried to substitute the formula for instantaneous velocity into the formula for average velocity.
Table 2-2: Comparison of (A) a typical textbook problem with (B) a context rich problem for an object on an inclined plane

Problem A. A Typical textbook style problem

A 5.0 kg block slides 0.5 m up an inclined plane to a stop. The plane is inclined at an angle of 20° to the horizontal, and the coefficient of kinetic friction between the block and the plane is 0.60. What is the initial velocity of the block?

Problem B. Context-rich problem

While visiting a friend in San Francisco, you decide to drive around the city. You turn a corner and find yourself going up a steep hill. Suddenly a small boy runs out on the street chasing a ball. You slam on the brakes and skid to a stop, leaving a skid mark 50 ft long on the street. The boy calmly walks away, but a policeman watching from the sidewalk comes over and gives you a ticket for speeding. You are still shaking from the experience when he points out that the speed limit on this street is 25 MPH.

After you recover your wits, you examine the situation more closely. You determine that the street makes an angle of 20° and that the coefficient of static friction between your tires and the street is 0.80. Your car’s information book tells you that the mass of your car is 1570 kg. You weigh 130 lb and a witness tells you that the boy had a weight of about 60 lbs and took 3.0 s to cross the 15-ft wide street. Will you fight the ticket in court?
Figure 2-3: Student group solution to the traditional textbook problem in Table 2-2A.58

A typical incorrect solution of a group for a standard textbook problem. The arrows show the progression of the mathematical solution.
Their solution is shown in Figure 2-3. Heller et al. estimate that two-thirds of the groups used this formula-based approach. They concluded that standard textbook problems were not effective in promoting group discussions that would help the student become better problem solvers.

Next, Heller and Hollabaugh compared textbook problems to real world problems to determine which characteristics of textbook problems encourage the use of novice strategies and which characteristics of real world problems require the use of an expert strategy. They found several characteristics of textbook problems that encourage students to use the formulaic approach described above “despite the instructor’s effort to teach a more effective strategy.” Textbook problems typically use idealized objects and events that have little or no connection to the student’s real world. They suggest that this reinforces the student’s tendency to memorize algorithms to deal with specific objects or situations. The unknown quantity is specified in the problem (usually in the last sentence) and all the other quantities needed to solve the problem are given (usually with the correct units). This encourages students to solve the problem by searching for formulas that have the right quantities and then plugging in numbers until a numeric answer is obtained. This numeric answer can then be checked in the back of the book.

On the other hand, solutions to real world problems are motivated by the solver wanting to know something about actual objects or events with which the solver is familiar. Before any calculations can be done, the solver must decide which quantities are useful for solving the question, which physics concepts and principles are relevant, what additional information is needed, and determine which information can be determined and which must be estimated. In other words, students solving a real world
problem have to think about the problem, try to understand what is going on, and make a number of decisions before reducing the problem to plug and chug mathematics. Since most textbook problems have removed the need for this type of analysis, they make algorithmic problem solving appear to be the correct way to solve problems.

To encourage students to use the prescribed problem solving strategy, Heller and her group created what they call “context-rich problems.” The context rich problems are designed to utilize many of the characteristics of real world problems. They are short stories that include a reason (although sometimes humorous or farfetched) for calculating specific quantities about real objects or events. They are more complex than typical textbook problems. In addition, they typically have one or more of the following characteristics:

1. The problem statement may not specify the unknown variable.
2. More information may be provided than is necessary to solve the problem.
3. Some information may need to be estimated or recalled.
4. Reasonable assumptions may need to be made to simplify the problem.

An example of a context rich problem is shown in Table 2-2B. This is the inclined-plane textbook problem rewritten in context-rich form. According to Heller and Hollabaugh,59

Because context-rich problems are complex and involve making decisions about physics concepts and principles new to beginning students, they are difficult and frustrating even for the best students. In cooperative groups, however, students share the thinking load and can solve these problems. Because decisions must be made, the context-rich problems forced the groups to discuss physics issues while practicing effective problem-solving techniques. The group practice enhanced the students’ ability individually [as well] …

The students have to pool what they know of the actual behavior of objects and the physics principles and concepts that describe this behavior to solve context-rich problems. Figure 2-4 is an example of a good group solution to the traffic ticket
problem shown in Table 2-2B. The following is a description from Heller and Hollabaugh of how the group came up with this solution,

The students first sketched the situation and discussed what variable was needed to answer the question: “Will you fight the traffic ticket in court?” They decided they should calculate the initial velocity of the car just before the brakes were applied to see if this velocity was above the speed limit of 25 mph. After drawing the kinematics diagram, they then discussed what information they needed to find the initial velocity. They decided they could ignore the information about the child, since “the car stopped before it hit the child.” They then spent several minutes drawing free body diagrams of the car and discussing whether they needed to use static friction, kinetic friction, or both. During this discussion, they referred several times to the friction experiments they were doing in the laboratory. Once this issue was resolved and the force diagram agreed upon, they systematically planned a solution following the planning procedure modeled during lectures.

Notice that this group focused on “what physics concepts and principles should be applied to this problem” rather than “what formulas should we use.” While the students attitudes towards using the prescribed strategy for context-rich problems improved, they still found using the strategy “annoying” or “frustrating” to use on simple textbook problems because the strategy required them to write down more than they thought was necessary. This reaction is particularly interesting since Heller and Hollabaugh note that “these students were not usually successful at solving these problems using the formulaic strategy they preferred.” However these same students did agree that the prescribed strategy was useful for solving the more difficult textbook problems in addition to individual and group context-rich problems.

Heller and Hollabaugh’s study strongly suggests one reason why students who do well on traditional textbook problems may not do as well on qualitative problems.
Figure 2-4. Student group solution to the context-rich problem shown in Table 2-2b.

Visualize:

\[ \begin{align*}
\text{D} & \quad 0 = 20^\circ \\
D & \quad 30 \text{ ft} \\
\text{speed limit} & \quad 35 \text{ mph} \\
\mu_k & \quad 0.60 \\
\rho & \quad 1570 \text{ kg} \\
\text{Deceleration} & \quad 30 \text{ lbs}
\end{align*} \]

Question: Is the speed faster or slower than 25 mph?

Physics Description:

Free-body Diagram

\[ \begin{align*}
W & \quad \text{weight of car and driver} \\
F_N & \quad \text{normal force} \\
F_k & \quad \text{kinetic force of friction} \\
v_i & \quad \text{initial velocity of car} \\
v_f & \quad \text{final velocity of car (b)} \\
t_i & \quad \text{initial time when brakes slammed on (b)} \\
t_f & \quad \text{final time when car stopped}
\end{align*} \]

Question: Is \( v_f \) less than 25 mph?

General Principles:

\[ \begin{align*}
W & = mg \\
F_a & = \mu_k F_N \\
v_f & = \frac{v_i + v_f}{2} \\
\Delta v & = a \Delta t
\end{align*} \]

Plan:

1. To find \( v_f \)
   \[ a = \frac{\Delta v}{\Delta t} \]
2. To find \( t_f \)
   \[ t_f = \frac{D}{v_i} \]
3. To find \( V \)
   \[ V = t_f \cdot t_i \]
4. To find \( a_i \)
   \[ \Delta s = a_i \cdot t_f + \frac{1}{2} a_i \cdot t_f^2 \]
5. To find \( s_i \)
   \[ s_i = \frac{W}{g} \]
6. To find \( F_a \)
   \[ F_a = \mu_k \cdot F_N \]
7. To find \( F_k \)
   \[ F_k = F_a - W_f \]
8. To find \( w_f \)
   \[ w_f = \frac{W_f}{m} \]

There are 8 equations and 8 unknowns.

Solve \#8 for \( W_f \), substitute into \#7 to find \( F_k \). Substitute \( F_k \) into \#6 to find \( F_a \). Solve \#5 for \( W_i \), substitute \( F_k \) and \( W_i \) into \#4 and isolate \( a_i \). Equate \#2 and \#3 and solve for \( t_f \). Substitute \( t_f \) and \( a_i \) into \#1 to find \( v_f \).

Execute (only last steps shown):

\[ \begin{align*}
v_f & = \sqrt{2 \left( \frac{D}{\cos \theta} \right) \cdot \left( \frac{g}{\sec^2} \cdot \sec \right)} \\
\text{Units:} & \quad \sqrt{\left( \text{ft} \right) \cdot \left( \text{ft} \right)} = \frac{\text{ft}}{\text{sec}} \cdot \frac{\text{ft}}{\text{sec}^2} = \text{ft/sec} \\
v_f & = \sqrt{2 \cdot 50 \cdot \left( \frac{30}{\sec^2} \right) \cdot (0.6 + 0.94 + 0.34)} \\
& = 53.8 \text{ ft/sec}
\end{align*} \]

Change to mph:

\[ \begin{align*}
v_f & = \left( \frac{0.528 \text{ ft}}{\text{sec}} \right) \cdot \left( \frac{1 \text{ mile}}{5280 \text{ ft}} \right) \\
& = 36.7 \text{ miles/hr}
\end{align*} \]

You were speeding -- you better pay the fine!
Because students don’t use physical reasoning to solve problems and their grades depend on how well they do on the typical textbook problems, they may not see conceptual understanding and physical reasoning as important. This result is consistent with the findings of Maloney and Siegler as well as the previously described research characterizing novice problem solvers.60

Maloney and Siegler gave novice undergraduate physics students a set of problems in which the students were asked to compare five objects and determine which objects had the largest and smallest values of either momentum or kinetic energy. Some of the questions were phrased in everyday language and some were phrased in explicit physics language asking about momentum and kinetic energy. The object of the study was to observe what strategies the students used to solve the problems and if the students had multiple strategies. Although Maloney and Siegler did observe students using multiple strategies, in both types of problems the strategies emphasized mathematical formulas, not physics concepts. They also found that while students tended to use momentum on most of the problems in everyday language, students tended to use a different strategy if the question asked about kinetic energy explicitly. This result is one of many that indicate that students approach problems differently depending on cues in the problem.

3. Cues, triggers, and models used by students

In 1984, Anzai and Yokoyama presented results from a series of three studies of novice student representations or models of three physics problems and how readily they could be persuaded to change their models.61 They defined 3 types of student models:
experiential models, false scientific models, and correct scientific models. Experiential models are models that are derived from the students’ experiences that do not involve scientific concepts or terms. False scientific models and correct scientific models both explicitly involve scientific concepts, terms, or relations but the false scientific model incorrectly characterizes the problem information.

The three physics problems they used are described below:

1. Yo-yo problem – what will happen if you pull the string of a wound yo-yo sitting on a horizontal surface, assuming the disks may roll but never slide?
2. Pulley problem – If two different mass blocks are hanging by a string on opposite sides of a pulley that is suspended from a spring balance, what does the balance read? Assume the pulley is massless.
3. Balance problem – For a two pan balance with a 1-kg mass on one side and a container (assumed to be massless) with 1 kg of water on the other, what happens to the balance when a 1 kg ball suspended by a string is completely immersed in the water?

In addition to these three problems, they also constructed augmented versions that gave physics cues on how to solve the problems. Anzai and Yokoyama used these problem in a series of studies using these problems with experts and novices including 216 university freshman enrolled in a college physics course.

The results of Anzai and Yokoyama’s studies can be summarized as follows:

- Experts would generate one model that they would use to solve the problem. The novices would generate several models (different from the expert models) and compare them before deciding which one to use to solve a problem. The novice models were primarily either experiential or scientifically false.
- The novices generated experiential models for the basic pulley problem and the basic yo-yo problem, but they generated false scientific models for the basic balance problem. With additional cues the novices were able to generate correct scientific models for the augmented yo-yo and pulley problems, but not for the balance problem. This result suggests that false scientific models are more insensitive to physical cues. Further study with the balance problem showed that while most students (86%) were able to understand the buoyant force on the ball, many of the students (44%) did not recognize the reaction force on the container of water.
Certain physical cues were more helpful than others in getting the students use the correct scientific model with the yo-yo and the pulley problem. In one case with the yo-yo problem, two different physically relevant cues were necessary. In another case with the pulley problem, tension cues improved student performance significantly but acceleration cues did not. These results suggest that novices have fragmented, incomplete knowledge and that the model they use to solve a problem strongly depends on the cues given in the problem.

In a related study of knowledge and information processing involved in solving mechanics problems, Hegarty argued that novice problem solvers have two distinct sets of knowledge from which they construct problem representations. The two types of knowledge are intuitive knowledge based on what people learn from observing the world around them and theoretical knowledge that is acquired from formal instruction. Hegarty claims that for novice problem solvers these two types of knowledge are not integrated and so novices must choose one or the other. In contrast, experts develop representations from a knowledge base where everyday and formal knowledge are integrated.

Hegarty then argues that novices’ conceptual knowledge changes in three ways as they learn mechanics:

1. The level of specificity of concepts broadens and is tied to underlying physical principles rather than to surface features.
2. Concepts change from qualitative to quantitative.
3. Concepts are applied more consistently to appropriate situations.

The problem solving studies discussed above indicate that students’ conceptual understanding and their conceptual models play a major role in how they look at and how they solve physics’ problems. In order to evaluate what students are learning, it is important to consider what we know about students’ understanding of physics concepts. It is also interesting to note that in some instances students separate what they know
from their own experiences, intuitive knowledge, and what they learn from formal instruction, theoretical knowledge. We will refer to the link between these two types of knowledge as a reality link, the link between what a student learns in classes and the student’s real world.

III. CONCEPTUAL UNDERSTANDING

Around the same time researchers began studying how people solve physics problems, physics education researchers such as McDermott, Clement, Viennot, Hestenes & Halloun and others began studying college students’ conceptual understanding of physics in the area of mechanics. Using a combination of demonstration interviews (see chapter 7) and specially designed problems (see chapter 6) they began studying students’ understanding of selected concepts in mechanics like velocity, acceleration, and force in specific contexts. Here again, they found that the key point is that there are significant differences between experts and novices. Like problem solving, the main difficulty in helping novices become experts is teaching them in ways that take into account their existing prior knowledge. This section will summarize what is known about the initial state of students’ conceptual understanding and how to help them learn a good functional understanding of physics concepts. The focus will be on the best understood and longest studied areas of students’ conceptual understanding, kinematics and Newton’s laws of motion.

A. Kinematics and Newton’s Laws

In the early 1980s, McDermott and other physics education researchers found that each student does not come into a physics course as a blank slate. The students
bring with them their own system of common sense beliefs and intuitions about how the
world works derived from extensive personal experience. Furthermore, they found that
many students’ common sense beliefs are very stable, incompatible with the physics
taught in the introductory course, and appear to outlast instruction. These research
results led Hestenes et al. to claim that “instruction that does not take the students initial
state of conceptual understanding into account is almost totally ineffective, at least for
the majority of the students.”69 Traditional instruction does little to change students’
common-sense beliefs causing some of them to misinterpret the course material.

However, many students have the same types of common-sense beliefs. This makes it
possible to design curriculum that can take the more prevalent common-sense beliefs into
account and help students learn concepts more effectively. There are three frequently
found types of student difficulties with common-sense beliefs of motion and force:
language, pre-Newtonian views of motion and force, and representations.

1. The language of physics

The language difficulty occurs mainly because many words used in the basic
description of mechanics are also used in common everyday language. The way students
use words like force, momentum, energy, acceleration, speed, and velocity in common
speech can cause difficulties in the context of the physics class. Many of these words
have different meanings and connotations in common speech and many are used
interchangeably without regard to the meaning they have for physicists. As a result,
students often use the language of physics either without understanding the meaning of
the words in the physics context70 and/or without differentiating between words for
related concepts. For example, many students have trouble differentiating between
distance, velocity and acceleration and equate them all with a generalized idea of “motion.”

2. Common sense beliefs on force and motion

In studying students’ common sense beliefs and building on the earlier studies, Halloun and Hestenes found from interviews and diagnostic concept tests (see chapter 4) that the majority of responses from 478 students on force and motion questions at the beginning of a University Physics class were consistent with either pre-Newtonian (83% on the diagnostic) or Newtonian (17%) models of motion.\(^2\) Moreover, nearly every student used a mixture of models of motion and appeared to apply the models inconsistently in different contexts. Interviews were conducted with 22 of the students to probe their common sense beliefs more deeply and verify the diagnostic results.

Halloun and Hestenes caution that the common-sense alternatives to Newtonian views are not just misconceptions. Many of these common-sense beliefs were held by some of the greatest intellectuals in the past including Galileo and even Newton.\(^3\) Many are reasonable hypotheses grounded in everyday experiences. For example, the common-sense belief that something must cause the motion of an object is due to the observation that some force must be applied to most objects like cars, trains, and boxes to keep them in motion at constant speed because of frictional effects.

The fact that the students apply their common-sense beliefs inconsistently implies that their knowledge structure coming into the class is fragmented, incoherent, and context dependent. Studies by Minstrell\(^4\) and diSessa\(^5\) indicate that the students’ common-sense beliefs can indeed be characterized as loosely organized, ill-defined bits and pieces of knowledge that are dependent upon the specific circumstance in question.\(^6\)
(This context dependence is related to the issue of cues and triggers discussed in the problem-solving section of this chapter.) diSessa has found that many student common sense beliefs are rooted in pieces of knowledge he calls “psychological primitives” or “p-prims.” P-prims are simple, general isolated pieces of mental models that are cued by particular situations and are used to explain the events of the situation. They are usually strongly tied to real world experiences. The common sense example discussed above is an expression of the “force as mover” p-prim and the “continuous force” p-prim. The “force as mover” p-prim is the belief that objects go in the direction they are pushed. This p-prim tends to be used when an object is given a short, instantaneous push. The “continuous force” p-prim is the belief that a force is required to keep an object moving. This p-prim tends to be used in situations where the force is continuous rather than instantaneous.

Studies using pre and post testing and/or interviews have shown that many students still have these common-sense beliefs after traditional instruction. In their interview study on students’ common sense beliefs, Halloun and Hestenes found that as a rule,

students held firm to mistaken beliefs even when confronted with phenomena that contradicted those beliefs. When a contradiction was recognized or pointed out, they tended at first not to question their own beliefs, but to argue that the observed instance was governed by some other law or principle and the principle they were using applied to a slightly different state. … Careful interviews of students who have just witnessed a demonstration are enough to make one dubious about the effectiveness of typical classroom physics demonstrations in altering mistaken physics beliefs. We doubt that a demonstration can be effective unless it is performed in a context that elicits and helps to resolve conflicts between common sense and specific scientific concepts.
This last finding of the ineffectiveness of typical demonstrations to resolve conflicts between common-sense beliefs and specific physics concepts has also been reported in studies by Kraus et al.\textsuperscript{78} as well as Redish et. al.\textsuperscript{79}

Recent studies by Francis and Larson\textsuperscript{80} at Montana State University have shown that even after students appear to have acquired a Newtonian view of linear force and motion, rotational analogs to the students’ common sense beliefs from linear motion appear when the students begin discussing rotational motion and dynamics. This suggests that students actually hold onto both the Newtonian concepts and the common sense beliefs and that their response will depend on what is triggered by the cues of the situation. It further suggests that in situations outside the context where they learn specific physics concepts, students have a tendency to revert to their common-sense beliefs.

3. Representations

In addition to the above mentioned difficulties with language and concepts, students also have difficulties with the abstract representations of physics such as graphs, equations, free-body diagrams, and vectors. One of the main goals of the hidden curriculum is to help students become fluent with the multiple representations of physics, a necessity for students trying to develop a robust, functional knowledge of physics. There is a lot of research on this issue by both math and physics education researchers. In this dissertation and this section, we will only consider student difficulties with understanding graphs since graphs are one of the most useful and powerful representations of ideas and data, both in class and in everyday life.
Several studies have found that students coming into introductory physics classes understand the basic construction of graphs but have difficulty applying their understanding to the tasks they encounter in physics. The two most common types of student errors are thinking of a graph as a literal picture of an object’s motion and confusing the meaning of the slope of the line with the height of the line. An example of the former is when a student asked to draw a velocity vs. time graph of a bicycle going along on a hilly road draws a velocity graph that resembles the hills and valleys traversed by the bicycle. An example of the latter is when students asked to find the point of maximum rate of change indicate the point of largest value. In general, students tend to find interpreting slopes more difficult than individual data points. They also have difficulty separating the meanings of position, velocity, and acceleration graphs.

Beichner at North Carolina State University developed a multiple-choice diagnostic test for studying the prevalence of student difficulties with the understanding of kinematics graphs. In a study of 900 students at both the high school and college level, there was no significant difference in the overall score between the high school and the college students. However, the calculus-based students did score significantly better than the algebra/trig students (with a mean of 9.8 vs. 7.4 out of a maximum of 21). In addition, he uncovered a consistent set of student difficulties with graphs of position, velocity, and acceleration. One of the main difficulties was that approximately 25% of the students believed that switching variables would not change the appearance of the graph. This is an indication that the students are interpreting the graphs as pictures. It is interesting to note that the students who could correctly translate from one variable to another had the best scores on this diagnostic test. Another result was that 73% of the
students correctly identified the slope of a line passing through the origin while only 25% were able to do so for a line that did not go through the origin. Approximately 25% of the students gave responses that are consistent with the slope/height mix-up described above. He also found that the students had trouble using the area under the curve to go from one graph to another. About one third of the students gave responses that used the slope of the line when they needed to find the area, and less then a third of the students were correctly able to determine the change in velocity from an acceleration graph.

Although Beichner did not use interviews to see why students answered the way they did, a similar study by Thornton of 10 multiple-choice velocity and acceleration graph questions did.85 Thornton found student error rates after traditional instruction of about 40% on the velocity questions and 70-95% error rates on the acceleration questions. The instructors of these classes felt that these questions were simple (they expected error rates of 5-10%) and that students who were unable to answer these questions correctly had a very poor understanding of kinematics. Thornton noted that the problem was not that the students were unable to read graphs. Almost all (95%) of the students could answer similar questions about distance graphs correctly and interviews with the students showed that the students were picking graphs consistent with their verbal or written explanations of velocity and acceleration. This implies that students’ difficulties with kinematics graphs are directly related to their conceptual difficulties with acceleration and velocity.

Research into students’ conceptual understanding has extended to many areas besides mechanics and continues to be an active area of PER. Some of the other areas include:
In the next section I will discuss research on helping students to improve their conceptual understanding.

**B. Mechanisms for Changing Students’ Beliefs**

In order to see how student beliefs can be changed, I need to discuss two mechanisms for learning introduced by Piaget, “assimilation” and “accommodation.”

Piaget defines assimilation as the process by which people learn new ideas that match or extend on their existing conceptual knowledge. He defines accommodation as the process where people learn new ideas that don’t fit into their existing conceptual knowledge either because the ideas are completely new or because the idea conflicts with what they already know. Students find it much easier to learn physics concepts that fit their view of how things work, i.e. to assimilate new ideas. Accommodation tends to be much harder because the students must change or rethink their existing views. One reason for this is that often students will perceive and interpret what they learn in a way that makes sense in terms of their existing views. For example, in the case of the demonstration studies described previously, many students tried to assimilate rather than accommodate what they observed by interpreting it either as an example that demonstrates what they believe or as a special case that is unrelated. This tendency to assimilate rather than accommodate is one reason that students’ conceptual knowledge may contain contradictory elements.
However, even though helping students learn to change their conceptual understanding is difficult, it is not impossible. Research has shown that it is possible to stimulate conceptual change for most of the students in a class. In his review of cognitive science research relevant to physics instruction, Redish notes that the mechanism for conceptual change appears to critically involve prediction and observation. “The prediction must be made by the individual and the observation must be a clear and compelling contradiction to the existing [conceptual knowledge].”

He notes Posner et al.’s suggestion that in order to change students’ existing conceptual understanding the proposed replacement must have the following characteristics:

- The replacement must be understandable.
- The replacement must be plausible.
- There must be a strong conflict with predictions based on the subject’s existing conceptual understanding.
- The replacement concept must be seen as useful.

Redish adds, “The clearer the prediction and the stronger the conflict, the better the effect.” However, since students can hold conflicting ideas in their models at the same time, it is crucial that the students be made to resolve the conflict for conceptual change to be effective.

Two of the three research-based curricula being studied in this dissertation, Tutorials developed by the McDermott PER group at University of Washington and Workshop Physics developed by Priscilla Laws at Dickinson College, were specifically designed to improve students’ conceptual understanding of physics based on the research discussed in this section. (Both curricula are described in detail in chapter 8.)
curricula use strategies that resemble Posner et al.’s four conditions of conceptual conflict and change.

In the Tutorial curricula, traditional recitations are replaced by groups of three and four students working through a specially designed worksheet by consensus. The tutorials use a strategy of elicit, confront, and resolve. The worksheet problems put the students into situations where they tend to use their common sense beliefs as part of their reasoning, then put them in a situation where they are forced to recognize an inconsistency between their reasoning and something they believe is true, and finally helps them to reconcile the inconsistency.

In the Workshop Physics curricula, the traditional lecture/lab/recitation format is abandoned in favor of an all-laboratory approach. Each laboratory unit is designed to take the collaborative student groups through a four-part process that can be summarized as follows:

1. Students make predictions about a physical situation.
2. The students perform discovery-oriented experiments and try to reconcile the differences between their experimental observations and their predictions.
3. The students develop definitions and equations from theoretical considerations.
4. Finally, they perform experiments to verify their theoretical model and apply their understanding of the phenomenon to the solution of problems.

While these two curricula seem very different, there are some common factors between them. Both of them build on students’ existing common sense beliefs to help them develop a physicists’ understanding of the concepts. They are also constructed on the principle that the students must be “engaged” in the learning process by thinking about situations, committing to a prediction of what will happen, discussing that
prediction with their fellow students, and resolving that prediction with what really happens.

IV. EXPECTATIONS AND EPISTEMOLOGY

Students bring more than just naive or novice views of problem solving and physics concepts to the introductory physics classroom. Each student, based on his or her own experiences, also brings to the class a set of attitudes, beliefs, and assumptions about what sorts of things they will learn, what skills will be required, and what they will be expected to do. In addition, their view of the nature of scientific information affects how they interpret what they hear. In this dissertation, I will use the phrase “expectations” to cover this rich set of understandings. In this dissertation, I consider what we might call students' cognitive expectations — expectations about their understanding of the process of learning physics and the structure of physics knowledge rather than about the content of physics itself. This is in contrast to more affective expectations such as students’ motivation, preferences, and feelings about science and/or scientists, etc. While these are also important, they are not part of this study and have been probed extensively elsewhere.98

As we saw in the studies of Hammer and Tobias, students’ expectations about what they were learning and what they needed to do to succeed in the class seemed to affect their views on building conceptual understanding and problem solving as well as what they took away from the class. In particular, some of the students’ expectations prevented the students from building a robust understanding of physics. Furthermore, it is reasonable to assume that one reason why many of Mazur's students learned to do the
traditional quantitative without learning the underlying concepts is related to these students’ expectations of what they were supposed to learn.

The role of these types of epistemological beliefs on adult learners in introductory undergraduate physics courses is not well understood. The few studies like Hammer’s that exist indicate the effects may be profound and very much related to students’ conceptual knowledge and problem solving skills. Furthermore, these studies indicate that like problem solving and conceptual understanding, introductory students’ epistemological beliefs often differ from those of experts and may hinder their learning of physics. This section will summarize the nature and results of four of the more pertinent studies.

A. **Previous Research on Cognitive Expectations in the Pre-College Classroom**

There are a number of studies of student expectations in science in the pre-college classroom that show that student attitudes towards their classroom activities and their beliefs about the nature of science and knowledge affect their learning. Studies by Carey, Linn, and others have demonstrated that many pre-college students have misconceptions both about science and about what they should be doing in a science class. Other studies at the pre-college level indicate some of the critical items that make up the relevant elements of a student's system of expectations and beliefs. For example, Songer and Linn studied students in middle schools and found that students could be categorized as having beliefs about science that were either dynamic (science is understandable, interpretive, and integrated) or static (science knowledge is memorization-intensive, fixed, and not relevant to their everyday lives). In a review of
student expectation studies in high school physics, Gunstone concludes: “The ideas and beliefs of learners about learning, teaching, and the nature of appropriate roles for learners and teachers are major influences is the likelihood of learners choosing to undertake the demanding and risky processes of personal and conceptual change.”

Expectation studies of high school mathematics classes by Schoenfeld and the third National Assessment of Educational Progress have found that students’ beliefs about mathematics and mathematical problem solving are shaped by their experiences in mathematics classrooms. Using national math assessments, surveys, interviews, and classroom observations, these two studies found that the majority of junior high school and high school students have the following beliefs about the nature of mathematics:

- Mathematics is mostly memorization
- Mathematics problems have one and only one correct solution and answer; the correct solution usually uses the rule the teacher has most recently demonstrated to the class.
- Students who have understood the mathematics they have studied will be able to solve any assigned math problem in five minutes or less. (Note: students with this belief will give up on a problem after a few minutes of unsuccessful attempts, even though they might have solved it had they persevered.)

Schoenfeld’s studies had two additional findings of typical student beliefs about mathematics that echo the expectations of Liza in Hammer’s study discussed earlier in this chapter:

- Ordinary students cannot expect to understand mathematics; they expect simply to memorize it and apply what they have learned mechanically without understanding.
- The mathematics learned in school has little or nothing to do with the real world.
In his classroom observations and interviews, Schoenfeld studied how student expectations affected their behavior in class and in problem solving. He concluded that, "Student's beliefs shape their behavior in ways that have extraordinarily powerful (and often negative) consequences."

These results led Edward Silver to use the phrase “hidden curriculum” to describe this unintentional by-product of formal mathematics education. In his words, since the students’ viewpoint represented by these statements is clearly inadequate, and potentially harmful to their future progress in mathematics, we need to focus our attention more clearly on those hidden products of the mathematics curriculum. Our students may realize greater educational benefits from our attention to the hidden curriculum of beliefs and attitudes about mathematics than from any improvement we could make in the “transparent” curriculum of mathematics facts, procedures, and concepts.

B. Studies of Young Adults’ Attitudes Towards Knowledge and Learning

Two important large scale studies that concern the general cognitive expectations of adult learners are those of Perry and Belenky et al. Perry tracked the attitudes of about 100 Harvard and Radcliffe students on epistemology, morals, and general world outlook throughout their college career. The students filled out an attitudinal survey at the beginning of their college careers and were interviewed at least once or twice a year during their four years at college. Extending on Perry’s study, Belenky et al. conducted in-depth interviews on similar issues, but with women from various walks of life. They tracked the views of 135 women in a variety of social and economic circumstances including 90 who were enrolled in one of six academic institutions varying from an intercity community college to a prestigious four-year women’s college. Twenty-five of the women in an academic setting were interviewed a second time one to five years later.
Both studies found evolution in the expectations of their subjects, especially in their attitudes about knowledge.\textsuperscript{111} Both studies frequently found their young adult subjects starting in a "binary" or "received knowledge" stage in which they expected everything to be true or false, good or evil, etc., and in which they expected to learn "the truth" from authorities. Both studies observed their subjects moving through a "relativist" or "subjective" stage (nothing is true or good, every view has equal value) to a "consciously constructivist" stage. In this last, most sophisticated stage, the subjects accepted that nothing can be perfectly known, and accepted their own personal role in deciding what views were most likely to be productive and useful for them.

The two studies also had two similar findings regarding the progression of their subjects through these stages. One, that the progression through the stages was not always linear and some of the subjects stayed in the binary or relativist stages. Two, the subjects did not usually progress from one stage to next without some type of cognitive conflict where their previous epistemology was inadequate for their current situation. This is very similar to Posner et al.’s conditions of conceptual change discussed earlier in this chapter.

Although the Perry and Belenky et al. studies both focused on areas other than science,\textsuperscript{112} most professional scientists who teach at both the undergraduate and graduate levels will recognize a binary stage, in which students just want to be told the "right" answers, and a constructivist stage in which the student takes charge of building his or her own understanding. Consciously constructivist students carry out their own evaluation of an approach, equation, or result, and understand both the conditions of validity and the relation to fundamental physical principles. Students who want to
become creative scientists will have to move from the binary to the constructivist stage. This is the transition that we will explore in chapter 10.

Another finding of the studies by Perry and Belenky et al. was that these expectation stages were not mutually exclusive and were often domain specific. An excellent introduction to the cognitive issues involved is given by Reif and Larkin\textsuperscript{113} who compare the spontaneous cognitive activities that occur naturally in everyday life with those required for learning science. They pinpoint differences and show how application of everyday cognitive expectations in a science class causes difficulties. Another excellent introduction to the cognitive literature on the difference between everyday and in-school cognitive expectations is the paper by Brown, Collins, and Duguid, who stress the artificiality of much typical school activity and discuss the value of cognitive apprenticeships.\textsuperscript{114}

C. Students Expectations in Undergraduate Physics

The most relevant and valuable study for the expectation questions in this dissertation is Hammer's Ph.D. thesis at Berkeley.\textsuperscript{115} In his dissertation (which followed the preliminary study described earlier in this chapter), Hammer interviewed six students several times each from the first semester of the calculus-based sequence for engineering students at the University of California at Berkeley. Four of the students volunteered and two more agreed to participate after specifically being selected in order to add "good" students to the group. All six students had taken physics in high school and scored at least 700 on the math SAT. Each student was interviewed for approximately ten hours during the same semester. In these interviews, Hammer used three types of
activities to probe student expectations: open and semi-directed discussions, problem solving, and direct questioning. For the problem solving activities, he asked the students to solve the problems out loud. The problems included three specifically chosen to address student common-sense beliefs. The interviews were taped and transcribed, and students were classified according to their statements and how they approached the problems.

Hammer proposed three dimensions along which to classify student expectations of physics knowledge: beliefs about learning, beliefs about content, and beliefs about structure. The three dimensions are described in Table 2-3 below.

In their beliefs about how physics is learned, the type A students feel they need to make sense of the course material for themselves, while the type B students feel they have to trust what they learned from the authorities, i.e. the teacher and the text, and simply believe what they were given. This dimension is based largely on the binary or received knowledge state of Perry and Belenky et al. However, Hammer notes that while a student in the relativist or subjective state might feel that all views are equally valid, they may recognize that using the views of the instructor and the textbook (the classroom authorities) on tests and exams is expedient for getting a good grade in the course. The quote from Mazur’s student about whether she should respond to the ungraded FCI according to what she was taught or by the way she really thinks is an example of this.

For their beliefs about the content of physics knowledge, the type A students feel that physics is coherent and fits together in a way they expect to understand and use. In contrast, type B students feel that knowing physics means remembering facts, and/or
they cannot solve a problem without knowing the “right” equation, i.e. they treat physics
knowledge as a set of independent and unrelated pieces of information. Some other
indicators of this dimension are that type A students believe that understanding

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Type A</th>
<th>Type B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Beliefs about learning</td>
<td>independent self-motivated, questions material until it makes sense to them</td>
<td>by authority takes what is given by instructor or text without evaluation</td>
</tr>
<tr>
<td>2: Beliefs about content</td>
<td>coherent believes physics can be considered as a connected, consistent framework</td>
<td>pieces believes physics can be treated as separate facts or “pieces”</td>
</tr>
<tr>
<td>3: Beliefs about structure</td>
<td>concepts focuses on conceptual understanding and connects formulas to intuition and underlying concepts</td>
<td>formulas focuses on memorizing formulas and using them in problem solving</td>
</tr>
</tbody>
</table>

formula derivations is important and check for consistency in their work while type B
students do not think derivations are important and are not bothered by apparent
inconsistencies.

For their beliefs about the structure of physics knowledge, the type A students
feel that the underlying concepts are what is essential and that the formulas are
expressions of those concepts, while the type B students feel that the formulas are
essential (at least for them), and the underlying concepts are not something that they
need to bother with. Note that these three dimensions need not be independent.

Hammer classified students by the number of times something they said or did in
the interviews indicated they held a particular type of expectation. He found that the
students’ expectations were generally consistent across all the activities in the interview.
Of the two students from the earlier study, Liza was classified as Type B while Ellen was type A. In this study, the two “good” students were both type A and other four fell into category B for all the variables. The type A students earned an A and an A+ in the course, while the four type B students grades ranged from A- to C+.

In addition, all four of the type B students displayed misconceptions on at least one of three selected problems in interviews. In general, they displayed more misconceptions than the type A students and were usually unable to resolve them when challenged by Hammer. The two type A students displayed no misconceptions on the selected problems and in general were able to resolve their own misconceptions when challenged. It is not surprising that the two good students received a higher grade and showed fewer misconceptions, since they were selected based on their high grades on the first midterm. It is surprising that they are characterized by expectations different from the subjects that did less well in the class and who were unable to resolve their misconceptions.

It is interesting to note the one of the type A students, code named Ken, is an interesting counter example to many instructors’ suppositions about successful physics students. Ken had a weaker math and science background than two of the type B students. Also, he was more concerned with getting a good grade in the class then in understanding physics but he felt conceptual understanding was essential to getting a good grade. Last, he did not value derivations except as support for his conceptual understanding. His main advantage over the two students with stronger backgrounds seems to have been his expectations, particularly his need to build a good conceptual understanding of physics.
Hammer also made the following three general observations about the involvement of expectations in the subjects learning.

1. [Type B subjects] were quite casual about making and breaking associations between different aspects of their knowledge; [type A subjects] were much more careful about building and modifying their understanding.

2. [Type B subjects] were quick to decide that they understood new information, while [the type A subjects] were more reflective and questioning.

3. [Type B subjects] were reluctant to spend time working on problems they did not know how to solve, while [the type A subjects] seemed to consider these the most interesting [problems]. In part, this appeared to reflect different goals: [the type A subjects] appeared to have a goal of understanding the material, while [the type B subjects] did not always consider understanding important.

Note that the first observation would lead to significant differences in the two types of students’ knowledge structures. The second observation is remarkably similar to the result of the studies by Chi et al.\textsuperscript{116} and Fergusson-Hessler and de Jong\textsuperscript{117} (discussed earlier in the problem solving section) that found that good students process readings and example problems more deeply than poorer students who just tend to read and accept. The third observation implies that the students’ personal course goals may be hindering them from reaching our goal of robust functional understanding.

In summary, Hammer found that his six students had expectations about physics knowledge and learning that could be classified based on indications over many hours of interviews. He found that these indications were consistent across a wide range of activities. His results indicate that expectations seem to be involved in whether the subjects were able to solve his interview problems and in whether they developed a coherent, conceptual understanding of the course material. These findings support expectations as a valid perspective for evaluating student learning and reasoning. Due to
the limited scope of his study, it is impossible to tell the extent to which the three 
expectation variables would be correlated or what the distribution of the categories 
would be in a typical class.

1 R. Thornton and D. Sokoloff, “Learning motion concepts using real time 
microcomputer-base laboratory tools,” Am. J. Phys. 58 (9), 858-867 (1990); L.C. 
McDermott, “Bridging the gap between teaching and learning: The role of research,” 
in AIP Conference Proceeding No. 399 The Changing Role of Physics Departments 
in Modern Universities: Proceedings of the International Conference on 
Undergraduate Physics Education, edited by E.F. Redish and J.S. Rigden (AIP Press, 
Woodbury NY, 1997), 139-166; R.R. Hake, “Active-engagement vs. traditional 
methods: A six thousand student study of mechanics test data for introductory 


3 A. van Heuvelen, “Learning to think like a physicist: A review of research based 


5 D. Hammer, “Two approaches to learning physics,” Phys. Teach. 27 (9) 664-670 
(1989).

6 S. Tobias, They’re Not Dumb, They’re Different: Stalking the Second Tier (Research 


9 See Ref. 4.

10 Mazur used the mechanics diagnostic test from Ref. 7.

11 In section I, there are quotes from papers and researchers as well as quotes from 
interview transcripts. The former are written in plain text while the latter are written 
in italics. Brackets [ ] are used to denote words added by this author to clarify a 
quote. Brackets in italics / / are used to denote words added to clarify a student 
quote added by the original author.
Liza and Ellen went to high schools in neighborhoods with similar socioeconomic backgrounds.

See Ref. 5.


See Ref. 6.


Both Eric and Jackie had not used calculus for some time but found that it came back to them quickly.


28 See Ref. 23.

29 See Ref. 24.


32 See Ref. 23.

33 See Ref. 26.

34 See Ref. 25.

35 See Ref. 25.

36 See Ref. 25.


38 See Ref. 24.


40 See Ref. 24.

41 See Ref. 39.
See Ref. 24.


See Ref. 39.


See Ref. 22.


See Ref. 50.


See Ref. 50.

Figures reproduced from Ref. 57.

See Ref. 57.


See Ref. 7.

See Ref. 8.


See Ref. 2.


See Ref. 8.

See Ref. 8.


See Refs. 1, 2, & 3.


See Ref. 83.


See Refs. 1 & 2.

See Ref. 92.


A classic example of this is the example from the third National Assessment of Educational Progress where of the 70 % of the students who performed the calculation correctly for a problem inquiring about the number of buses for a task roughly two thirds of them wrote “31 remainder 12” or “31”. Less than one third wrote down the correct answer of “32”. See Ref. 105.


W.F. Perry, *Forms of Intellectual and Ethical Development in the College Years* (Holt, Rinehart, and Wilson, NY, 1970).


This brief summary is an oversimplification of a complex and sophisticated set of stages proposed in each study.

Perry specifically excludes science as “the place where they do have answers.” See Ref. 109.


116 See Ref. 23.

117 See Ref. 25.
PART II. RESEARCH METHODS AND ASSESSMENT: HOW DO WE DETERMINE WHAT STUDENTS ARE LEARNING?

Chapter 3. Overview of Methods in Physics Education Research

Part II of this dissertation is intended to help acquaint the reader with the methods and limitations of PER, in particular, the methods used in this investigation. This chapter provides a general overview of the models and methods of PER. The remaining chapters of Part II discuss in detail the specific research methods used in this dissertation.

I. PHYSICS EDUCATION RESEARCH

PER involves research into the teaching and learning of physics, curriculum and software development, and innovative instruction. Lillian C. McDermott of the University of Washington Physics Education Group has broken the process in which her group uses research to guide curriculum development into three parts:

1.) conducting systematic investigations of student understanding
2.) applying the results to the development of specific instructional strategies to address specific difficulties, and
3.) designing, testing, modifying, and revising the materials in a continuous cycle on the basis of classroom experience and systematic investigations with the target population.

This process is represented schematically as a cyclic process in Figure 3-1. This cycle represents an iterative approach where research leads to changes to curriculum which is then implemented in the classroom. Classroom instruction is then evaluated through research which leads to more changes in the curriculum. McDermott’s research group went through several iterations of this cycle in the development of the materials for both
the tutorial curriculum\textsuperscript{3} to supplement calculus-based introductory physics lecture (discussed in chapter 8) and the *Physics by Inquiry* curriculum\textsuperscript{4} to teach K-12 teacher science through discovery.

At University of Maryland, we have added an axle to McDermott’s wheel to represent the model of how students think and learn. This highlights the importance of the model in all aspects of this cycle. The model both guides and is informed by the research and development cycle. Since, as we discussed in chapter 2, our model of student learning focuses on changes in the student, the basic problem in PER is then the transformation of a system $S$, i.e. the student, from an initial state $S_i$ to a desired final state $S_f$ where the student can do things they could not do before.\textsuperscript{5} This suggests that student learning should be studied in a way analogous to the way physicists study physical processes; namely, measurements need to be made to determine the students’ initial and final states to understand the transformation to help build and extend the
model of student leaning. Intermediate measurements, if possible, are greatly desired to improve our understanding of the transformation and the model.

II. PHYSICS EDUCATION RESEARCH METHODS

In the past twenty years, physics education researchers have used a variety of techniques to evaluate what students know and what they are learning. These methods include the following:

1. Observing students in class and office hours
2. Measuring student and faculty habits, attitudes, and beliefs with a survey or questionnaire.
3. Measuring student learning using a multiple-choice test designed using the results of physics education research on commonly found errors to specify attractive distractors.
4. Measuring student learning using long-answer exam questions -- problems or open-expression questions in which students explain and discuss their answers.
5. Measuring student learning through recorded problem interviews.

Most instructors use observations of students to some degree. One of the key findings of physics education research is that one wants to understand student difficulties. It is important to not just listen to what the students are saying about the course material, but to draw out the students and see what they really think. Classroom observations can be very helpful in this regard. However, it is difficult to make substantial observations on more than a small fraction of a class and the observations are dependent on available opportunities.

While surveys and questionnaires are the simplest and most commonly used method of evaluation by instructors in the form of typical course evaluations, it is important to distinguish between instruments that measure aspects of course satisfaction
and instruments that deal with issues more directly related to student learning. Although both student and faculty satisfaction is important in motivating student work and presumably therefore student success, the link between satisfaction and learning is highly indirect. Indeed, students whose primary goal is a good grade may find higher satisfaction in a course that produces a good grade without improved learning, since improved learning often requires time and painful effort. However, several physics education researchers including the author have developed and used surveys to learn more about student habits, attitudes, and beliefs that have a more direct effect on how and what students learn. The survey developed for this dissertation is discussed in chapter 5.

Multiple-choice tests are easy to deliver, but building a useful and reliable instrument requires substantial effort. While the results can be highly suggestive, multiple choice tests can be difficult to interpret. They have a tendency to overestimate the student's learning since they can sometimes be answered correctly by means of incorrect reasoning or by "triggered" responses that fail to represent functional understanding. On the other hand, the use of common misconceptions as distractors produces "attractive nuisances" that challenges the students' understanding. Students that get the correct answer despite this challenge are likely to have a good understanding of the topic in question. We expect therefore that this method does give some indication of the robustness of a student's possession of and confidence in correct knowledge.

Long-answer problems are easy to deliver as part of a course quizzes and exams, but the analysis can be time consuming. Student answers must be read in detail and classified by the understanding displayed. Unlike instructors whose goal may be to
certify student success and therefore focus on the student’s correct answers, researchers need to pay greater attention to student errors. The errors are often informative about how the students are thinking about physics. The functionality of student knowledge is rather well-tested by this approach since the student is being asked to produce the desired knowledge within the context of a problem and without the most common and automatic triggers. It has the defect that students occasionally give answers too incomplete or ambiguous to let instructors or researchers see what they are thinking.

The interview method is the most effective approach since it permits the researcher to observe in detail the functionality of the student's knowledge by the presentation of a variety of contexts. The researcher can follow up suggestive responses with more detailed and individually designed questions, but it is highly time consuming. In addition to the recording time (usually one or more hours per student), the recordings must be transcribed and analyzed. This approach is thus impractical for evaluating the distribution of student knowledge throughout a large class. However, because many students share a relatively small number of difficulties, a small number of interviews can usually reveal most of the common student problems in great detail.

Other evaluation methods used by physics education researchers include student journals, grades, and retention within the introductory sequence. Grades and retention are difficult measures to use unless proper controls are used to account for variations in population, instructors, and time. Student journals are becoming an increasingly common tool for physics education researchers in smaller classes. They can be very revealing for seeing how the students view both the course material and the course if the students are given proper guidelines. However, while journal entries can be a useful
research tool, they can also be time consuming to read and analyze making them impractical for large classes.

III. OVERVIEW OF RESEARCH METHODS USED IN THIS STUDY

In order to evaluate the success of a particular research-based curriculum, we must decide what we mean by "success." This plays an important role in determining our approach to evaluation. What is meant by success is, in turn, determined by our model of student understanding and learning. As discussed in chapter 2, the critical element of our model for this application is to help students gain the knowledge and skills needed to improve their ability as problem solvers. In this study, I am evaluating the curricula in terms of student’s conceptual understanding and their expectations or cognitive attitudes and beliefs.

In terms of conceptual understanding, the student may "have" an item of knowledge, that is, be able to recall it in response to a narrow range of triggers, but be unable to recall and apply it in a wide range of appropriate circumstances. Since our goal is help students achieve a robust functional understanding, this is what we want our evaluations to test for. Our evaluation of students’ conceptual understanding with multiple-choice concept tests, open-ended conceptual problems, and interviews is presented in chapter 9.

As we saw in chapter 2, students’ expectations can have a strong influence on what students take away from an introductory course. Here, a successful curriculum would be one that supports or encourages student expectations that are favorable for building a robust understanding of physics. Based on the studies of Perry and Belenky et
and my own observations, student expectations can become more favorable over time. To measure this change, we need instruments that can recognize changes in expectations. While interviews are used to assess the expectations of a small sample of individual students, I have participated in the development of a survey instrument to determine the distribution of student expectation coming into the introductory class and to see how the distribution changes as the students progress through the sequence. The Maryland Physics Expectation (MPEX) survey has been developed by members of the Physics Education Research Group at the University of Maryland over the past five years. Our evaluation of student expectations with the MPEX survey and interviews is presented in chapter 10.

The instruments and methods used in this study are described in detail in the next four chapters. Chapter 4 will discuss the Force Concept Inventory (FCI) and the Force and Motion Concept Evaluation (FMCE). The development and validation of the MPEX survey is contained in chapter 5. A discussion of how exam and quiz problems can be used to evaluate instruction can be found in chapter 6. Our interview methods for studying students’ expectations and conceptual understanding are described in chapter 7.


Chapter 4. Multiple Choice Concept Tests: 
The Force Concept Inventory (FCI)

I. CHAPTER OVERVIEW

In the early 1980s, McDermott, Viennot, and other physics education researchers found that each student comes into a physics course not as a blank slate but brings into the classroom a system of common sense beliefs and intuitions about how the world works derived from extensive personal experience (this is discussed in more detail in chapter 2). Furthermore, they found that these common sense beliefs were very stable and often incompatible with the physics being taught in the introductory course. Traditional instruction does little to change these beliefs and they cause some students to misinterpret the course material. This result was pieced together by researchers who studied student understanding of isolated concepts in mechanics.

At the same time Hestenes and Halloun at Arizona State University began developing an instrument called the Mechanics Diagnostic Test (MDT) that measured not the student’s initial knowledge of Newtonian force but the discrepancy between the students’ common sense beliefs and their belief in the Newtonian force concept. In 1992, an improved version of the MDT was published as the Force Concept Inventory (FCI). Because the questions are written in plain language and easily understood by non-physics students, the FCI can be given at the beginning and as well as at the end of a course to see if the students improve. Because the instrument was included in the published study, which also described how to interpret the results, the FCI could be used by any physics instructor to evaluate their own students. Many physics faculty, including Eric Mazur as described in chapter 2, have overcome the initial “not my class” response
to reports of students’ difficulties with conceptual understanding after using the MDT and/or the FCI and seeing exactly how poorly their own students fare. The value of these two instruments has led to the development of other multiple-choice concept tests in mechanics and other content areas of the introductory course.\(^7\) One of the other mechanics tests is the Force Motion Concept Evaluation (FMCE),\(^8\) an instrument similar to the FCI that looks at a smaller set of concepts and makes heavy use of graphical and pictorial representations.

The FCI and the FMCE are the two most commonly used physics concept tests in use today. Every class involved in this study has used one of these two tests to monitor students’ improvements in conceptual understanding. Almost all evaluations of improved introductory mechanics classes reported at national AAPT meetings make use of one of these tests. Yet, recently questions have been raised concerning what tests like the FCI actually measure.\(^9,10\)

This chapter will review the development of these tests, discuss what they measure, and present some results concerning limitations of the FCI and how the two tests compare. The so-called 4-H controversy will also be discussed.\(^11,12,13,14\)

### II. DEVELOPMENT OF THE FCI

Halloun and Hestenes developed the MDT for the expressed purpose of evaluating introductory physics instruction objectively. They share Reif’s view\(^15\) that student learning should be viewed as a transformation from an initial state to a final state. Thus, the MDT was intended to be used as a pre-course test to assess students’ initial knowledge state and as a post-course test to measure the effect of instruction
independent of other assessments such as exams or homework. Because Newtonian mechanics is the central theme of the first course in most introductory physics sequences and it is an essential for the rest of the sequence, Halloun and Hestenes restricted their instrument to concepts related to Newtonian force.

Halloun and Hestenes developed the MDT in successive iterations over three years and administered it to over 1000 students in introductory college physics. The test questions were initially selected to assess students’ qualitative conceptions of motion and its causes. The questions were designed to identify common misconceptions or non-Newtonian common sense beliefs noted in the literature. The early versions required written responses. Later multiple-choice versions used the most common written student responses that were indicative of non-Newtonian common sense beliefs as distractors. A student’s overall score was taken as a measure of that student’s qualitative understanding of Newtonian force.

The FCI is similar in design to the MDT and produces similar overall scores when used in comparable classes. In fact, roughly half the questions from the MDT remain unchanged in the FCI. The main advance in the FCI from the MDT comes from a systematic analysis and explicit taxonomy of the basic concepts of Newtonian concepts and students’ common sense beliefs. The FCI is designed to cover these concepts more comprehensively and facilitate the interpretation of the results. For example, the FCI can identify students’ difficulties with each of Newton’s laws of motion and can help identify the common sense beliefs associated with each of these difficulties.¹⁶

<table>
<thead>
<tr>
<th>Concept</th>
<th>Inventory Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Kinematics</td>
<td></td>
</tr>
<tr>
<td>Velocity discriminated from position</td>
<td>20E</td>
</tr>
<tr>
<td>Acceleration discriminated from velocity</td>
<td>21D</td>
</tr>
<tr>
<td>Constant acceleration entails</td>
<td></td>
</tr>
<tr>
<td>parabolic orbit</td>
<td>23D;24E</td>
</tr>
<tr>
<td>changing speed</td>
<td>25B</td>
</tr>
<tr>
<td>Vector addition of velocities</td>
<td>(7E)</td>
</tr>
<tr>
<td>1. First Law</td>
<td></td>
</tr>
<tr>
<td>with no force</td>
<td>4B;(6B);10B</td>
</tr>
<tr>
<td>velocity direction constant</td>
<td>26B</td>
</tr>
<tr>
<td>speed constant</td>
<td>8A;27A</td>
</tr>
<tr>
<td>with canceling forces</td>
<td>18B;28C</td>
</tr>
<tr>
<td>2. Second Law</td>
<td></td>
</tr>
<tr>
<td>Impulsive force</td>
<td>(6B);(7E)</td>
</tr>
<tr>
<td>Constant force implies constant acceleration</td>
<td>24E;25B</td>
</tr>
<tr>
<td>3. Third Law</td>
<td></td>
</tr>
<tr>
<td>for impulsive forces</td>
<td>2E;11E</td>
</tr>
<tr>
<td>for continuous forces</td>
<td>13A;14A</td>
</tr>
<tr>
<td>4. Superposition Principle</td>
<td></td>
</tr>
<tr>
<td>Vector sum</td>
<td>19B</td>
</tr>
<tr>
<td>Canceling forces</td>
<td>(9D);18B;28C</td>
</tr>
<tr>
<td>5. Kinds of Force</td>
<td></td>
</tr>
<tr>
<td>5S. Solid contact</td>
<td></td>
</tr>
<tr>
<td>passive</td>
<td>(9D);(12 B,D)</td>
</tr>
<tr>
<td>Impulsive</td>
<td>15C</td>
</tr>
<tr>
<td>Friction opposes motion</td>
<td>29C</td>
</tr>
<tr>
<td>5F. Fluid contact</td>
<td></td>
</tr>
<tr>
<td>Air resistance</td>
<td>22D</td>
</tr>
<tr>
<td>buoyant (air pressure)</td>
<td>12D</td>
</tr>
<tr>
<td>5G. Gravitation</td>
<td></td>
</tr>
<tr>
<td>acceleration independent of weight</td>
<td>1C;3A</td>
</tr>
<tr>
<td>parabolic trajectory</td>
<td>16B;23D</td>
</tr>
</tbody>
</table>
Table 4-2. Taxonomy of misconceptions probed by the Force Concept Inventory (FCI). Belief in the misconceptions is suggested by selection of the corresponding FCI item.\textsuperscript{17}

<table>
<thead>
<tr>
<th>Misconception</th>
<th>Inventory Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Kinematics</td>
<td></td>
</tr>
<tr>
<td>K1. Position-velocity undiscriminated</td>
<td>20B,C,D</td>
</tr>
<tr>
<td>K2. velocity-acceleration undiscriminated</td>
<td>20A;21B,C</td>
</tr>
<tr>
<td>K3. nonvectorial velocity composition</td>
<td>7C</td>
</tr>
<tr>
<td>1. Impetus</td>
<td></td>
</tr>
<tr>
<td>I1. impetus supplied by 'hit'</td>
<td>9B,C;22B,C,E;29D</td>
</tr>
<tr>
<td>I2. loss/recovery of original impetus</td>
<td>4D;6CE;24A;26A,D,E</td>
</tr>
<tr>
<td>I3. impetus dissipation</td>
<td>5A,B,C;8C;16C,D;23E;27C,E;29B</td>
</tr>
<tr>
<td>I4. gradual/delayed impetus build-up</td>
<td>6D;8B,D;24D;29E</td>
</tr>
<tr>
<td>I5. circular impetus</td>
<td>4A,D;10A</td>
</tr>
<tr>
<td>2. Active Force</td>
<td></td>
</tr>
<tr>
<td>AF1. only active agents exert forces</td>
<td>11B;12B;13D;14D;15A,B;18D;22A</td>
</tr>
<tr>
<td>AF2. motion implies active force</td>
<td>29A</td>
</tr>
<tr>
<td>AF3. no motion implies no force</td>
<td>12E</td>
</tr>
<tr>
<td>AF4. velocity proportional to applied force</td>
<td>25A;28A</td>
</tr>
<tr>
<td>AF5. acceleration implies increasing force</td>
<td>17B</td>
</tr>
<tr>
<td>AF6. force causes acceleration to terminal velocity</td>
<td>17A;25D</td>
</tr>
<tr>
<td>AF7. active force wears out</td>
<td>25C,E</td>
</tr>
<tr>
<td>3. Action/Reaction Pairs</td>
<td></td>
</tr>
<tr>
<td>AR1. greater mass implies greater force</td>
<td>2A,D;11D;13B;14B</td>
</tr>
<tr>
<td>AR2. most active agent produces greatest force</td>
<td>13C;11D;14C</td>
</tr>
<tr>
<td>4. Concatenation of Influences</td>
<td></td>
</tr>
<tr>
<td>CI1. largest force determines motion</td>
<td>18A,E;19A</td>
</tr>
<tr>
<td>CI2. force compromise determines motion</td>
<td>4C;10D;16A;19C,D;23C;24C</td>
</tr>
<tr>
<td>CI3. last force to act determines motion</td>
<td>6A;7B;24B;26C</td>
</tr>
<tr>
<td>5. Other Influences on Motion</td>
<td></td>
</tr>
<tr>
<td>CF. Centrifugal force</td>
<td>4C,D,E;10C,D,E</td>
</tr>
<tr>
<td>Ob. Obstacles exert no force</td>
<td>2C;9A,B;12A;13E;14E</td>
</tr>
<tr>
<td>Resistance</td>
<td></td>
</tr>
<tr>
<td>R1. mass makes things stop</td>
<td>29A,B;23A,B</td>
</tr>
<tr>
<td>R2. motion when force overcomes resistance</td>
<td>28B,D</td>
</tr>
<tr>
<td>R3. resistance opposes force/impetus</td>
<td>28E</td>
</tr>
<tr>
<td>Gravity</td>
<td></td>
</tr>
<tr>
<td>G1. air pressure-assisted gravity</td>
<td>9A;12C;17E;18E</td>
</tr>
<tr>
<td>G2. gravity intrinsic to mass</td>
<td>5E;9E;17D</td>
</tr>
<tr>
<td>G3. heavier objects fall faster</td>
<td>1A;3B,D</td>
</tr>
<tr>
<td>G4. gravity increases as objects fall</td>
<td>5B;17B</td>
</tr>
<tr>
<td>G5. gravity acts after impetus wears down</td>
<td>5B;16D;23E</td>
</tr>
</tbody>
</table>
III. EVALUATION OF FCI RESULTS

Unlike most multiple choice tests, the FCI and the earlier MDT distractors come from common-sense student responses to open-ended versions of the questions. This forces the student to choose between the Newtonian concepts and the common sense alternatives. To help interpret FCI results, Hestenes et al. have included two taxonomy tables, one for the Newtonian force concept (Table 4-1) and one for the common sense alternatives (Table 4-2).\textsuperscript{18}

In Table 4-1, the Newtonian force concept is decomposed into six conceptual dimensions: kinematics, first law, second law, third law, superposition principle, and kinds of force. Most physics instructors will agree that all six dimensions are required for a complete understanding of the Newtonian force concept. Each dimension is further broken down into the isolated concepts that characterize that dimension. Each concept is listed with the FCI items in which it appears and the associated Newtonian response. Note that with the exception of item 12, there is only one Newtonian response to each item.\textsuperscript{19} Although each dimension is probed by questions of more than one type, most of the concepts in each dimension are only probed by one or two questions. This is an important point we will come back to at the end of this chapter.

IV. VALIDATION AND RELIABILITY OF THE MDT & THE FCI

A. The MDT

Content validity establishes that the items or questions are relevant to what is being studied and that the subjects’ interpretation of both the questions and answers matches what the developers intended. The content validity of the MDT was established
in four different ways. First, early versions of the test were critiqued by physics professors and physics graduate students. Second, the MDT was taken by graduate students to verify agreement on the correct answers. Third, interviews conducted with introductory physics students verified that they understood the questions and the multiple-choice responses. Fourth, the responses of thirty-one students who received A’s in the introductory physics class were checked for common misunderstandings of individual items.

Reliability indicates that the instrument results or the results from a subsection of the instrument are reproducible for a given group of subjects. The reliability of the MDT was established through interviews and a statistical analysis. A sample of students who had already taken the test were interviewed. During the interviews, the students repeated their responses on the MDT “virtually without exception.” The students gave reasons for their choices and were not easily swayed from their answers as the individual items were discussed. This is an indication that the students’ interview responses reflect stable beliefs as opposed to tentative, random, or flippant responses.

The statistical analysis used the Kuder-Richardson formula, otherwise known as the KR-20. The KR-20 is a special case of the Cronbach Alpha coefficient (see the more detailed discussion on reliability in chapter 5). Both are methods for determining the internal consistency of the test items. Both methods are measures of the intercorrelation of the test items, that is, how well the responses to the items in the test contribute to the overall test score. The KR-20 is used for tests where the items are scored right or wrong. The KR-20 was used on pre-course and post-course MDT results from different, but comparable groups. The KR-20 values obtained were 0.86 for
the pre-course results and 0.89 for the post-course results. A test with a KR-20 value greater than 0.70 is considered reliable enough for measurements of groups. A test with a KR-20 greater than 0.80 is considered reliable enough for measurement of individuals.21 The KR-20 values for the MDT indicate high reliability and suggest that even measurements of individuals should be considered reliable.

B. The FCI

Hestenes, Wells, and Swackhamer did not follow the formal procedures to establish the validity or reliability of the FCI because the FCI is not substantially different from the MDT whose validity and reliability had been established as discussed above. They back up this claim by noting that the FCI and the MDT share a common design, that 14 of the 29 questions on the FCI came unchanged from the MDT, and the results of the FCI and MDT on comparable classes are very similar.

However, Hestenes et al. did interview 20 high school students and 16 first-year graduate students about their responses to the FCI. The responses of the high school students were very predictable and many of the responses sounded as if the students were reciting the results of the studies that led to publication of the MDT.22 They had firm reasons for most of their choices, although there was some vacillation among alternatives. Few non-Newtonian responses were made by students who understand the Newtonian concept involved (false negatives), but Newtonian responses without Newtonian reasoning were “fairly common” (false positive). Because of the false positives and the relatively few false negatives, Hestenes et al. state that except for high score (> 80% or so) the FCI score “should be regarded as an upper bound on a student’s Newtonian understanding.”23
The 16 first year graduate students were beginning the graduate mechanics course at Arizona State University. They participated in in-depth interviews on the questions they missed on the inventory. Although some of the difficulties were attributed to language difficulties both with the foreign students and the native English speakers, the interviews confirmed difficulties with Newton’s 2\(^{nd}\) and 3\(^{rd}\) laws, buoyancy, and friction that had been indicated by the FCI.

In addition, Hestenes et al. demonstrated the reproducibility of the FCI, the mark of a good experimental measurement. It is difficult to give the FCI to the same students twice over a period of time long enough for the students not to remember the questions but short enough that they haven’t learned anything or changed the way they think about forces. But, if the FCI is a reliable test, then classes with similar populations (comparable classes) should get comparable FCI scores. Hestenes et al. found that the post-course averages from classes with over a thousand students in introductory classes taught in the traditional style by seven different professors at Arizona State University were remarkably similar. The post-course averages were all between 60 and 64\% correct. The reproducibility of FCI results from comparable classes has since been observed by others including Hake (see Hake’s 6000 Student Study and the h-factor discussion in section V. below) and myself (see chapter 9).

More recently, Hestenes and Halloun\(^{24}\) have used interviews with students and other tests to compare students FCI scores to other measures of Newtonian skills including problem solving.\(^{25}\) Based on their results, Hestenes and Halloun suggest an interpretation of FCI scores that is consistent with a three-stage model of conceptual understanding in learning Newtonian mechanics.
Students who score below 60% on the FCI are classified as stage I. Stage I student thinking can typically be described in terms of the following characteristics:

- undifferentiated concepts of velocity and acceleration,
- lack of a vectorial concept of velocity,
- belief that there are other influences on motion besides forces,
- inability to reliably identify passive and active agents of force on an object, and
- fragmented and incoherent concepts about force and motion.

Students who score between 60% and 85% on the FCI are in stage II. Hestenes and Halloun suggest that an FCI score of 60% be considered as the entry threshold to Newtonian thinking. In stage II, students are developing coherent dynamics’ concepts, including vectorial concepts of velocity, acceleration, and force. An FCI score of 85% is interpreted as the threshold to stage III and mastery of the Newtonian force concept.

Students in stage III develop a complete Newtonian interaction concept including a full understanding of Newton’s third law. Hestenes and Halloun express confidence in “identifying students with scores above [85%] as confirmed Newtonian thinkers.”

In addition, Hestenes and Halloun believe their results indicate that not only is the conceptual development of students influenced by the order in which the concepts are introduced in instruction, but that there is a natural order in which the concept are most easily learned. A similar study of student understanding of kinematics by Thornton also suggests the existence of a natural order for learning mechanics concepts.
V. RESULTS FROM FCI/MDT STUDIES

A. Halloun, Hestenes, and the MDT (1985)

Using the MDT in conjunction with a math-skills diagnostic test of their own design, Halloun and Hestenes found that the initial knowledge state of the students as measured by the two pretests is a significant predictor of student performance in traditional introductory physics classes. In combination the two tests accounted for about 42% of the variance of exam scores, that is 42% of the variability in exams scores can be predicted by the two diagnostic tests. The MDT alone accounts for 31% of the variance. Differences in students’ gender, age, academic major, and academic background were found to have little or no effect. They also found that the pretest MDT scores for calculus-based classes (51%, 51%, 50%, and 53%) were significantly better than the scores for algebra/trig-based classes (37%, 37%, 40%) with relatively narrow distributions for both. That the calculus-based classes did better is not surprising, but considering the simplicity of the tests, the math and physics diagnostics average scores were thought to be low for both types of classes. The reader is reminded that a low score on the MDT does not just mean that students don’t know or understand the concepts of Newtonian mechanics, but that common sense misconceptions may be firmly in place. Based on my own study of FCI scores from both calculus-based and algebra/trig-based courses from 17 colleges and universities, the scores Halloun and Hestenes found in 1985 are typical pre-course scores for calculus-based and algebra/trig-based courses for schools of moderate selectivity.

The teaching styles of the four instructors in the calculus-based introductory classes in the Halloun and Hestenes study were very different. Professor A was a
Theorist whose lectures emphasized the conceptual structure of physics with careful definitions and orderly logical arguments. Professor B expended a great deal of time and energy to incorporate many demonstrations in his lectures to help students develop physical intuition. Professor C emphasized problem solving and modeled one problem after another in his lectures. Professor D, as an experimental physicist teaching physics for the first time, tended to follow the book closely in his lectures. All four professors were known as good teachers according to informal peer opinion and student evaluation. In fact, Professor B won two awards for outstanding teaching. The students of professors A, B, and C increased their scores on the MDT by 13%; the students of professor D increased by 11%.

Despite the wide variety of teaching styles of the four instructors, the result on the MDT was essentially the same. Halloun and Hestenes therefore conclude from this result “that the basic knowledge gain under traditional lecture instruction is essentially independent of the professor.” They comment that this result is consistent with observations by physics instructors that “the most strenuous efforts to improve instruction hardly seem to have any effect on the general student performance.”

Halloun and Hestenes interpret these results in the following way. The low initial level and small gain of students’ basic knowledge as measured by the MDT indicate that throughout the course students are operating with a defective vocabulary and belief system. This implies that students are continually misunderstanding the course material presented in class. Furthermore, in their words,

The small gains in basic knowledge ‘explain’ the high predictive validity of the mechanics pretest; the student’s ability to process information in the course depends mainly on his initial knowledge state and hardly
improves throughout the course. The high predictive [power] of the [MDT] pretest is not intrinsic to the test; rather, it indicates a failure of [traditional] instruction. The more effective the instruction is in altering the basic knowledge state [of the students], the lower the predictive [power] of the pretest.30

B. The Hestenes, Wells, and Swackhamer Study (1992)

Hestenes, Wells, and Swackhamer used the FCI as a pre/post course evaluation of high school and university introductory physics classes.31 They found low pre-course FCI scores for most classes and smaller post-course FCI scores for traditional classes. The classes using research-based pedagogy tended to have higher post-test scores.

Their study of high school classes included 19 high school teachers involved in an NSF physics instruction enhancement program run by Wells and Hestenes. The FCI was used as a pre/post evaluation of the high school teachers’ classes. Hestenes et al. compared the FCI scores with ratings of the socioeconomic level of the students at the teacher’s schools and with a ranking of the teacher’s competence as measured by academic background, concept test score, and teaching experience. The results appeared independent of socioeconomic level of the school but this could be due to selection bias since the students who take physics tend not to be typical of the student population at the school. With one exception, the results tend to be independent of the teacher competence ranking. One instructor scored 39% on the Mechanics Diagnostic; that instructor’s students only scored 33% on the post-course test, the lowest post score in the study. This implies that the students’ FCI scores may only be affected by the quality of the instructor in extreme cases.

As part of the enhancement program, the teachers took a workshop on the teaching methods of Wells. The method uses computer-assisted laboratory-oriented
instruction with no lectures, but with substantial class discussion. The high school
teachers’ classes were tested before and after the workshop. The classes showed little or
no improvement. Upon further examination, Hestenes et al. found that in their first year
of instruction with the Wells’ method the teachers focused on the mechanics of the
method and not on the pedagogy. This suggests that technology by itself cannot
improve instruction. The best that technology can do is enhance the effectiveness of
good pedagogy.

C. Hake’s 6000 Student Study and the h-factor (1993-97)

A detailed study of pre/post MDT and FCI results nationwide by Hake compares the
performance of 62 introductory physics classes (N = 6542 students).\textsuperscript{32} The student
populations sampled include high schools, colleges, and universities. Hake's results show
an interesting uniformity. When the class's gain on the FCI (post-test average - pre-test
average) is plotted against the class's pre-test score, classes of similar structure lie
approximately along a straight line passing through the point (100,0). This is shown
schematically in the Hake plot shown in Fig. 4-1.\textsuperscript{33} Traditional classes lie on the line
closest to the horizontal axis and show limited improvement. The middle line represents
classes with active engagement (see chapter 2 for a discussion of active engagement).
The steepest line represents classes with active engagement and a research-based text.
The negative slope of the line from a data point to the point (100,0) is a figure of merit:

\[ h = \frac{(\text{class post-test average} - \text{class pre-test average})}{(100 - \text{class pre-test average})} \]

The interpretation of this is that two classes having the same figure of merit, \( h \), have
achieved the same fraction of the possible gain.
The 62 classes are classified as either “traditional” or “active engagement.”

These terms are defined as follows:

*Active Engagement:* Courses that incorporate at least in part research-based teaching methods where the students are actively involved in discussion, discovery and/or analysis of the course material. The students work on these activities in cooperative groups or for at least one hour per week in class and in some cases as many as six hours per week. In these active learning activities, the teaching emphasis is on helping the students answer their own questions and think about what they are learning.

*Traditional:* Courses that do not make use of active engagement activities. Typically, the primary activity in the lecture and recitation components of a traditional class involves students listening to a lecture presentation or answers to their questions by an instructor. Students may make predictions for demonstrations but do not discuss them with their peers or write them down. Hands-on activities like laboratories follow fairly detailed instructions that minimize decision making.

---

**Figure 4-1.** Schematic of the Hake plot. A class’s average pre-test FCI score is plotted on the horizontal axis, the pre- to post-test gain is plotted on the vertical axis. Since the maximum class average is 100%, every data point must lie in the shaded region. The three lines of constant fractional gain \( h \) described in the text are shown in the figure below. (This figure is reproduced from “On the effectiveness of active-engagement microcomputer-based laboratories” by Redish, Saul, and Steinberg in the American Journal of Physics. 34)
The main distinction of an active engagement activity is whether or not a majority of the students are actively discussing, arguing, or testing the course material. The classes in Hake’s study were categorized based on information from the instructors either through private communications with Hake or through published accounts of the curriculum. Hake found that active engagement classes had significantly higher fractional gain than traditional classes. A Hake plot of the FCI/MDT scores from his survey is shown in Figure 4-2. The average fractional gains for both types of classes are:

- Traditional Classes (14 classes) \( \langle h \rangle = 0.23 \pm 0.04 \) (std. dev.)
- Active Engagement Classes (48 classes) \( \langle h \rangle = 0.48 \pm 0.14 \) (std. dev.)

where \( h \) is averaged over classes, not students.

To test the significance of this result, I performed a two-tailed t-test with pooled variance on Hake’s data.\(^{35}\) Assuming that these two measurements are drawn from similar populations, the difference between the two types of classes is statistically significant since the probability that this difference in means is due to random chance is less than 0.1%. This result clearly shows that the active engagement classes are achieving significantly greater fractions of the possible gain on the FCI. All but seven of the forty-eight active engagement classes had higher \( h \)’s than all of the traditional classes. Hake noted that case studies showed that these 7 active engagement classes with relatively low \( \langle h \rangle \)’s (0.20 < \( \langle h \rangle \) < 0.28) had various implementation problems.
Percent gain is plotted vs. percent average pretest score. The data shown below represents pre- and post-test results from 62 courses, 14 traditional lecture courses (N = 2084 students) and 48 active engagement classes (N = 4458 students). The slope line labeled $<g>_{48IE}$ represents the average fractional gain $h$ for the 48 active engagement classes. The slope line labeled $<g>_{14T}$ represents the average fractional gain $h$ for the 14 traditional lecture classes.
with the active engagement methods. These problems included:

- insufficient training of instructors new to [active engagement] methods,
- failure to communicate to students the nature of the science and learning,
- lack of grade incentives for taking [active engagement] activities seriously,
- a paucity of exam questions which probe the degree of conceptual understanding induced by [active engagement] methods, and
- use of active engagement in only isolated components of the course.\textsuperscript{37}

Note that several of these problems regarding communication and assessment are related to the issues of the “hidden curriculum” discussed throughout this dissertation.

Hake’s inference that $h$ is a measure of course effectiveness is bolstered by examining the correlations between the average normalized gain $<h>$, the average unnormalized gain $g$ ($g \equiv$ average post-test score - average pre-test), and the average pre-test scores averaged over all 62 classes. There is a low correlation between $h$ and the pre-course score ($r = 0.02$). This implies that students’ knowledge coming into the course does not strongly influence normalized gain, $h$. There is also a negative correlation between the unnormalized gain (post-course score - pre-course score) and the pre-course score ($r = -0.49$). This negative correlation indicates that classes that start with smaller pre-course scores tend to have larger unnormalized gains than classes of the same category with larger pre-course scores. The latter correlation is quite evident from the plot of Hake’s FCI/MDT survey data shown in Figure 4-2.

Hake collected this data to address the question, “Can the classroom use of [active engagement] methods increase the effectiveness of introductory mechanics courses well beyond that attained by traditional methods?” Note that this question is similar to the research questions explored by this dissertation, but Hake limits his study
to the results of multiple-choice tests. (There are also design differences between the
two studies discussed below.) To answer this question, Hake solicited pre/post FCI and
MDT data from the physics education community through talks at colloquia and
meetings and e-mail postings on the PHYS-L and PhysLrnR listserves.\(^{38}\) Hake
acknowledges that his method of data solicitation “tends to pre-select results which are
biased in favor of outstanding courses which show relatively high gains on the FCI;”
because relatively low gains are

neither published not communicated except by those who wish to use the
results from a traditional course as a baseline … or possess unusual
scientific objectivity and detachment. Fortunately, several in the latter
category contributed data to the present survey for courses in which
active engagement methods were used but relatively low gains were
achieved.\(^{39}\)

However, Hake also notes that this survey includes all the pre/post FCI results from
traditional classes that were known to him.

In addition, while Hake controlled or took into account several kinds of
systematic error, not all the data Hake presents is matched so that the result represents
real changes in the students and not differences in which students took the FCI.\(^{40}\) Hake
estimates that the error in the normalized gain \(h\) is probably less than 5% for classes with
20-50 students and less than 2% for classes with more than 50 students. In my own
studies of FCI results reported in chapter 9, I found that the matched and unmatched
results are not significantly different. Despite the known flaws in Hake’s study, his result
of a very significant difference in \(<h>\) for traditional and active engagement classes is still
quite impressive.
Prior to this dissertation, Hake’s study was the only one to compare concept test results from classes using a variety of active-engagement methods with traditional instruction. In chapter 9, I present FCI results from approximately 2000 students from calculus-based introductory physics classes using three research-based curricula using active-learning methods and traditional instruction from eight of the ten schools participating in this study. Where possible, results are presented from traditional classes and active engagement classes at the same school. The selection bias in Hake’s data was avoided since data was taken from all first term classes in the sequence being studied at a particular school. The data are matched unless explicitly stated otherwise.

VI. DISCUSSION OF WHAT IS MEASURED BY THE FCI

A. The 4-H Controversy (1995)

The so-called 4-H controversy began in March 1995 when Huffman and Heller reported on the results of a factor analysis of FCI data.\textsuperscript{41} They questioned what is measured by the FCI and the dimensions of the Newtonian force concepts defined by Hestenes \textit{et al}. and shown in Table 4-1. Hestenes and Halloun defended their interpretation of the test and were in turn rebutted by Heller and Huffman.

The point of debate in this controversy is quite subtle. Hestenes, Halloun, Heller and Huffman all acknowledge that the FCI is a useful test for evaluating instruction and as a diagnostic tool for classes. All four also agree that Hestenes \textit{et al}. have demonstrated the reliability (the results for similar classes are reproducible), the face validity (physics instructors agree that Halloun et al.’s six dimensions are necessary for a Newtonian force view and that the Newtonian choice for each item is correct), and
context validity (the developers interviewed students to confirm the validity of the student responses to individual items) of the FCI. What is in dispute is the construct validity of the FCI, that is whether or not the FCI actually measures the student force concept and the six conceptual dimensions of force.

To understand the nature of the controversy, it is necessary to understand that a factor analysis is a statistical method for finding structure in a set of data by analyzing the inter-item correlations. The factors are those groups of items that show relatively strong correlations with each other. A more detailed description can be found in the next chapter in the section on reliability. Factor analysis is often used to check tests and surveys to see if all of the questions are correlated with the main issue being addressed or if all the questions along a claimed dimension are correlated to one another. If, as in the case of the FCI, a test was designed to measure an issue like student understanding of Newtonian force and the developers claimed the test measured student understanding across six dimensions as well as overall understanding of Newtonian Force, then one might expect that the students responses to correlate both within dimensions and overall. Huffman and Heller did a factor analysis of FCI data taken at the end of the mechanics section of introductory physics classes at the University of Minnesota and at a suburban high school in Minneapolis.

Their factor analysis yielded one factor that corresponded to one of Hestenes et al’s Newtonian force dimensions (the Newton third law dimension) and part of a second (the kinds of force dimension). In general, they found that the student responses to the FCI in general were weakly correlated and therefore not more than loosely related. That is, the statistics of the student responses to the FCI did not correspond to Hestenes et
al.’s taxonomy of the six conceptual dimensions shown in Table 4-1 that comprise the Newtonian force concept. Also, one would expect that since the FCI items all concern the Newtonian force concept, the student responses for the different items should correlate with one another. Because Huffman and Heller found that all of the items are only loosely related to one another, they were unable to statistically show that the force concept defined by Hestenes et al. is the same as the student force concept. In short, Huffman and Heller were unable to statistically verify that the students’ responses showed a structure that was consistent with the defined structure of Newtonian force.

However, two things should be kept in mind. First, statistics are estimates and they are only one way to look at data. They are not the only way. Hestenes et al. interviewed students to verify not only that the students reasoning and responses were consistent in responding to the FCI items but also that the FCI result reflected the degree to which the students appeared to hold the Newtonian force concept as defined. Second, the weak correlations of the student responses are consistent with our understanding of how students structure their physics knowledge. These results come from physics education research on students’ conceptual understanding and problem solving (see chapter 2).

The findings of Hestenses and Halloun, Minstrell, and diSessa among others indicate that student knowledge structures are often not coherent, well defined, consistent, or logically organized. The students’ knowledge structure can be characterized as knowledge fragments where each fragment refers to a specific idea or situation. Both Minstrell’s facets and diSessa’s phenomenological primitives or p-prims can be characterized in this way.
In addition, student knowledge and reasoning is often context dependent. Almost every physics instructor has observed that many of their students fail to apply concepts and reasoning they have seen in class and homework assignments if they are given a problem in an unfamiliar context. In another example, physics education researchers have observed that students who appear to have acquired a Newtonian understanding of linear motion will revert to their common sense beliefs when first confronting rotational motion. As we discussed in chapter 2, this implies that student knowledge does not exist in a common sense state or a Newtonian state, but rather that the common sense beliefs and the Newtonian concepts form a superposition state where the probabilities depend on what is triggered by the context of the situation or problem.

This is the heart of the 4-H controversy. The issue is not whether the Newtonian force taxonomy is a correct representation of the essential components of Newtonian force, but whether the taxonomy represents how students think about Newtonian force. The weak correlations between the items of the FCI found by Huffman and Heller suggest that if the FCI questions are measuring students’ knowledge of the Newtonian force concept then “the FCI measures bits and pieces of students knowledge that do not necessarily form a coherent force concept.” However, it is still useful to use the Hestenes et al. taxonomy of the Newtonian Force concept and their taxonomy of misconceptions to see what aspects of Newtonian force students approach correctly and where and how the students use their common sense beliefs. One need only keep in mind that the two taxonomy structures may not represent how the students think about Newtonian force. (Please see the crystal axes analogy in chapter 5 on page 180.)
B. Steinberg & Sabella’s Comparison of FCI Responses & Exam Problems (1997)

Noting that only a few studies have been published on how student performance on the FCI correlates with other measures of students conceptual understanding, Steinberg and Sabella recently reported on a comparison of student responses on the FCI and two qualitative exam problems. The two qualitative exam problems are shown in Figure 4-3. The concepts needed to solve these problems correspond to concepts addressed by several questions on the FCI by design as shown in Table 4-3 below. The elevator problem was written by Steinberg; the two-cart problem was written by the author of this dissertation. The two problems were given on final exams in consecutive semesters one week after the students took the FCI. Because of the relatively small number of student responses, Steinberg and Sabella suggest that the results of their study may not be statistically significant. However, they feel that the results “give an indication of the types of issues that should be important to instructors using multiple-choice instruments.”

In their analysis, Steinberg and Sabella looked at how individual students did the exam problems and the comparable FCI problems. They found that while there was a correlation between student performance on the exam and the comparable FCI questions, for certain students and certain questions the responses and the reasoning varied greatly. For example, both part A of the elevator problem and FCI question 18 ask about the forces on an elevator moving with constant velocity. Only 54% of the students correctly answered FCI question 18, although 90% of the students answered part A of the elevator problem correctly even though the situations are identical. The majority of incorrect responses indicated a net force in the direction of motion which
Figure 4-3a. Elevator problem that corresponds to several FCI questions from Steinberg and Sabella, “Performance on multiple-choice diagnostics and complementary exam problems,” in *The Physics Teacher*.

**Exam problem 1:** Ignore all friction and air resistance in this problem.

A. A steel ball resting on a small platform mounted to a hydraulic lift is being lowered at a constant speed, as shown in the figure at right.

i. Draw a free body diagram of the ball.
   Describe each type of force.

ii. Compare the magnitudes of the forces you have drawn.
   Explain your reasoning.

B. As the ball is moving down, a bullet moving horizontally hits the exact center of the ball (see figure at right) and then ricochets straight back. This causes the ball to immediately fall off the platform.

i. Draw a free body diagram of the ball after it is no longer in contact with the bullet or the platform. Describe each type of force.

ii. A vector that represents the velocity of the ball just before the bullet hits is shown below. Draw vectors that could represent the velocity at each of the 2 other times indicated. The scales of the 3 vectors should be consistent with each other. Explain your reasoning.
Exam Problem 2: Two carts, A and B (\(\text{Mass}_A > \text{Mass}_B\)), are placed on a table then stuck together with Velcro. Using pulleys, two small blocks, C and D (\(\text{mass}_C < \text{mass}_D\)), are connected by light strings to the carts as shown below. Initially, the carts are held in place. Ignore all friction in this problem.

At \(t = 0\), the carts are released.

A. Draw and label two separate free-body diagrams, one for each cart, for a time after the carts start moving but before the Velcro pulls apart.

B. Rank all the horizontal forces from both your diagrams by magnitude, from largest to smallest. Explain the reasoning that you used to rank the forces.

Table 4-3. Comparison of concepts needed to solve exam problems and FCI items

<table>
<thead>
<tr>
<th>Concept</th>
<th>Problem</th>
<th>Comparable FCI items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newton’s 1(^{st}) Law</td>
<td>Elevator Problem – Part A</td>
<td>Question 18</td>
</tr>
<tr>
<td>Newton’s 2(^{nd}) Law</td>
<td>Elevator Problem – Part B(i)</td>
<td>Questions 9 &amp; 22</td>
</tr>
<tr>
<td>Motion Diagrams</td>
<td>Elevator Problem – Part B(ii)</td>
<td>Questions 12 &amp; 28</td>
</tr>
<tr>
<td>Newton’s 3(^{rd}) Law</td>
<td>2 Cart Problem – Part B</td>
<td>Questions 2, 11, 13, &amp; 14</td>
</tr>
</tbody>
</table>
is inconsistent with Newton’s first law of motion. On the FCI questions and the part of
the elevator problem that ask students to identify the forces acting on an object in motion
after a brief impulse, only two thirds of the students answered the three virtually identical
items consistently. The motion diagrams use different representations in the FCI
questions and on the exam although the situations described are physically similar. While
43% of the students had responses that were consistent, another 43% wrote answers on
the exam that corresponded to FCI choices that were different from the ones they
selected.

On the FCI questions and the part of the exam problem in figure 4-3b that checks
students’ understanding of Newton’s third law (N3), the student responses on the FCI
questions and on the exam problem correlate, but not perfectly. Steinberg and Sabella
note that only 65% of the students who answered three or four of the FCI items
correctly were able to identify and compare the N3 force pair, 53% of those who
answered two of the FCI items correctly, 29% of those who answered only one of the
four FCI items correctly, and 14% of the students did not answer any of the FCI items
correctly. Moreover, on FCI question 13, which describes a situation very similar to the
exam problem where two objects remain in contact and accelerate uniformly for the
entire motion, roughly 50% of the students answered the two consistently. Twenty-one
of the students who answered the exam problem incorrectly clearly stated their
reasoning. Of these, only six answered FCI question 13 in a way that was consistent
with their reasoning on the exam.

Steinberg and Sabella offer the following explanations as to why the performance
of a large minority of students does not seem to correlate:
1. Student models for physical systems can be ill formed and inconsistent.

2. The FCI looks colloquial, the exam formal.

3. The FCI is multiple-choice, the exam is open-ended.

   All three of these arguments have merit. The first explanation recognizes that students often respond differently to problems that appear to physicists to be basically the same problem with different surface features. This is consistent with the earlier discussion of the structure of students’ knowledge and the roles of cues for triggering different responses. The third explanation is simply recognition that students sometimes respond differently to questions where they have to come up with an answer and an explanation on their own as opposed to selecting from a set of choices. The choices can trigger responses that students might not come up with on their own.

   The FCI questions are worded like real world experiences, the qualitative exam problems sound more like problems studied in a physics class. In the former, the students are asked a question about forces; in the latter, they are asked to draw free body diagrams and rank the forces. The wording and nature of the exam problems is more likely to trigger a classroom physics response. Another difference between the FCI and the exam questions in this study was the context in which they were taken by the students. The students took the FCI in the last week of classes and were told they would receive credit for participation but not for the correctness of their answers. In the final exam, the students know their grades depend on their demonstrating what they learned in class. I suggest that this makes it much more likely for students to use their formal physics knowledge on the exam rather than their common sense beliefs. I will discuss this point in chapter 9.
VII. THE FORCE AND MOTION CONCEPT EVALUATION (FMCE)

The FCI gives a good indication of a student’s overall understanding of the Newtonian concept of force and it can be used to gauge common-sense beliefs held by the class as a whole. However, because some issues are addressed by only a few questions in one or two contexts, it is not a precise enough instrument to determine which common-sense beliefs are held by a single student or under what circumstances they are used. In this sense the FCI is a survey instrument. The results of a survey instrument can indicate possible problems but lack the resolution to be a reliable measure of a single student. A diagnostic instrument may address fewer issues than a survey instrument but covers each issue thoroughly in multiple contexts.

To aid their development of MBL laboratory and demonstration curricula, Thornton and Sokoloff wanted an instrument that would give more information about the common sense beliefs of each student than was possible with the FCI, i.e. a diagnostic test of Newtonian concepts. They developed the Force and Motion Conceptual Evaluation (FMCE) to address this need (a copy of the FMCE can be found in Appendix B). The FMCE does not cover as much material as the FCI (for example, there are no questions on circular motion) but uses more questions for each concept and approaches them in a few different contexts. Also, the FMCE places more emphasis on students’ understanding of graphical representations of velocity, acceleration, and force. A few of the schools participating in this study used the FMCE instead of the FCI to measure students’ conceptual understanding of mechanics.

A few of the FMCE questions are identical to items on the FCI; the remainder were developed through student interviews and testing open-ended versions of the
questions with explanations. Like the FCI, the FMCE distractors come from the most common student responses to open-ended versions of the questions. To test the validity, Thornton and Sokoloff have given the FMCE to hundreds of physics faculty, compared student responses on multiple-choice versions and open-ended with explanation versions of the FMCE, and asked additional questions on exams to compare with the FMCE results. The faculty agree with the interpretation and Newtonian response to all 43 questions. There is an extremely high correlation between the student responses to the various styles of questions, particularly the multiple-choice and open-ended with explanation versions of the FMCE questions (>90%). In addition, the pre and post results have been very stable and repeatable, comparing equivalent classes at several different schools for both traditional and enhanced instruction.

Thornton and Sokoloff have reported FMCE results that are very similar to the FCI results discussed earlier in the chapter, i.e. that traditional classes show little gain (less than 10% increase in the total score) while classes using research-based curricula show much greater gains. Thornton and Sokoloff note that patterns of individual student responses are usually consistent with either a Newtonian understanding or specific common-sense beliefs about motion. Their results also show that many students do not apply a consistent model to questions on force and motion. In their words, “many students view speeding up, slowing down, moving at constant speed, and standing still to be independent states of motion that do not require a consistent relationship between force, acceleration, and velocity.”

Since a few schools only used the FMCE to measure conceptual understanding, one may ask whether the FMCE and the FCI are comparable. In the 1995 spring
semester at University of Maryland, the students in the first semester of the introductory physics for engineering majors sequence took the FCI on the first day of class and the FMCE in recitation sections later in the same week. I found the overall scores have a correlation (Pearson product coefficient) of 0.79 which is high enough for a direct comparison of the two scores.

In chapter 9, the results from the velocity graph and the Newton 3 questions from the FMCE will be discussed in detail in addition to the overall FMCE results.

VIII. WHAT DO MULTIPLE CHOICE CONCEPT TESTS TELL US ABOUT STUDENTS’ CONCEPTUAL UNDERSTANDING?

The discussion in this chapter indicates that tests like the FCI appear to measure students’ understanding of basic concepts compared with their common-sense beliefs. Both the FCI and FMCE have proved their validity and reliability through extensive testing and comparisons with interviews and open-ended versions of the questions. In addition the results have been shown to be very robust in classes at many institutions. The results of pre- and post-course testing indicate that classes taught with traditional lecture instruction show small gains on these tests, while classes using research based active-engagement activities resulted in significantly better gains. Validation studies have shown that the results are consistent with students’ understanding of basic Newtonian concepts; although factor analysis suggests that student knowledge is fragmented. Students’ conceptual knowledge can appear to be a mixture of both Newtonian and common-sense beliefs.
In addition, while the FCI results were strongly correlated with the exam problems in the Steinberg and Sabella study, the responses for certain students and certain problems differed greatly. This implies that tests like the FCI may not necessarily be a good measure of students’ ability to use concepts in problems, i.e. test the functionality of their conceptual knowledge. Furthermore, student responses to questions that test their understanding of concepts appear to be triggered by cues in the context of the problem in question and the test itself. More research is needed to see how the FCI compares with measures of students thinking and reasoning using physics concepts.

The FCI has played an important role of convincing the physics education community of the extent of student difficulties with conceptual understanding of Newtonian force and limited effect of traditional lecture instruction. However, while the FCI is a useful measure of students’ conceptual understanding, there is a tendency in the community to rely solely on tests like the FCI. This over-simplifies the view of both student learning and assessment. The FCI should be one part of a broad-based comprehensive assessment.


7 Some examples are the *Heat and Temperature* concept test and the *Circuits* concept test written by Thornton and Sokoloff (unpublished, but available from the authors) as well as the Wave Concept Test being developed by Michael Wittmann at University of Maryland. There are efforts currently underway by Van Heuvelen, Heiggelke, Maloney, and Beichner to write a concept test for electricity and/or magnetism.


11 The 4 H’s are Halloun, Hestenes, Heller, and Huffman. The controversy name comes from two articles on this topic published in the same issue of *The Physics Teacher*. One article was a response by Halloun and Hestenes to an earlier article by Heller and Huffman and the other article was Heller and Huffman’s counter-response.

12 See Ref. 9.


16 Private communications with Pat Heller at University of Minnesota and Edward Redish at University of Maryland.
The correct Newtonian response to item 12 (d) requires an understanding of buoyancy that escapes even some physics graduate students. Because of this difficulty and since buoyancy is not even covered in some introductory classes, response (b) is also considered a Newtonian response.


See Ref. 33.


Phys-L is a listserv for undergraduate and high school physics instructors to discuss issues related to physics teaching. PhysLrnR is a listserv for discussing experimental and theoretical physics education research topics.

See Ref. 31.

Private communication with R. Hake, October 1997.

See Refs. 9, 11, 13 & 14.


See Ref. 14.

See Ref. 10.

See Ref. 46.


50 See Ref. 8.

51 Up to forty percent of physics faculty initially fail to pick the correct answer for question #6, although they change their response to the correct answer when challenged. Thornton and Sokoloff have been unable to come up with a new wording for this problem and suggest that question 6 be evaluated in conjunction with other questions on force and acceleration.

52 See Ref. 44.
Chapter 5. Measurements of Expectations:
The Maryland Physics Expectations (MPEX) Survey

I. OVERVIEW

A. What are Expectations?

In chapter 2, several examples were used from the Physics Education Research literature to demonstrate that students do not come into the introductory physics class as blank slates. In the previous chapter, we showed that concept tests like the FCI\(^1\) and FMCE\(^2\) have determined that many students have common sense beliefs about physics concepts that can hinder their learning of the physics concepts taught in the introductory class. However, the discussion in chapter 2 showed that it is not only physics concepts that a student brings into the physics classroom. Each student, based on his or her own experiences, brings to the physics class a set of expectations about what sorts of things they will learn, what skills will be required, and what they will be expected to do. In addition, their view of the nature of scientific information affects how they interpret what they hear. As we saw in chapter 2, students’ expectations can have a strong effect on what they get out of a physics class.

B. Why Study Expectations?

While it is important for students to learn a functional understanding of the physics concepts, it is also important for the students to develop expectations favorable for developing a deep, functional understanding of physics. As we saw in the examples of Mazur,\(^3\) Hammer,\(^4\) and Tobias\(^5\) in chapter 2, the students’ expectations have a significant effect on what the students do to learn physics, their idea of what physics is, the type of understanding they build, and ultimately what they get out of the class. In
many traditional introductory physics classes, some students’ expectations and their
desire for a good grade with a minimum of work may lead them to a false sense of what
learning physics is about. Like Hammer’s student Liza in chapter 2, these students may
believe that by memorizing formulas and problems solutions and using what they’ve
memorized to solve typical end of chapter problems on exams, that they have
successfully learned physics in addition to doing well on the exams.

C. Why a Survey?

The FCI and other concept tests like it, as described in the previous chapter, have
played a major role in convincing physics instructors of the validity of studies of the
nature and persistence of students’ common sense beliefs in traditional instruction as well
as evaluations of curricula designed to improve students’ conceptual understanding.
There are several reasons why tests like the FCI have become an almost indispensable
assessment instrument to physics instructors. Unlike many research-based assessment
methods, they can easily be used and interpreted by instructors who are not physics
education researchers. They can be used to roughly determine the distribution of
Newtonian and common sense beliefs over a whole class. In addition, they can be used
as a pre/post evaluation tool to see if and how student responses change. Several
curricula that take into account the students’ common sense beliefs have shown
significant improvements in students’ conceptual understanding as measured by multiple-
choice concepts tests compared to traditional instruction (see chapter 4).

However, while a great deal is known about student common sense beliefs about
concepts in introductory physics courses, very little is known about the nature and
effects of student expectations in physics classes. The few studies that exist have mainly
used individual student interviews. What they have taught us about student learning is alarming, but we need to learn how serious and how widespread the problems demonstrated by these studies are. If inappropriate expectations play a significant role in the difficulties our students commonly have with introductory calculus-based physics, physics education researcher and physics instructors need way to track and document them. In particular, we need an instrument equivalent to study this issue. It needs to be convenient to give and easy to analyze for a physics instructor who is not a physics education specialist to use with their own classes as a pre/post evaluation. This would give us a sense of the distribution and evolution of these expectation “misconceptions” in the introductory physics classroom.

In this chapter, I describe our development of the MPEX survey, a 34-item Likert-scale (agree-disagree) survey that probes students’ cognitive expectations. The MPEX survey is one of several instruments being developed to meet this need. Note that because of their brevity, these instruments are surveys and not diagnostics. (See the discussion on this issue in chapter 4, page 129.) While they are useful as an instrument for learning how broad and prevalent expectation misconceptions are in a class, they are often less reliable measures than diagnostics for evaluating individual students.

D. Chapter Layout

In this chapter, I describe the development, the structure, and testing for reliability and validity of the MPEX Survey. The MPEX survey results from introductory physics classes are presented in chapter 10. In section II, I describe the development of the MPEX survey over the last five years by Redish, Steinberg, and the author. In section III, I describe both the 34 survey items and the construction of seven
dimensions of expectations. Section IV contains results on two tests of the validity of
the MPEX survey deals, the results of the survey with five calibration groups and the
results from interviews with students taking the survey and going over their responses.
Tests of the MPEX survey reliability including factor analysis, Cronbach alpha, and
reproduction of results are presented in section V together with a brief discussion of the
survey data’s statistical uncertainty.

II. DEVELOPMENT OF THE SURVEY

This study began as an attempt to extend Hammer's work. Detailed interviews
are too slow and expensive to allow the determination of distribution functions – how
many students in various classes hold particular views. We want to know such things as
what is the distribution of students’ expectations coming into the introductory physics
class, whether students in universities or junior colleges have different attitudes or
distributions of attitudes, and whether the distribution of attitudes found in a study at a
particular college or university are representative of what is found throughout the
country. To determine these distribution functions requires an efficient and easily
delivered instrument — a questionnaire or short answer test. A major goal of the MPEX
project has been to develop and evaluate such an instrument.

We also want to know the role of dynamics. While Hammer did not observe
students changing their views, the studies by Perry\textsuperscript{10} and Belenky \textit{et al.}\textsuperscript{11} (see the
discussion on expectations in chapter 2) found that many of their young adult subjects
were able to evolve more sophisticated expectations about general knowledge.
Although neither Perry’s nor Belenky \textit{et al.’s} studies focused specifically on scientific
situations, these two studies give us hope that many students can be moved from Hammer's category B to the more sophisticated and scientific category A. A second goal of this project is to understand whether such transitions are possible and to begin to specify what activities in a university physics class can help such transitions take place.

Redish and Steinberg began to develop the MPEX survey in the 1992 fall quarter at the University of Washington. The first version of the expectations survey was delivered in the spring quarter of 1993 to students in the three-quarter calculus-based introductory physics sequence at the University of Washington. The students in first-quarter course were surveyed at the beginning of the quarter and the students in the third-quarter course were interviewed at the end of the quarter. Responses from more than 100 students were obtained from each course. These students in the introductory calculus-based physics class were given a survey with 51 statements about the nature of physics, the study of physics, and their relation to it. They rated these statements on a five point Likert scale from strongly disagree (1) to strongly agree (5). The survey was also given to eight faculty members who had taught the class before. The survey items were chosen as a result of a detailed literature review, discussions with physics faculty, and the designers' combined 35 years of teaching experience.

Upon his return to University of Maryland, Redish gave a new version of the survey to students in an experimental class of the engineering physics sequence. Based on the analysis of the University of Washington results, this second version of the survey was pared to 35 statements. My involvement with this project began with the second semester of this sequence. Up to this point, the surveys were given to the students as paper and pencil instruments that had space for comments. For the second version of the
survey, I validated the survey items by interviewing students and by examining students’ written comments from the class surveys. Beginning with the third version of the survey and widespread distribution with other schools, Redish and I developed a scantron version of the survey to allow for easy processing of large numbers of completed surveys from the 17 participating colleges and universities.

We validated the survey items from the current version of the survey (version 4.0) in a number of ways: by discussion with other faculty and physics education experts, through student interviews, by giving the survey to a variety of "experts", and through repeated delivery of the survey to groups of students. I conducted over 125 hours of interviews with over 100 students at eight of the participating colleges and universities.

The MPEX survey was iteratively refined through testing in more than 17 universities and colleges during the last four years. The final version of the survey presented here has 34 items and typically takes twenty to thirty minutes to complete. The survey items are designed so that those students with sophisticated expectations will agree with some items and disagree with others. The results of the MPEX survey given in the introductory courses at ten colleges and universities using traditional and research-based curricula are given in chapter 10 of this dissertation. (The curricula and their implementations at the ten schools are described in chapter 8.)
III. CHOOSING THE ITEMS OF THE MPEX SURVEY

A. Cluster Descriptions

The cognitive beliefs that we have referred to as "student expectations" clearly are complex and contain many facets. We decided to focus on six issues or dimensions along which we might categorize student attitudes about learning and physics. Three of these are taken from Hammer's study and we have added three of our own. Building on the work of Perry and Songer and Linn,\textsuperscript{15} Hammer proposed three dimensions along which to classify student beliefs about the nature of learning physics:

1. \textit{Independence} — beliefs about learning physics — whether it means receiving information or involves an active process of reconstructing one's own understanding;

2. \textit{Coherence} — beliefs about the structure of physics knowledge — as a collection of isolated pieces or as a single coherent system;

3. \textit{Concepts} — beliefs about the content of physics knowledge — as formulas or as concepts that underlie the formulas.

In the MPEX survey, we also seek to probe the three additional dimensions described below:

4. \textit{Reality Link} — beliefs about the connection between physics and reality — whether physics is unrelated to experiences outside the classroom or whether it is useful to think about them together;

5. \textit{Math Link} — beliefs about the role of mathematics in learning physics — whether the mathematical formalism is just used to calculate numbers or is used as a way of representing information about physical phenomena;

6. \textit{Effort} — beliefs about the kind of activities and type of work necessary to make sense out of physics — whether they expect to think carefully and evaluate what they are doing based on available materials and feedback or not.

The extreme views associated with each of these variables are given in Table 5-1. We refer to the extreme view that agrees with that of most mature scientists as the “expert” or “favorable” view, and the view that agrees with that of most beginning students as the
“novice” or “unfavorable” view. The survey items that have been selected to probe these six expectation dimensions are given in the right hand column of the table. We

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Favorable</th>
<th>Unfavorable</th>
<th>MPEX Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>independence</td>
<td>takes responsibility for constructing own understanding</td>
<td>takes what is given by authorities (teacher, text) without evaluation</td>
<td>1, 8, 13, 14, 17, 27</td>
</tr>
<tr>
<td>coherence</td>
<td>believes physics needs to be considered as a connected, consistent framework</td>
<td>believes physics can be treated as unrelated facts or &quot;pieces&quot;</td>
<td>12, 15, 16, 21, 29</td>
</tr>
<tr>
<td>concepts</td>
<td>stresses understanding of the underlying ideas and concepts</td>
<td>focuses on memorizing and using formulas</td>
<td>4, 19, 26, 27, 32</td>
</tr>
<tr>
<td>reality link</td>
<td>believes ideas learned in physics are relevant and useful in a wide variety of real contexts</td>
<td>believes ideas learned in physics has little relation to experiences outside the classroom</td>
<td>10, 18, 22, 25</td>
</tr>
<tr>
<td>math link</td>
<td>considers mathematics as a convenient way of representing physical phenomena</td>
<td>views the physics and the math as independent with little relationship between them</td>
<td>2, 6, 8, 16, 20</td>
</tr>
<tr>
<td>effort</td>
<td>makes the effort to use information available and tries to make sense of it</td>
<td>does not attempt to use available information effectively</td>
<td>3, 6, 7, 24, 31</td>
</tr>
</tbody>
</table>
refer to the collection of survey items designed to probe a particular dimension as a “cluster.” Note that there is some overlap, as these dimensions are not independent variables.

B. Survey Items & Responses

The survey has undergone several iterations of development and testing. A copy of the current pre-course versions (version 4.0) of both the scantron and paper & pencil survey forms are included in Appendix C. Note that the pre- and post-course versions of the forms have the same survey items but with different tenses to differentiate between what students expect to do at the beginning of the sequence and what they have done at the end of the course. In the current section, each of the thirty-four survey items are presented and interpreted. The items are listed under their corresponding clusters.

1. Student beliefs about learning physics: The independence cluster

One characteristic of the binary thinker, as reported by Perry and Belenky et al., is the view that knowledge comes from an authoritative source, such as an instructor or a textbook, and it is the responsibility of that authority to convey this knowledge to the student. This is a key element of the transmissionist view of learning, that knowledge is presented to the student who learns it and repeats it back on assignments and exams. The more mature students understand that developing knowledge is a participatory process. They understand that they must actively think about what they are learning to build an understanding of the course material. Hammer classifies these two extreme views as “by authority” and “independent.” The survey items 1, 8, 13, 14, 17, and 27 are designed to probe students’ views along this dimension.
Item 1. All I need to do to understand most of the basic ideas in this course is just read the text, work most of the problems, and/or pay close attention in class.

At first glance, item 1 seems like a fairly innocuous question that all students should agree with. After all this is what instructors expect most students to do to succeed in the course. It is only when one realizes that with the word “just” there is nothing in this statement about understanding the material that the point of this item becomes clear. As students become more independent in building their own understanding, they start to disagree with this item. Disagreeing with this item is a strong indication that the student is moving beyond “binary” thinking.

Item 8. In this course, I do not expect to understand equations in an intuitive sense; they just have to be taken as givens.

All instructors would expect their students to disagree with this item. However, many students attempt to learn physics by memorizing equations and mathematical problem solutions without understanding. Some students do this because they feel this is the most efficient way to learn the material. Others like Hammer’s student Ellen do it because they haven’t been able to understand the material and this is their only way to succeed. However, because many students recognize that they should disagree with this statement, agreement with this statement is a strong indication of learning by authority. Since in this view, concepts are divorced from the equations, this item is also part of the Math Link cluster.

Item 13. My grade in this course is primarily determined by how familiar I am with the material. Insight or creativity has little to do with it.

As Tobias’ student observers commented, many students believe that if you know physics, there is one and only one right way to do things like solving problems,
particularly at the introductory course level. When students believe this, they do not see the value of insight or creativity because they are focusing on learning the “right” answers. They see science as a set of facts laid down by authority rather than as a process to understand physical phenomenon. Students who agree with this item are emphasizing learning the course material rather than trying to understand and think about the course material.

Item 14. *Learning physics is a matter of acquiring knowledge that is specifically located in the laws, principles, and equations given in class and/or in the textbook.*

While learning physics involves learning the laws, principles, and equations, it is more than that. Learning physics also involves building a functional understanding of the laws, principles, and equations including understanding their implications and interconnections. “Independent” students who are actively building their own understanding of the material will disagree with this statement.

Item 17. *Only very few specially qualified people are capable of really understanding physics.*

Many students come into the introductory physics class believing that they cannot do physics but maybe they can learn something about it. One of our goals of instruction is that students learn how to do physics themselves, that is, to make observations, to generalize from what they observe and construct models, and to make and test predictions rather than to just receive physics knowledge. We expect that students who believe that they can do physics to disagree with this statement.

Item 27. *Understanding physics basically means being able to recall something you've read or been shown.*

148
Item 27 is an example of the extreme “by authority” view. We would hope that most students would disagree with this item. Since such a view precludes conceptual understanding, item 27 is also part of the concepts cluster.

2. Student beliefs about the structure of physics knowledge:

The coherence cluster

Most physics faculty feel strongly that students should see physics as a coherent, consistent structure. A major strength of the scientific worldview is its ability to describe many complex phenomena with a few simple laws and principles. Students, like Liza from Hammer’s study in chapter 2, who emphasize science as a collection of facts, fail to see the integrity of the structure, an integrity that is both epistemologically convincing and useful. The lack of a coherent view can cause students many problems, including a failure to notice errors in their reasoning and an inability to evaluate a recalled item through crosschecks. Survey items 12, 15, 16, 21, and 29 have been included in order to probe student views along this dimension.

Item 12. Knowledge in physics consists of many pieces of information each of which applies primarily to a specific situation.

Item 12 reflects the view of students who focus on memorizing information. These students have a great deal to memorize because there is so much material and so many different situations. Their focus is on the equations used in all these different situations, not on the more general equations and principle from which the situation specific equations can be derived. These students tend not to see physics knowledge as a consistent framework and they don’t see the connections and underlying themes in the
course material. A student who sees either the framework, the connections, or the underlying themes should disagree with this statement.

Item 15. In doing a physics problem, if my calculation gives a result that differs significantly from what I expect, I'd have to trust the calculation.

As we saw in the discussion of Hammer’s doctoral study in chapter 2, the type B students were very casual about making and breaking associations between different aspects of their knowledge. Since they don’t expect physics to be coherent or even to really make sense, these students tend to trust their calculation more and change their intuition to suit a particular problem. The type A students saw physics as more coherent and were much more cautious about modifying their understanding. They tend to trust their intuition more than their calculations.

Item 16. The derivations or proofs of equations in class or in the text have little to do with solving problems or with the skills I need to succeed in this course.

The derivations of equations show how the equations are related to the coherent framework. They show where they come from and how they relate to the main principles. A student who agrees with item 16 does not see either the relationship or the coherent structure as useful. While this is usually an indication of the students’ expectations, it can also be an indication of the types of physics knowledge valued in the class. Since derivations are an important part of the relationship between concepts and equations, this item is also part of the math link cluster.

Item 21. If I came up with two different approaches to a problem and they gave different answers, I would not worry about it; I would just choose the answer that seemed most reasonable. (Assume the answer is not in the back of the book.)
Coming up with two different answers using two different approaches indicates that something is seriously wrong with at least one of the solutions and perhaps with the students’ understanding of physics and how to apply it to problems. A student who sees physics as a set of many equations that apply to many different situations and believes that there is only one right solution would assume that the less reasonable answer was produced from using an incorrect equation. A student who sees physics as a coherent and consistent whole will disagree with item 21.

Item 29. A significant problem in this course is being able to memorize all the information I need to know.

A sophisticated student will realize that the large number of different equations and results discussed in a physics text can be structured and organized so that only a small amount of information needs to be memorized and the rest can be easily rebuilt as needed. Item 29 is part of a probe into whether or not students see this structure or are relying on memorizing instead of rebuilding. However, if students are permitted to use a formula sheet or if exams are open book, they may not perceive memorization as a problem. This does not mean that they see the coherence of the material. If extensive information is made available to students during exams, item #29 needs to be interpreted carefully.

3. Student beliefs about the content of physics knowledge: The concepts cluster

The group of items selected for the concepts cluster (items 4, 19, 26, 27, and 32) are intended to probe whether students are viewing physics problems as simply a mathematical calculation or the application of an equation, or whether they are aware of the more fundamental role played by physics concepts and principles in complex problem
Item 4. *Problem solving in physics basically means matching problems with facts or equations and then substituting values to get a number.*

Item 19. *The most crucial thing in solving a physics problem is finding the right equation to use.*

Item 26. *When I solve most exam or homework problems, I explicitly think about the concepts that underlie the problem.*

Item 27. *Understanding physics basically means being able to recall something you've read or been shown.*

Item 32. *To be able to use an equation in a problem (particularly in a problem that I haven’t seen before), I need to know more than what each term in the equation represents.*

The intent of these items ranges from statements of extreme novice views like item 27 to the more sophisticated views expressed in items 19 and 26. Students who disagree with item 27 can at least distinguish between memorizing and understanding. Students who rely heavily on equation manipulation to get through the introductory physics course will agree with item 4. However, it is interesting to compare the results of items 4 and 19. A more experienced student could disagree with item 4 and yet still agree with item 19 either because of or despite the use of the words “most crucial.” As students become more experienced with complex problem solving, the importance of finding the correct equation decreases and more emphasis is placed on a better understanding of the problems and the conceptual issues involved. Another sign of sophistication in problem solving is the explicit use of physics concepts and principles as queried by item 26. Many novice students will write out equations in problem solutions with little or no idea of the relation between the equations and the concepts or principles. These students often apply whatever equation they can find or remember that seems to
have the right quantities to solve the problem. These students tend to disagree with item 32.

4. Student beliefs about the connections between physics and the real world:

   The reality-link cluster

   Although physicists believe that they are learning about the real world when they study physics, the context dependence of cognitive responses (see ref. 5) opens a possible gap between faculty and students. Students may believe that physics is related to the real world in principle, but they may also believe that it has little or no relevance to their personal experience. This can cause problems that are both serious and surprising to faculty. The student who does a calculation of the speed with which a high jumper leaves the ground and comes up with 8000 m/s (as a result of recalling numbers with incorrect units and forgetting to take a square root) may not bother to evaluate that answer and see it as nonsense on the basis of personal experience. When an instructor produces a demonstration that has been “cleaned” of distracting elements such as friction and air resistance, the instructor may see it as displaying a general physical law that is present in the everyday world but that lies “hidden” beneath distracting factors. The student, on the other hand, may believe that the complex apparatus is required to produce the phenomenon, and that it does not occur naturally in the everyday world, or is irrelevant to it. A failure to make a link to experience can lead to problems, not just because physics instructors want students to make strong connections between their real-life experiences and what they learn in the classroom, but because learning tends to be more effective and robust when linked to real and personal experiences.
The four items included as the reality-link cluster (items 10, 18, 22, and 25) do not just probe whether the students believe the laws of physics govern the real world. Rather, these items probe whether the students feel that their personal real-world experience is relevant for their physics course and vice versa. In our interviews, we observed that many students show what we would call, following Hammer, an “apparent reality link.” That is, they believe that the laws of physics govern the behavior of the real world in principle, but that they do not need to consider that fact relevant or necessary to their physics course.

Item 10. *Physical laws have little relation to what I experience in the real world.*

Item 18. *To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed.*

Item 22. *Physics is related to the real world and it sometimes helps to think about the connection, but it is rarely essential for what I have to do in this course.*

Item 25. *Learning physics helps me understand situations in my everyday life.*

Item 10 is fairly obvious and most, but not all, students usually disagree with it at the beginning of the course. Note that the question does not ask if physical laws are related to the real world but if they are related to the students’ experiences in the real world. Even many students with novice views will disagree with this item. As students start thinking about what they are learning, they usually start to see examples of physics in everyday situations and agree with Item 25. More sophisticated students will agree with item 18 and disagree with item 22. While many students see examples of physics in everyday life, only a few use their own experiences to help them understand the physics they are learning or in solving problems. While many students find that thinking about
physics in connection with the real world is useful, only a fraction of them feel that this
helps them in most introductory physics courses.

We expect at least some students to respond as if they believe physics and
physical laws are closely related to their experiences, but this connection is not
something they use to help them learn physics in or out of class. These students are
aware of the connection and some even know that in ideal circumstances they should be
using the link, but they don’t actively integrate their experiences and the physics they are
learning.

5. Student beliefs about the role of mathematics in learning physics:

The math-link cluster

An important component of the calculus-based physics course is the development
of the students' ability to use abstract and mathematical reasoning in describing and
making predictions about the behavior of real physical systems. Expert scientists use
mathematical equations as concise summaries of complex relationships among concepts
and/or measurements. They can often use equations as a framework on which to
construct qualitative arguments. Many introductory students, however, fail to see the
deeper physical relationships present in an equation and instead use the math in a purely
arithmetic sense, i.e. as a way to calculate a numeric answer. When students have this
expectation about equations, there can be a serious gap between what the instructor
intends and what the students infer. For example, a professor may go through extensive
mathematical derivations in class, expecting the students to use the elements of the
derivation to see the structure and sources of the relationships in the equation. The
students, on the other hand, may not grasp what the professor is trying to do and reject it
as irrelevant “theory.” Students who fail to understand the derivation and structure of an
equation may be forced to rely on memorization — an especially fallible procedure if
they are weak in coherence and have no way to check what they recall. The survey
items probing students’ apparent expectations\textsuperscript{21} of the role of mathematics are 2, 6, 8, 16,
and 20.

Item 2. \textit{All I learn from a derivation or proof of a formula is that the formula
obtained is valid and that it is OK to use it in problems.}

Item 6. \textit{I spend a lot of time figuring out and understanding at least some of
the derivations or proofs given either in class or in the text.}

Item 8. \textit{In this course, I do not expect to understand equations in an intuitive
sense; they just have to be taken as givens.}

Item 16. \textit{The derivations or proofs of equations in class or in the text have little
to do with solving problems or with the skills I need to succeed in this
course.}

Item 20. \textit{If I don’t remember a particular equation needed for a problem in an
exam there’s nothing much I can do (legally!) to come up with it.}

Items 2, 6, and 16 deal with the different roles of derivations in student learning.

Many students find looking at derivations done by the professor in class or in the
textbook to be useful, but don’t actually work derivations out themselves. And while
students might find derivations useful for learning physics, some do not see them as
necessary for doing well in the course.

Some students don’t try to understand the equations, they just use them to solve
problems where the variables match the conditions of the problem. Students who use
this approach should agree with item 8. One disadvantage of this approach on exams is
that if you forget the correct equation, you are either stuck or forced to use a different
approach. However, if a student understands the equation as a relationship and
remembers where it comes from and how it connects to the concepts, they could rebuild the forgotten equation. Students who have this ability should disagree with item 20.

6. **Student beliefs about studying physics: The effort cluster**

Many physics lecturers do not expect most of their students to follow what they are doing in lecture during the lecture itself. They expect students will take good notes and figure them out carefully later. Unfortunately, many students do not take good notes and even those who do may rarely look at them. When physics begins to get difficult for students, most instructors expect them to try to figure things out using a variety of techniques — working through the examples in the book, trying additional problems, talking to friends and colleagues, and in general trying to use whatever resources they have available to make sense of the material. Some students, on the other hand, when things get difficult, may be at a loss for what to do. Some students do not have the idea that if they do not see something right away, there are steps they can take that will eventually help them make sense of the topic. An important component of the tools that help build understanding is making the effort to go over the book and the class activities (lecture in the traditional course format). Another important component is the appreciation that one's current understanding might be wrong, and that mistakes made on homework and exams can give guidance in helping to correct one's errors. This dimension is probed by items 3, 6, 7, 24, and 31 on the survey.

- **Item 3.** *I go over my class notes carefully to prepare for tests in this course.*
- **Item 6.** *I spend a lot of time figuring out and understanding at least some of the derivations or proofs given either in class or in the text.*
- **Item 7.** *I read the text in detail and work through many of the examples given there.*
Item 24. *The results of an exam don’t give me any useful guidance to improve my understanding of the course material. All the learning associated with an exam is in the studying I do before it takes place.*

Item 31. *I use the mistakes I make on homework and on exam problems as clues to what I need to do to understand the material better.*

Items 3, 6, and 7 describe activities beyond doing the homework that help students develop a good understanding of the material. We expect that good students will agree with these items. Items 24 and 31 ask if students make the effort to use the graded exams and homework as feedback to debug their understanding of physics or problem solving methods.

7. Other expectation issues

Not all the items in the survey are part of the clusters. Since this is a survey instrument to study student expectations, we have included the following items to probe addition expectation issues:

Item 5. *Learning physics made me change some of my ideas about how the physical world works.*

As we saw in chapters 2 and 4, many students come into introductory physics courses with common-sense beliefs about how the world works that are incompatible with what they learn in the course. Students who develop a good conceptual understanding of physics will need to reconcile what they learn with what they thought they knew about how things work in the physical world. Even students who come into an introductory physics class with a more-physics like view find new applications and subtleties that help them see the world in new ways. These students should agree with this item.
Item 9. The best way for me to learn physics is by solving many problems rather than by carefully analyzing a few in detail.

Item 9 is subtler. As mentioned above, as students become more sophisticated learners, they shift their emphasis from the equations to the concepts and principles. The studies on expert and novice problem solvers by Chi et al.23 (discussed in chapter 2) suggest that these students also begin to classify problems not by their surface features but by the concepts needed to understand the problem and reach a solution. Thus, these students often find solving a few problems in great detail teaches them how to use concepts effectively for problem solving. On the other hand, students who focus more on the equations like to solve many problems so they are better prepared to apply them in many different situations. We would expect that as students become more sophisticated learners, they will disagree with this item.

Item 11. A good understanding of physics is necessary for me to achieve my career goals. A good grade in this course is not enough.

Since most of the students surveyed are not physics majors and few non-majors take physics out of interest, most of the students surveyed are taking the introductory physics because it is required by their program. While many of these students just look at physics as another general education requirement, the more sophisticated students see how understanding physics will be useful in their careers. These students will agree with this item. In the pre-course survey, this item can be used to gauge the distribution of majors in a class. The more engineering and physical science majors in a class, the higher the percentage of student responses agreeing with this item.

Item 23. The main skill I get out of this course is learning how to solve physics problems.
Item 30. The main skill I get out of this course is to learn how to reason logically about the physical world.

Items 23 and 30 are unusual in that both responses are necessary to evaluate student expectations on the course objective. Novice students who emphasize mathematical problem solving will agree more with item 23 than 30. As students become more sophisticated learners, they agree less with item 23 and more with item 30. A comparison of the pre and post course distributions on these two items is an indication of the course goals communicated to the students as well as their expectations.

Item 28. Spending a lot of time (half an hour or more) working on a problem is a waste of time. If I don't make progress quickly, I'd be better off asking someone who knows more than I do.

Most physics instructors recognize that having their students struggle with a problem and working through it on their own helps the students piece their knowledge together and build their confidence. Students who recognize the value of this struggle in building their own understanding should disagree with this item.

Item 33. It is possible to pass this course (get a "C" or better) without understanding physics very well.

This item tell us more about students’ perception of the class rather than the physics expectations of the students. Ideally, understanding physics should be required to pass an introductory physics course. Students who feel this way should disagree with this item. Some false positives (“disagrees”) can be expected from students who confuse familiarity with understanding.

Item 34. Learning physics requires that I substantially rethink, restructure, and reorganize the information that I am given in class and/or in the text.
In science we make observations, test predictions, and evaluate what we find. Often as we learn more, we need to rethink, restructure, or reorganize what we thought we knew when it seems to contradict what we learn. Reflection is an important tool for building a consistent, coherent knowledge framework. Often students in introductory physics will reach a similar point when they learn something new that either contradicts what they already knew or helps them to think about what they know in different ways. Students who reflect and rethink to add what they learn to what they know should agree with this item. A favorable response on this item suggests a sophisticated learner.

8. Additional survey questions

In addition to the 34 survey items, we also collected data with the survey on the students’ background, study habits, and self-appraisal of skills. These background questions ask about the students’ major, whether or not they had physics before, and the students’ math courses in high school and college. Questions on study habits inquire into how much time students spend studying each week, preparing for exams, and working with others. The students are then asked to rate themselves with regard to understanding course materials, useful skills, and mathematical ability. This background data has not been used for the analysis in this dissertation but will be used in later work.

IV. VALIDITY

Any measurement instrument, particularly an attitude measurement like the MPEX survey, needs to be checked for both validity and reliability. A measurement instrument is validated when one can demonstrate that it measures what was intended. Reliability is an indication of to what extent the instrument measurements are free of
unpredictable kinds of error. Reliability is discussed in section VI. In the current section, I discuss the results of two kinds of construct validity procedures. Construct validity determines how well the instrument measures the construct being measured, in this case the students’ cognitive expectations. For our purposes, construct validity is important for verifying that the MPEX survey is measuring cognitive attitudes and beliefs about the course and learning physics. However, since we believe that student expectations play a major role in what students learn in an introductory course, it would be useful to see if the MPEX survey results correlate with other measures of student learning such as grades and the FCI. This correlation analysis will be performed in later work.

A. Surface Validity - Measurements of Calibration Groups

In order to test whether the survey correctly represents elements of the hidden curriculum, it was given to a variety of students and physics instructors. The “expert” response was defined as the response that was given by a majority of experienced physics instructors who have a high concern for educational issues and a high sensitivity to students. Redish, Steinberg, and I conjectured that experts, when asked what answers they would want their students to give, would respond consistently.24

1. The calibration groups

We tested the response of a wide range of respondents by comparing five groups:

Group 1: engineering students entering the calculus-based physics sequence at the University of Maryland,

Group 2: members of the US International Physics Olympics Team
Group 3: high school teachers attending the two-week Dickinson College Summer Seminar on new approaches in physics education

Group 4: university and college teachers attending the two-week Dickinson College Summer Seminar on new approaches in physics education

Group 5: college faculty who are part of a multi-university FIPSE-sponsored project to implement Workshop Physics at their home institutions.

The University of Maryland students (UMD) are a fairly typical diverse group of primarily engineering students at a large research university. The entering class average on the FCI is 51.1% ± 2.4% (standard error). The number of students in the sample is N = 445.

The US International Physics Olympics Team (POT) is a group of high school students selected from applicants throughout the USA. After a two week training session, five are chosen to represent the US in the International Physics Olympics. In 1995 and 1996, this group trained at the University of Maryland in College Park and we took the opportunity to have them complete survey forms. The total number of respondents in this group is N = 56. Although they are not teachers, they have been selected by experts as some of the best high school physics students in the nation. Our hypothesis was that they would prove to be more expert than the average university physics student, but not as expert as our groups of experienced instructors.

The physics instructors who served as our test groups were all visiting Dickinson College. Attendees came from a wide variety of institutions. Many have had considerable experience in teaching, and all of them were sufficiently interested in educational development to attend a workshop. We separated them into three groups. Group 3 — high school teachers (HS) attending a two-week summer seminar (N = 26),
Group 4 — college and university teachers (College) attending the two-week summer seminar (N=56), and Group 5 — college and university teachers implementing Workshop Physics (Expert) in their classroom (N = 19). The teachers in Group 5 were committed to implementing an interactive engagement model of teaching in their classroom. We asked the three groups of instructors to respond with the answer that they would prefer their students to give after instruction. We expected these five groups to show an increasing level of agreement with answers that we preferred.

2. The responses of the calibration groups

The group we expected to be the most sophisticated, the group 5 instructors, agreed strongly as to what were the responses they would like to hear from their students. On all but three items, ~80% or more of this group agreed with a particular position. These three items, numbers 7, 9, and 34, had a strong plurality of agreement, but between \( \frac{1}{4} \) and \( \frac{2}{3} \) of the respondents chose neutral. We define the preferred response of Group 5 as the expert response. We define a response in agreement with the expert response as “favorable” and a response in disagreement with the expert response as “unfavorable”. A list of the favorable responses to the survey items is presented in Table 5-2.

Although the survey itself uses a five point Likert scale (strongly disagree = 1 to strongly agree = 5),\(^{25}\) we have chosen to group the survey responses into three categories: agree, disagree, and neutral. Someone who responds either “agree” or “strongly agree” for a survey item is considered to agree with that item. Someone who responds either “disagree” or “strongly disagree” on a survey item is considered to disagree with that item. Someone who responds “neutral” for or does not answer a
survey item is considered to be neutral with regard to that item. This was done for two reasons. One, it is not clear that the intervals of the Likert scale are the same for every respondent with regard to any particular survey item. For example, one person who agrees with a particular survey item might have the same expectations as another person who strongly agrees with the same item. Because of this, Redish and I feel that the three-point scale does not unduly reduce the resolution of the survey. The second reason is to sharpen the interpretation of the data. Because we are looking for shifts in student expectations, changes from agree to disagree or even agree to neutral are more significant than changes from strongly agree to agree.

Table 5-2. Prevalent responses of our expert group for the MPEX Survey items. Where the respondents did not agree at the >80% level, the item is shown in parentheses and the majority response is shown. The response "A" indicates agree or strongly agree. The response "D" indicates disagree or strongly disagree.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>A</td>
<td>D</td>
<td>A</td>
<td>A</td>
<td>(A)</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>A</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>(D)</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>A</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>A</td>
<td>A</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>(A)</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>A</td>
<td>A</td>
<td>D</td>
<td>D</td>
<td>(A)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5-1. A-D plot for the calibration groups, average of all survey items. The percentage of respondents agreeing with the majority of expert views (favorable) is plotted against the percentage disagreeing the those views (unfavorable).

Table 5-3. Percentages of the calibration groups giving favorable / unfavorable responses on Overall and Cluster MPEX survey.

<table>
<thead>
<tr>
<th>MPEX clusters</th>
<th>Experts</th>
<th>College</th>
<th>HS</th>
<th>POT</th>
<th>UMD Pre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>87 / 6</td>
<td>81 / 10</td>
<td>73 / 15</td>
<td>68 / 18</td>
<td>54 / 23</td>
</tr>
<tr>
<td>Independence</td>
<td>93 / 3</td>
<td>80 / 8</td>
<td>75 / 16</td>
<td>81 / 12</td>
<td>47 / 34</td>
</tr>
<tr>
<td>Coherence</td>
<td>85 / 12</td>
<td>80 / 12</td>
<td>62 / 26</td>
<td>79 / 8</td>
<td>53 / 24</td>
</tr>
<tr>
<td>Concepts</td>
<td>89 / 6</td>
<td>80 / 8</td>
<td>71 / 18</td>
<td>73 / 13</td>
<td>54 / 27</td>
</tr>
<tr>
<td>Reality Link</td>
<td>93 / 3</td>
<td>94 / 4</td>
<td>95 / 2</td>
<td>64 / 20</td>
<td>61 / 14</td>
</tr>
<tr>
<td>Math Link</td>
<td>92 / 4</td>
<td>84 / 9</td>
<td>67 / 21</td>
<td>85 / 8</td>
<td>67 / 22</td>
</tr>
<tr>
<td>Effort</td>
<td>85 / 4</td>
<td>82 / 6</td>
<td>68 / 13</td>
<td>50 / 34</td>
<td>67 / 13</td>
</tr>
</tbody>
</table>
To display our results in a concise and easily interpretable manner, we introduce the *agree-disagree* (A-D) or *Redish plot*. In this plot, the percentage of respondents in each group answering favorably is plotted against the percentage of respondents in each group answering unfavorably. Since the fraction of students agreeing and disagreeing must add up to less than or equal to 100%, all points must lie in the triangle bounded by the corners (0,0), (0,100), (100,0). The distance from the diagonal line is a measure of the number of respondents who answered neutral or chose not to answer. The closer a point is to the upper left corner of the allowed region, the better the group's agreement with the expert response.\(^{26}\)

The results on the overall survey are shown in Fig. 5-1. In this plot, the percentages are averaged over all of the items of the survey, using the preferred responses of calibration group 5 as favorable. The groups' responses are distributed from less to more favorable in the predicted fashion.\(^{27}\)

Although the overall results support the contention that our survey correlates well with an overall sophistication of attitudes towards doing physics, the cluster results show some interesting deviations from the monotonic ordering. These deviations are quite sensible and support the use of clusters as well as overall results. In order to save space and simplify the interpretation of results, the data is presented in Table 5-3. Displayed in this table are the percentages of each group's favorable and unfavorable responses (in the form favorable/unfavorable). The percentage of neutrals and those not answering can be obtained by subtracting the sum of the favorable and unfavorable responses from 100.
From the table we see that most of the fraction of respondents agreeing with the favorable response tends to decrease monotonically from group 1-5 with a few interesting exceptions. The high school teachers (group 3) are farther than their average from the favorable corner in the coherence and math clusters, while the Physics Olympics team is closer to the favorable corner in those categories than their average. These results are plausible if we assume that high school teachers are less concerned with their students forming a coherent and a mathematically sophisticated view of physics than are university teachers. The results also agree with Redish and Steinberg’s personal observations\(^{28}\) that the members of the POT are unusually coherent in their views of physics and exceptionally strong in their mathematical skills.

Note also that the Olympics team results are very far from the favorable corner in the effort cluster. The main discrepancies are in items 3 and 7. These items represent highly traditional measures of effort (reading the textbook, going over one's lecture notes) which we conjecture are not yet part of the normal repertoire of the best and brightest high school physics students before they enter college. Redish, Steinberg, and I also conjecture that most of them will have to learn to make these kinds of efforts as they progress to increasingly sophisticated materials and the level of challenge rises.

This analysis of both the overall responses of the calibration groups and the variations in the ordering confirms that the MPEX survey provides a quantitative measure of characteristics which experts hope and expect their students to have.

### B. Validation with Student Interviews

I conducted more than 120 hours of videotaped student interviews in order to validate that our interpretation of the survey items matched the way they were read and
interpreted by students. I asked students (either individually or in groups of two or three) to describe their interpretations of the statements and to indicate why they responded in the way that they did. In addition, students were asked to give specific examples from class to justify their responses. The protocols for these interviews (the MPEX Survey Protocol and the Open MPEX Protocol) are discussed in chapter 7.

From these interviews, we found that students are not always consistent with their responses to what appear to us to be similar questions and situations. We feel that this does not represent a failure of the survey, but properly matches these students’ ill-defined understanding of the nature of physics. One reason for this was described by Hammer (see the section of expectations in chapter 2). He observed that some students in his study believed that professional physicists operated under the favorable conditions, but that it sufficed for these students to behave in the unfavorable fashion for the purposes of the introductory course.\(^{29}\) We refer to this type of characteristic as an “apparent” expectation. This is only one aspect of the complex nature of human cognition. We must also be careful not to assume that a student exists in one extreme state or another. A student's attitude may be modified by an additional attitude, as in Hammer's observations, or even exist simultaneously in both extremes, depending on the situation that triggers the response.\(^{30}\) This is one reason why considerable care must be taken in applying the results of a limited probe such as our survey to a single student.

We are also aware that students’ self-reported perceptions may not match the way they actually behave. However, the interviews suggest that if a student's self-perception of the learning characteristics described in Table 5-1 differs from the way that student actually functions, the self-perception has a strong tendency to be closer to the
side chosen by experts. We therefore feel that while survey results for an individual student may be misleading, survey results of an entire classroom might understate unfavorable student characteristics.

Because of the length of this dissertation, I have included in Appendix F selected interview responses for four survey items from nine of the nearly 100 students I have interviewed. The responses represent all the interviews conducted during a site visit at Nebraska Wesleyan University (NWU) two weeks before the end of the second semester. I have selected this set for two reasons. First, these were the only MPEX interviews conducted with a good sample from a sizable calculus-based class that used the most current version (version 4.0) of the survey. Second, two additional controls were added to address concerns raised in the earlier interviews.

- To prevent the interview itself from unknowingly influencing the student responses, the students completed the survey before starting the interview (within 24 hours of the interview).

- This student sample was representative of the entire class. At least three students were interviewed from the top third, middle third, and bottom third of the class. The students were rated by the instructor based on their overall grade at the time of the interviews.

The class was taught with the Workshop Physics curriculum. A description of the curriculum and details of the implementation at NWU can be found in chapter 8.

Below, I discuss student responses from the four MPEX survey items with regard to how students are interpreting the survey items and whether their answers make sense. (A detailed analysis of the interviews in terms of the expectations of these students for evaluation of student learning and the curriculum is presented in chapter 10.) I selected items 2, 6, 14, and 22 because they are a good cross section of both the
expectation issues and the difficulties in interpreting the survey data. (Note that all interviewed student names used in this dissertation are code names.)

Items 2 and 6 are two of the items that look at students’ perception of derivations. Item 2 asks students if derivations are good for anything beyond demonstrating the validity of an equation. Item 6 asks the students if they do or go over derivations themselves.

All nine students’ responses indicated they interpreted item 2 correctly. For example:

John: *I put "disagree (favorable response)." I'm not a lover of derivations, by any means; but I think that it can tell you a lot about linking two concepts together. It also creates the idea that science can be a unified thing, and that previous information is applicable to new information. And so, therefore, I think that it is — it's useful.*

Leb: *I said disagree, (favorable response) because, once again, I think it helps to know where the formula came from and then to — helps you to know why it works and why you can apply it that way.*

However, some of the students’ responses are illustrative of the complexities involved in student expectations. For example, two of the student responses were very clearly indicative of an apparent expectation, although both students gave an unfavorable response to item 2.

Amy: *I do agree with that (unfavorable response). Okay. … [Can you elaborate a little bit on that? Why you agree with that statement.] When you get into the beginning, you’re introduced some kind of equation. See, Dr. A won’t give you the equation until he derives it. ... But if — You know, if he got up there and was — I don’t know. I guess I probably would be negative towards it. You can use that proof, no matter what, anyway; because once you get experience with any kind of proof at all, you know you can plug and chug. And I mean that's fact. But what I get out of watching someone derive a derivation or a proof is it gets me closer to being that type of person who'll be able to do it, myself. I'm not, by nature, somebody who does that. I think Dr. A is and always has been.*
Krystal: *I said I agree* (unfavorable response), *because I'm poor in mathematics, but not necessarily for everybody.* I mean, those students — or a lot of students, I'm sure, would understand what they got and know where everything came from. But, again, I just — since I haven't had much math, usually, when he does all that on the chalkboard or on the overhead screen, it's Greek to me. I don't know where he got it from. I just write the final thing down and, you know, that's what I need.

Amy recognizes that watching the professor model derivations is useful, but she doesn’t see derivations as useful in this class or see herself as being able to do them. Krystal, sees derivations as useful for people who understand the mathematics but not her.

One thing that comes out clearly from the favorable responses is that the derivations are helping some students see the relationship between the concepts and the formulas. This is clearly seen in John’s response above.

All but one of the eight student responses to item 6 are consistent with the intended interpretation. For example:

Charlie: *And I agreed with this* (favorable response). *Whenever I -- If there’s a formula presented in the text,* or Dr. Wehrbein presents one on the board, I always try and understand where it came from. And if I don't right offhand -- like, if he writes something and goes through it too fast, I'll try and take all the notes I can. And later, I'll go over it again and see if I can make it all make sense in my head. And it's only at that point that I feel like I understand it.

Krystal: *I said three, “neutral.”* Somewhat, again, because I’m a poor math student, so -- and it's also a phobia. Because when I see a derivation, I usually just turn the page and don't try to thumb through it. [Okay. So, that tells me why you didn't agree with it. Why didn't you disagree with it?] Well, because I also ... *I mean I don't always just flip the page, because I don't give up that easily. I usually will try and get partially through it. But, then, if I get lost, I usually give up then.*

John: *“Disagree* (unfavorable response).* I don't spend too much time trying to figure them out. If you do it once, to show you where the formula came from. I know where it came from. I understand the concepts linking the two together. I usually don't try to go through it mathematically, because then it becomes more of a math exercise.*
It is interesting to note that while 5 of the nine students think derivations are generally useful for learning physics, three of these five students including John do not regularly work out derivations on their own. Only Kim’s interpretation of item 6 is questionable since she spends a lot of time going over derivations with the instructor outside of class. In addition, Kim’s response may indicate a difficulty that appears regularly in a small fraction of the students interviewed, usually in an interview near the beginning of the sequence. These students seem to confuse derivations with derivatives. Also note that although Krystal does do derivations but she does not spend a lot of time on them which is consistent with a neutral response.

Item 14 deals directly with the issue of what does it mean to learn physics. As we saw earlier, item 14 looks like a statement that students should agree with until you think about what you really want students to learn. Students who disagree with this item recognize their own role in building an understanding of physics that makes sense to them. For example,

Hannah: *I think I should disagree with this* (favorable response), *because it's more than just acquiring knowledge. You also have to put it all together. It just can’t be a bunch of facts*

Most students who agree with this item see learning physics as memorizing and applying the facts and formulas from lecture and the textbook. All nine students were asked this question. Of the nine, only Krystal and Ramsey misinterpreted the question although Ramsey changed his mind when asked to reconsider.

Both Krystal and Ramsey initially use the laboratory component of the course in their response and disagreed with item 14. For example, Hannah responded:
I said I disagree, because you need to be able to set up the experiment, too, which is something that isn't really given to you anywhere. Like -- Like you said, that was something that I had to learn first of all, where some people didn't. I don't know. Some people who really had a great high school curriculum, or just this natural instinct at how to set up circuits; but I didn't know how. So first I had to be taught that. Then I could continue to learn.

The fact that both Ramsey and Krystal mention the laboratory as an important part of learning in the courses is not unusual since the students in a Workshop Physics class spend at least three quarters of their time on lab activities. Even so, this interpretation was unusual and not seen in other interviews, even at the other two schools using the Workshop Physics curriculum.

John and Leb gave neutral responses to item 14 which makes sense in light of their explanations which show that they have mixed expectations. This is particularly clear in John’s response,

John: Three, "neutral." Again, I think it depends on the type of person you are. Learning physics as a whole – really learning physics – is more than just ... principles, equations, laws. It's ... in understanding and incorporating all this. But to pass a course, learning physics, I think that's all you really need to get by.

John sees that there is more to learning physics than just the laws, principles and equations but he also sees that this view is not essential in this class. Despite the obviousness of item 14, almost all students who disagreed with this item in interviews had strong constructivist expectations. However, some constructivist students who express mixed expectations do not disagree with this statement like Leb and John.

The student responses to item 22 at Nebraska Wesleyan were unusually favorable. (The MPEX reality link survey responses overall for Nebraska Wesleyan were also unusually favorable. Please see the discussion in the workshop physics site
visits section of chapter 10 to see why.) All the student responses indicate that the students are interpreting item 22 correctly. For example,

Charlie: *And I disagree with this one* (favorable response). *I think any time you can relate it to the real world – the world around you, you can understand it a little bit better, because it familiarizes it with yourself a little bit more. And it seems to apply a little bit more, becomes a little more important.*

Several of the students like Ramsey emphasized the role of laboratory activities in connecting physics to the real world.

Ramsey: *And I marked "disagree" for the reasons I stated before* (favorable response) – *that we're exposed to experiments or demonstrations that you can find in the real world. So I think that learning about those things are essential for what we're doing.*

Only John’s response was questionable. His response is almost a false unfavorable since he sees the connection as beneficial, though not essential. However, this response and its categorization is consistent with his overall expectation profile discussed in chapter 10.

The interview results demonstrate that while there are occasional misinterpretations of survey items, the effect is small and tends to overstate the number of favorable student expectation responses.

VI. UNCERTAINTY & THE STATISTICS OF SHIFTS

Every finite set of data contains fluctuations, which have no real significance but arise from the variability of a particular sample. In this dissertation, my research questions involve comparisons of groups – experts and novices, novice students at different institutions, and students at the beginning and end of their first semester of physics. In order to compare these groups, we compare their averaged responses (agree
vs. neutral vs. disagree). In order for us to understand whether two responses are significantly different, we have to have some model of the random variable in our sample.

Our interviews, our intuitions, and many discussions in the cognitive literature suggest that a human attitude is a highly complex object. As we noted above, some students gave clear evidence in interviews of being in two contradictory states at the same time. What this implies is that the random variable we should be averaging is itself a probability, rather than a set of well-defined values. Unfortunately, the average of probabilities may depend significantly on the structure of the constraints and parameterization of the probabilities, as is well known from quantum statistics. Since detailed models of student attitudes do not yet exist, we will estimate our shift significance by using a cruder model.

Let us assume that a class is drawn from a very large homogeneous group of students and that in the large population, a percentage $p_0$ of responses to an item or cluster will be favorable and a percentage $q_0$ will be unfavorable with $p_0 + q_0 \approx 1$. (For now, we will ignore the possibility of neutral responses.) In a finite sample of $n$ students, we want to know what is the probability of finding $n_1$ favorable and $n_2$ unfavorable responses with $n_1 + n_2 \approx 1$. Using the Gaussian approximation to the binomial distribution, we get that the probability of finding fractions $p = n_1/n$ and $q = n_2/n$ is

$$P(p) = Ae^{-\frac{(p-p_0)^2}{2\sigma^2}}$$

where $A$ is a normalization constant and the standard deviation

$$\sigma = \sqrt{\frac{p_0q_0}{n}}$$
For this distribution, the probability that a sample will have a mean that falls within 1\(\sigma\) of the true mean, \(p_0\), is 0.684 and the probability that a sample will fall within 2\(\sigma\) of the true mean is 0.954.

Since the fraction of neutral responses tends to be small, and since the binomial model is crude for this set of data, we treat our trinomial data as if it were approximately binomial by renormalizing the observed \(p\) and \(q\) into \(p' = \frac{p}{p+q}\) and \(q' = \frac{q}{p+q}\). We consider a difference or shift in means to be significant if it is at less than the 5% probability level, that is, if the difference or shift is greater than twice \(\sigma\) (where \(\sigma = \sqrt{\frac{pq}{n}}\)). Because of the crudeness of this model, we consider differences of 2\(\sigma\) to be significant. For example, at values of \(p = 60\%, q = 20\%\) for \(N = 450\), we get \(\sigma \sim 2\%\). We would therefore consider a 4\% shift to be significant for the University of Maryland engineering students. For \(N = 50\), those values of \(p\) and \(q\) give \(\sigma \sim 6\%\). We therefore consider a 12\% shift to be significant for the Physics Olympics team and the college instructors attending the Dickinson Summer Seminar. Note that for a given sample size, \(\sigma\) doesn’t change much over the typical values of \(p\) and \(q\) seen in Tables 10-3 and 10-4.

V. RELIABILITY

Reliability tests are measures of the random errors inherent in instrument measurement. In this section, I will describe the results of three types of reliability measurements: factor analysis, Cronbach alpha, and reproducibility of results. Since we have created clusters that are subsets of overall student expectations, I have used factor analysis to see if these clusters correspond to structures reflected in the student data.
The Cronbach alpha coefficient is a standard test of internal consistency, commonly used for Likert-scale attitude surveys.\textsuperscript{33}

As any physics instructor preaches, the ultimate mark of a good measurement is that it is reproducible. In this case, this means reproducibility of the measurements of a class rather than the measurements of individual students. The reader is reminded that the MPEX survey was not designed to be used a diagnostic instrument for individual students, but rather to measure the distribution of student expectations in a class. However, reproducing the MPEX survey results from the same class sample twice presents at least two major difficulties. The first difficulty is the logistics in giving the survey twice to a class over a period of time long enough that the students do not remember how they responded before, but not so long that the student attitudes have changed. The second difficulty is that many students have mixed expectations, i.e. hold contradictory expectations simultaneously. These students can be thought of as existing in an expectation superposition state similar to the model of fragmented students’ knowledge structures discussed in chapter 2. This implies that the measured expectation response can change in some circumstances. Unfortunately there are not enough survey items to accurately discriminate among various mixed expectation states.

Instead of trying to compare two measurements from the same set of students, I compared results from pre-course surveys of multiple classes of the same course from the same school assuming that the incoming students’ expectations do not change significantly from one year to the next. The survey results from a particular course at a particular school would then be considered reproducible if the distribution of the means
of the survey results were consistent with the standard deviation \( \sigma \) given by the binomial probability theory in the previous section.

A. Factors & Clusters

Our seven MPEX clusters were written to make sense of students’ epistemology and learning beliefs in way that could be used by an instructor to monitor different aspects of the hidden curriculum in the introductory physics course. The clusters are based on the previous work of Hammer and others as well as our own observations. While Hammer found that students’ expectations could be categorized by a researcher, he did not believe that his categories necessarily represented how expectations were structured in the students’ mind. Because so little is known about the cognitive expectations of undergraduate students in introductory calculus based-physics classes, there is little reason to expect that our seven dimensions represent dimensions of expectations in the students’ minds.

The situation is analogous to the strain on a cubic crystal.\(^{34}\) Consider the two dimensional case of a hexagonal piece of crystal as shown in Figure 5-2 below. Suppose 3 clamps are placed on the crystal so that external forces are exerted on all six sides. Although the internal structure of the crystal is easily described using two basis vectors, the external forces are most easily described in terms of those being exerted on the three pairs of parallel sides. The problem is most easily considered as two coordinate axes or three coordinate axes depending on your perspective and what

---

Figure 5-2. Two-dimensional view of hexagonal piece of cubic crystal.
you are trying to do. If I want to study how the overall crystal reacts to external forces, then I would want to use the (overcomplete) coordinate system that corresponds to the symmetry axes of the hexagonal shape of the crystal. If I want to study how individual crystal cells react to forces, then I would want to use a coordinate system based on the crystal structure.

Because so little is known about student expectations, the situation is like looking at the hexagonal block of cubic crystal but the internal structure is unknown. In making our measurements, we define our coordinate system or clusters along dimensions we can observe easily and which are easily related to external influences. In the crystal analogy, this coordinate system also makes it easier to relate the changes in the measurement of the crystal to the external forces being applied.

In a similar way, the MPEX clusters are easily related to various aspects and goals of instruction such as emphasis on linking physics to everyday life or on the coherence of physics. Like the external forces on the crystal, the seven MPEX expectation dimensions are interrelated. In addition, we can try to diagonalize the matrix obtained from our external measurements to learn more about the internal structure, i.e. how student expectations are structured in the student’s mind.

One way to determine how student expectations are structured is to use factor analysis to look for underlying structure in the data. Factor analysis is a statistical method of using correlations to reduce a large number of variables to a smaller number of factors that might help to explain the data more easily. The procedure is similar to diagonalizing a matrix and finding the eigenvalues and eigenvectors.
However, to properly interpret the results of a factor analysis one must be aware of the assumptions and limitations of this type of analysis. Factor analysis assumes that the test items in question are linearly related to a set of uncorrelated, i.e. orthogonal and independent, factors. While factor analysis is a powerful technique for looking for interrelationships in a data set, the resulting factors are purely mathematical constructs that may not have any real meaning if the linear relationship described above does not exist. Also, as the linear relationship may not be unique, one must have additional justification of the factors before trusting the results of a factor analysis.

It is common practice in the constructions of tests and surveys that use sub-scales like our clusters to use factor analysis to see if these sub-scales are reflected in the data. For a factor analysis of the MPEX survey data, I used my largest single sample of data, pre-course surveys from the University of Minnesota calculus-based introductory physics sequence. A course description can be found in chapter 9. I performed a principal-component analysis with varimax rotation using the SPSS statistics application. A plot of the eigenvalues vs. number of factors shown in Figure 5-3 indicates that perhaps three, four, or five factors account for significant fractions of the variance. As can be seen in table 5-4, the percentage of variance accounted for by the fourth factor is marginal. Note that 3 factors would account for 29% of the total variance; four would account for 33%. For a typical survey instrument, the factors usually account for roughly 30% of the variance.

An analysis was run on the data for three, four, and five factors. The full results are shown in Appendix H. A summary of the results is shown in Table 5-5. Note that
Figure 5-3. Scree plot: the # of Eigenvectors vs. the # of factors. The Scree plot is used to estimate how many factors should be extracted. The factor number at the knee should be within ±1 of the correct number of factors. This graph suggests there are 4 ± 1 factors.
Table 5-4. Percentage of variance associated with the extracted factors. Ideally one wants to maximize the variance with the fewest factors.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>Pct. of Variance</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.46</td>
<td>16.1</td>
<td>16.1</td>
</tr>
<tr>
<td>2</td>
<td>2.75</td>
<td>8.1</td>
<td>24.1</td>
</tr>
<tr>
<td>3</td>
<td>1.70</td>
<td>5.0</td>
<td>29.2</td>
</tr>
<tr>
<td>4</td>
<td>1.45</td>
<td>4.3</td>
<td>33.4</td>
</tr>
<tr>
<td>5</td>
<td>1.33</td>
<td>3.9</td>
<td>37.3</td>
</tr>
<tr>
<td>6</td>
<td>1.26</td>
<td>3.6</td>
<td>41.0</td>
</tr>
</tbody>
</table>
Table 5-5. Results of factor analysis for three-, four-, and five-factor extraction.\textsuperscript{38}

<table>
<thead>
<tr>
<th>MPEX items</th>
<th>3 factor extraction</th>
<th>4 factor extraction</th>
<th>5 factor extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor loading</td>
<td>Factor loading</td>
<td>Factor loading</td>
</tr>
<tr>
<td>25</td>
<td>-0.66</td>
<td>25</td>
<td>0.71</td>
</tr>
<tr>
<td>18</td>
<td>-0.59</td>
<td>18</td>
<td>0.65</td>
</tr>
<tr>
<td>17</td>
<td>0.58</td>
<td>10</td>
<td>-0.60</td>
</tr>
<tr>
<td>10</td>
<td>0.57</td>
<td>30</td>
<td>0.58</td>
</tr>
<tr>
<td>20</td>
<td>0.52</td>
<td>31</td>
<td>0.47</td>
</tr>
<tr>
<td>8</td>
<td>0.52</td>
<td>22</td>
<td>-0.42</td>
</tr>
<tr>
<td>30</td>
<td>-0.51</td>
<td>11</td>
<td>0.38</td>
</tr>
<tr>
<td>31</td>
<td>-0.51</td>
<td>5</td>
<td>0.36</td>
</tr>
<tr>
<td>15</td>
<td>0.47</td>
<td>24</td>
<td>-0.29</td>
</tr>
<tr>
<td>22</td>
<td>0.47</td>
<td>26</td>
<td>0.27</td>
</tr>
<tr>
<td>29</td>
<td>0.40</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>11</td>
<td>-0.40</td>
<td>8</td>
<td>0.61</td>
</tr>
<tr>
<td>5</td>
<td>-0.36</td>
<td>16</td>
<td>0.60</td>
</tr>
<tr>
<td>16</td>
<td>0.35</td>
<td>20</td>
<td>0.58</td>
</tr>
<tr>
<td>1</td>
<td>-0.35</td>
<td>2</td>
<td>0.57</td>
</tr>
<tr>
<td>26</td>
<td>-0.34</td>
<td>15</td>
<td>0.52</td>
</tr>
<tr>
<td>24</td>
<td>0.29</td>
<td>17</td>
<td>0.49</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>21</td>
<td>0.42</td>
</tr>
<tr>
<td>19</td>
<td>0.64</td>
<td>9</td>
<td>0.38</td>
</tr>
<tr>
<td>4</td>
<td>0.56</td>
<td>28</td>
<td>0.38</td>
</tr>
<tr>
<td>14</td>
<td>0.52</td>
<td>29</td>
<td>0.36</td>
</tr>
<tr>
<td>27</td>
<td>0.52</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>23</td>
<td>0.52</td>
<td>4</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>0.51</td>
<td>14</td>
<td>0.61</td>
</tr>
<tr>
<td>13</td>
<td>0.45</td>
<td>19</td>
<td>0.58</td>
</tr>
<tr>
<td>12</td>
<td>0.40</td>
<td>23</td>
<td>0.56</td>
</tr>
<tr>
<td>28</td>
<td>0.35</td>
<td>27</td>
<td>0.49</td>
</tr>
<tr>
<td>21</td>
<td>0.31</td>
<td>1</td>
<td>0.40</td>
</tr>
<tr>
<td>9</td>
<td>0.24</td>
<td>13</td>
<td>0.39</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>0.75</td>
<td>6</td>
<td>0.75</td>
</tr>
<tr>
<td>7</td>
<td>0.59</td>
<td>7</td>
<td>0.58</td>
</tr>
<tr>
<td>34</td>
<td>0.49</td>
<td>34</td>
<td>0.49</td>
</tr>
<tr>
<td>3</td>
<td>0.47</td>
<td>3</td>
<td>0.46</td>
</tr>
<tr>
<td>32</td>
<td>0.44</td>
<td>32</td>
<td>0.44</td>
</tr>
<tr>
<td>33</td>
<td>-0.28</td>
<td>12</td>
<td>0.38</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>33</td>
<td>-0.27</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
four of the factors in the results from four factor extraction and five factor extraction have substantial overlap. Phenomenographic\textsuperscript{39} examination of the survey items corresponding to the four-factor case by Wittmann,\textsuperscript{40} Redish, and the author found themes in the first four factors:

1. The items in this factor are all related to student beliefs about building an intuitive conceptual understanding to learn physics.
2. The items in this factor all deal with how students think about the material they learn in introductory physics.
3. The items in this factor concern whether physics is mainly problem solving or building understanding.
4. The items in this factor ask about what students do to learn physics.

Not surprisingly, when I compared the factors and the clusters, I found that the four factors and the seven clusters do not correspond exactly. A comparison of the factors and the clusters is shown in Table 5-6. Note that in the four-factor table, several items from each cluster correspond to a single factor. The split in the effort cluster represents differences in students’ perceptions. This split suggests that students see going over notes, derivations, and the text as different from going over graded homework and exams.

The fact that the structures of the factors and clusters do not correspond exactly does not invalidate either. Going back to the example of pairs of vertical forces applied to the cut crystal. The clusters correspond to the force pairs; the factors correspond to the crystal’s internal structure. The two may be related, but they are two different ways of looking at the same situation. And while it is reasonably straightforward to design curriculum to deal with the students difficulties indicated by the clusters, designing curriculum in terms of the factors would be more difficult. Also, as with the case of the
Table 5-6. Comparison of factors and clusters

### 4 Factor extraction:

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>No Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>1,13,14,27</td>
<td>8,17</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Coherence</td>
<td>—</td>
<td>15,16,21,29</td>
<td>—</td>
<td>—</td>
<td>12</td>
</tr>
<tr>
<td>Concepts</td>
<td>4,19,27</td>
<td>—</td>
<td>26</td>
<td>32</td>
<td>—</td>
</tr>
<tr>
<td>Reality Link</td>
<td>—</td>
<td>—</td>
<td>10,18,22,25</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Math Link</td>
<td>—</td>
<td>2,8,16,20</td>
<td>—</td>
<td>6</td>
<td>—</td>
</tr>
<tr>
<td>Effort</td>
<td>—</td>
<td>—</td>
<td>24,31</td>
<td>3,6,7</td>
<td>—</td>
</tr>
<tr>
<td>No Cluster</td>
<td>23</td>
<td>9,28</td>
<td>5,11,30</td>
<td>33,34</td>
<td>—</td>
</tr>
</tbody>
</table>

### 3 Factor extraction:

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>No Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>13,14,27</td>
<td>1,8,17</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Coherence</td>
<td>12,21</td>
<td>15,16,29</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Concepts</td>
<td>4,19,27</td>
<td>26</td>
<td>32</td>
<td>—</td>
</tr>
<tr>
<td>Reality Link</td>
<td>—</td>
<td>10,18,22,25</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Math Link</td>
<td>2</td>
<td>8,16,20</td>
<td>6</td>
<td>—</td>
</tr>
<tr>
<td>Effort</td>
<td>—</td>
<td>24,31</td>
<td>3,6,7</td>
<td>—</td>
</tr>
<tr>
<td>No Cluster</td>
<td>9,23,28</td>
<td>5,11,30</td>
<td>33,34</td>
<td>—</td>
</tr>
</tbody>
</table>
FCI factor analysis discussed in the last chapter, dimensions of expectations are not completely independent. They are very much interrelated. Furthermore, as discussed earlier, just as many students can hold two contradictory conceptual ideas in their minds simultaneously, they can also hold mixed expectations with regards to the physics class. This analysis of the structure of students’ cognitive expectations and their relation to the survey and the clusters is suggestive, but at this point our analysis is still preliminary.

B. Cronbach Alpha

One estimate of the reliability of a diagnostic or survey test instrument is to look at the internal consistency of the items. As mentioned earlier, the Cronbach alpha coefficient is a standard method for estimating the internal consistency of a test where the items are scored across a range of values like a Likert scale.\(^{41}\) It measures how well the responses to the items in a test or a subset of questions in a test (items in a cluster or factor for example) contribute to measurements of a single construct or issue like students’ understanding of Newtonian force, students general attitudes towards science, or students’ epistemological beliefs with regards to science. The Cronbach alpha coefficient is commonly used to estimate the reliability of both the total test score and/or the scores from subsets of test items like the clusters or factors discussed above. Tests or subtests having an alpha coefficient \(\geq 0.70\) are considered to be reliable for group measurements.\(^{42}\)

The coefficient is also used to determine if it is more appropriate to evaluate test measurements in terms of the overall test score or the subset scores.\(^{43}\) If the alpha coefficient for the overall score is smaller than the coefficients for the subtests, the test
items within the subtests correlate more strongly with the subtests than with the overall score. This is an indication that the test is measuring more than one variable or issue. In this case it is not appropriate to construct an overall score and the test results should only be discussed in terms of the subtest scores.

The Cronbach alpha coefficient is based on the idea that the responses to a group of questions addressing a single issue should correlate strongly with one another. The formula for calculating the Cronbach alpha coefficient\(^{44}\) is

\[
\alpha = \left( \frac{k}{k - 1} \right) \left( 1 - \frac{\sum s_i^2}{s_t^2} \right)
\]

where

- \(k\) = the number of test items,
- \(\Sigma\) = the sum over all test items,
- \(s_t^2\) = the variance\(^{45}\) of the total test scores, and
- \(s_i^2\) = the variance of the \(i\)th item on the test.

If the responses to the test items are substantially intercorrelated, then \(s_t^2\) will be considerably larger than when they are not intercorrelated. However, if the responses to the items are uncorrelated, the variance for each item is independent. This means the summed variance term is unaffected by the correlation of the test item responses. Thus alpha approaches zero when the item responses are not intercorrelated and becomes larger as the item responses become more and more correlated. If the standard deviation of the test items is one, the Cronbach alpha reflects the average inter-item correlation, i.e. how well the test data items inter-correlate with one another and the overall score. If the standard deviation of the items is not one, then the Cronbach alpha reflects the average inter-item covariance.\(^{46}\) Note that since the average of the standard deviations
of the MPEX survey items for the data set used to calculate the Cronbach alpha coefficient is $0.93 \pm 0.10$, the result can be interpreted as the inter-item correlation.

Note that negative inter-item correlations can yield a reduced or negative alpha coefficient. So, data from surveys like the MPEX that have a mixture of negative and positive items (items where people with a particular expectation should disagree in some cases and agree in others) need to be adjusted so that all the inter-item correlations are positive. In this case, the MPEX data used in the factor analysis described above was transformed so that 5 was the extreme favorable response and 1 was the extreme unfavorable response for all items. Items where the extreme favorable response had been 1 were transformed by the equation $x' = 6 - x$. The full five point Likert scale was used for this analysis. The SPSS program was used to calculate the alpha coefficients for the overall survey score, the clusters, and the factors. The results are shown in Table 5-7. In some cases, the alpha coefficient for a group of items could be increased by removing some items from the grouping in question. The maximum obtainable alpha’s from the selected items are also shown in Table 5-7.

The table shows that while the clusters and factors have coefficients ranging from 0.47 to 0.78, the overall survey has an alpha coefficient of 0.81. This indicates that the overall survey score is estimated to be a more reliable measurement of expectations than the clusters or factors. Since its alpha coefficient is $> 0.70$, the overall MPEX score is therefore a useful and reliable measure. (Item groupings with alpha coefficients of $\geq 0.70$ are considered to be well intercorrelated and reliable.)
Table 5-7. Cronbach Alpha for the overall, cluster, and factor MPEX results

i.) Cronbach Alpha for overall MPEX survey result = 0.806

Number of students = 417   Number of survey items = 34
Fall 1995 University of Minnesota pre course data

ii.) Cronbach Alpha for the n = 3 and n = 4 factors using the same data

3 Factor Extraction

<table>
<thead>
<tr>
<th>Factor</th>
<th>Items</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1, 5, 8, 10, 11, 15, 16, 17, 18, 20, 22, 24, 25, 26, 29, 30, &amp; 31</td>
<td>0.78</td>
</tr>
<tr>
<td>2</td>
<td>2, 4, 9, 12, 13, 14, 19, 21, 23, 27, &amp; 28</td>
<td>0.69</td>
</tr>
<tr>
<td>3</td>
<td>3, 6, 7, 32, 33, &amp; 34</td>
<td>0.54</td>
</tr>
</tbody>
</table>

4 Factor Extraction

<table>
<thead>
<tr>
<th>Factor</th>
<th>Items</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5, 10, 11, 18, 22, 24, 25, 26, 30, &amp; 31</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>2, 8, 9, 15, 16, 17, 20, 21, 28, &amp; 29</td>
<td>0.72</td>
</tr>
<tr>
<td>3</td>
<td>1, 4, 13, 14, 19, 23, &amp; 27</td>
<td>0.65</td>
</tr>
<tr>
<td>4</td>
<td>3, 6, 7, 32, 33, &amp; 34</td>
<td>0.54</td>
</tr>
</tbody>
</table>

iii.) Cronbach Alpha for the MPEX clusters using the same data

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Items</th>
<th>Alpha</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>1, 8, 13, 14, 17, 27</td>
<td>0.48</td>
<td>8, 13, 14, 17, 27</td>
</tr>
<tr>
<td>Coherence</td>
<td>12, 15, 16, 21, 29</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>Concepts</td>
<td>4, 19, 26, 27, 32</td>
<td>0.49</td>
<td>4, 19, 27</td>
</tr>
<tr>
<td>Reality Link</td>
<td>10, 18, 22, 25</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>Math Link</td>
<td>2, 6, 8, 16, 20</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>Effort</td>
<td>3, 6, 7, 24, 31</td>
<td>0.47</td>
<td>3, 6, 7, 31</td>
</tr>
</tbody>
</table>
Still, this does not mean that MPEX clusters with alpha coefficient < 0.70 are unreliable. The Cronbach alpha analysis assumes that all the items in a grouping are measuring the same construct in a similar way and therefore all items should correlate. Unfortunately, this is not consistent with the overall survey design, particularly in the clusters. The survey was designed so that as a student’s expectations increase in sophistication, the student will give more favorable responses for the items in a given cluster. The responses of students in a mixed state for some items in a given cluster are expected to negatively correlate. For example, some items are designed so that only students with extremely favorable expectations will respond favorably to these items. Because there are only a few items like this in the survey, the effect of these items on the overall MPEX result is small. The effect on the clusters is greater because of the smaller number of items in the clusters.

Also, as we discussed earlier, the clusters were intended to give instructors a way of interpreting the results in terms of student response to instruction. That does not mean that the clusters represent the way the students associate the issues associated with the items. For example, while student response to the items in the effort cluster that concern studying for the class are correlated, the items regarding learning from mistakes on exams and homeworks do not correlate with these items. These items correlate better with items dealing with how students construct their understanding.

However, the construction of the factors is based on the intercorrelations of the test items. Therefore, factors like factors 4 and 5 with alpha coefficients < 0.65 should not be considered reliable and should not be used for interpreting test results.
C. Repeatability

For any experimental measurement in physics, a measurement is not considered reliable unless it is reproducible. Since the MPEX survey was designed to measure the distribution of student expectation in a class, one method for evaluating repeatability is to give the survey to a class twice over a period of a few weeks. While the data I took at the end of one quarter and the beginning of the next quarter could be used for this purpose, there is a problem. Between the two measurements, the students study and prepare for the final exam. This can have a strong effect on student expectations. In addition, we have noticed another effect that makes a pre/post comparison of this kind questionable. First, students tend to respond more favorably to the MPEX survey items at the beginning of the second course in the sequence than at the end of first. This effect is particularly noticeable in the effort cluster. This is an example of students’ responding according to what they think and hope they will do compared to their evaluation of what they have really done.

A better test of repeatability is to compare MPEX results of the initial state of the students coming into the introductory physics sequence from several classes for the same sequence at the same school over a two to three year period of time. In this case, the student responses will be classified as favorable, unfavorable, or neutral. Recall that a favorable response is one that agrees with a majority of our experts.

For this comparison I wanted a data set that met the following conditions so that the class populations would be as similar as possible:

1. At least six classes of one course from one school were surveyed. (The sample needs to include enough classes to get a reasonable estimate of the distribution of mean responses.)
2. The classes are all from the main sequence over a period of two years and none are evening extension classes. (Off sequence classes often have varying mixtures of students who are taking the course early, students who delayed, and students who are repeating the course. Thus it is more difficult to say that any given class taught off-sequence is typical or has a population similar to a main sequence course. Evening extension classes often have populations that are significantly different from normal day classes. These are often older, returning students with different views of learning and motivation.)

3. The same version of the MPEX survey was used in all the classes (This ensures that the wording of all the items was the same for all classes and that none of differences are due to the change over from a pencil and paper survey to the scantron version.)

4. The average class size is a large enough sample to account for local fluctuations (The relatively small number of classes surveyed, N < 13, does not adequately account for the differences due to small sample size, i.e. classes with fewer than 50 students. In this study I have observed large significant fluctuations in both the initial MPEX results on some items and the initial FCI scores between physics classes taught in the same semester in the same school when average class size was < 50 students.)

Only data from four of the schools participating in the study met the first condition of pre-course data from at least 6 classes being given the MPEX survey at the beginning of the introductory physics sequence: University of Maryland (9 classes), University of Minnesota (10 classes), Dickinson College (6 classes), and Nebraska Wesleyan University (6 classes). Descriptions of the courses and schools can be found in chapter 8. However, of these four schools, only the data from University of Minnesota met all four conditions. University of Maryland and Dickinson College failed to meet conditions two and three. Both Dickinson College and Nebraska Wesleyan University have small classes and failed condition four. In addition, the six classes at Dickinson were surveyed over three years using versions 3.0 and 3.5 of the survey. Although versions 3.0 through 4.0 of the MPEX survey are very similar, there are some small wording changes on some items. At University of Maryland, classes from on and off sequence were surveyed with MPEX versions 3.0, 3.2, and 3.5. The data from all of
these schools were analyzed to see how weakening the conditions would affect the repeatability of the measurements.

The MPEX item results for these four schools were translated to a binomial distribution by renormalizing the observed $p$ and $q$ into $p' = \frac{p}{p+q}$ and $q' = \frac{q}{p+q}$ where $p$ is the observed percentage of favorable responses and $q$ is the observed percentage of unfavorable responses. The overall MPEX results and the cluster results were recalculated from $p'$ and $q'$. Then the measured averages and standard deviations for the individual survey items, the overall survey, and the seven clusters for each school were calculated from the normalized results for each class from that school. An estimated standard deviation $\sigma'$ calculated for a Gaussian distribution from binomial probability theory was calculated for each item, each cluster, and the overall survey. The overall survey and cluster results for the four schools are shown in Table 5-8 with both the measured $\sigma$ and the estimated $\sigma'$. The results from the University of Minnesota data for the survey items, the overall survey, and the clusters show that the distribution of actual results in all three cases is consistent with the distribution spread estimated using the binomial Gaussian distribution. The results for Nebraska Wesleyan University and Dickinson College for the overall survey and the clusters are also consistent with the distribution spread predicted using the binomial Gaussian distribution but for some of the individual items $\sigma' < \sigma \leq 2\sigma'$. The measured overall and cluster results for University of Maryland from courses both on and off sequence using up to three different versions of some survey.
questions were roughly consistent with the predicted distribution spread. While the spread in some of the items is consistent with the predicted spread, on six of the survey items $\sigma > 2\sigma'$.

These results indicate that if the four conditions for comparing classes with similar populations are met, the distribution of all normalized survey results are consistent with the distribution spread estimated using binomial probability theory. If the class sizes are small, there may be small differences in the spread of the distribution for some of the survey items, but not in the distribution of the overall survey and cluster results. The classes from University of Maryland, where the distribution included data from both on and off sequence classes and results from versions 3.0-3.5 of the survey, had larger differences on some of the survey items, but the distribution of overall survey and cluster results were roughly consistent with the predicted standard deviation. Thus, the MPEX survey results of classes at the beginning of the introductory physics sequence are reproducible and reliable. The overall and cluster results are more robust than the results of individual survey items.

VII. SUMMARY

We saw in chapter 2 that students’ expectations can have significant influence on what they learn in an introductory physics class. To better understand the role of student expectations in introductory physics, we developed the MPEX survey to study the distribution and evolution of student expectations through the introductory sequence. The MPEX survey items look at many expectation issues that affect student learning, but mainly focus on the following six dimensions described in more detail in Table 5-1:
independence, coherence, concepts, reality link, math link, and effort. The MPEX survey can determine the pre-course distribution of student expectations in a class both overall and for each of the six dimensions listed above. Measurements taken later in the sequence can show how the distribution changes during the sequence.

The MPEX survey was tested for both validity and reliability. Validity was demonstrated by studying the results of calibration groups and over 120 hours of student interviews. The results of these two tests indicate that both faculty and students agree with our interpretation of the MPEX survey items and what we consider to be an expert response. Although the interview results indicated that a few students occasionally misinterpreted some items, the interview study showed that the misinterpretation error is small and generally tends to overstate the favorable nature of the measured expectations. The survey was tested for reliability by comparing pre-course results for similar main sequence classes at a particular school. The measured distribution was comparable or less than the estimated distribution width for the overall survey results and each of the seven dimensions listed above.

---


5 S. Tobias, *They’re Not Dumb, They’re Different: Stalking the Second Tier* (Research Corporation, Tucson AZ, 1990).
To really determine what a person believes regarding a particular issue, current testing practice in the social science suggests that is necessary to ask at least five or six questions on that issue. Questions on related issues may not be sufficient to fully understand the belief in question. This is one reason why the reliability coefficients for sub-scales of survey instruments tend to be smaller (less reliable) than those for sub-scales of diagnostic instruments. See D.J. Mueller, *Measuring Social Attitudes: A Handbook for Researchers and Practitioners* (Teachers College, Columbia University, New York, 1986).

10 W. F. Perry, *Forms of Intellectual and Ethical Development in the College Years* (Holt, Rinehart, & Wilson, NY, 1970).


12 R. Likert, “A technique for the measurement of attitudes,” *Archives of Psychology* No. 140 (1932).

13 We refers to Edward R. Redish, Richard N. Steinberg, and the author.

14 A copy of the current version of both the pre and post versions of the survey can be found in Appendix C.

Indeed, some student comments lead us to suspect that formula sheets may have the tendency of confirming student expectations that formulas dominate physics. Their interpretation is that although memorizing lots of formulas is important for professionals, they do not need to do so for the current course. Thus, many faculty may be encouraging precisely that attitude they hope to discourage when they permit the use of formula sheets on exams. Redish, Steinberg, and I are not aware of any research that shows the effect of formula sheets on student perceptions of the coherence of the material.

Note that this is an area where students' beliefs about their abilities may surpass their actual abilities. More detailed investigations will require direct observation of student behavior on solving physics problems.

Redish has referred to this failure as a lack of parsing skills. These students, when faced with a complex sentence that they do not understand, will try reading it over and over again until it becomes familiar -- but they still may not understand it. They seem to lack the ability to decompose a complex sentence into its constituent parts in order to make sense of it. See E. F. Redish, "Is the computer appropriate for teaching physics?", *Computers in Physics* 7, 613 (December 1993).


The device of plotting three numbers whose sum is fixed in a triangle is well known in elementary particle physics as a Dalitz plot. In our case, the percentage responding agree, disagree, and neutral must add up to 100%.

Note that we have included all items, including those marked with parentheses in Table 3. As remarked above, even though the agreement on these items is not as strong, there is still a strong plurality of our experts in favor of the indicated responses. The shift in the position of the overall items resulting from removing these items is on the order of a few percent and the relative order of the groups is not modified.
Private communication with E.F. Redish and R.N. Steinberg, Summer 1995-96.

See Ref. 6

The ability of an individual to hold conflicting views depending on circumstances is a fundamental tenet of our learning model. See ref. 3 and R. Steinberg and M. Sabella, “Student performance on multiple choice questions vs. open-ended exam problems”, Phys. Teach. 35 (3) 150-155 (1997) for more discussion of this point.

“Homogeneous” in this case does not of course mean that we assume the students are identical. Rather, it means that the students are “equivalent” — that they are characteristic of the students who are to be found in “that type of class in that type of school.”

We choose this reduction from two independent variables to one because the primary variations we observe tend to maintain a fairly constant proportion of neutral responses.

D. J. Mueller, Measuring Social Attitudes (Teacher’s College, Columbia University, New York, 1986).

This crystal analogy was developed by E.F. Redish to describe the relation between students’ mental models and beliefs to what is measured by assessment, private communication.

There is a method of factor analysis to study non-orthogonal factors that uses oblique rotations.


Private communication with Gilley Shama (Spring 1997).

Factor loading is the correlation of the item with the factor.

In phenomenographic analysis, the researcher looks at the data obtained and develops categories from the data to classify the results. For more information on phenomenographic analysis, see F. Marton, “Phenomenography — a research approach into investigating different understandings of reality,” J. Thought 21 (3), 28-49 (1986).

Michael Wittmann is another senior graduate student in the University of Maryland Physics Education Research Group.

See Ref. 33.


See Ref. 33.

Variance is the square of the standard deviation.

See Ref. 36.
Chapter 6. Open Ended Written Assessments: Quizzes and Exams

I. CHAPTER OVERVIEW

While surveys and diagnostics are useful for determining the students’ initial state and changes that occur during the introductory sequence, more traditional assessments like exams and quizzes also have a role in determining what students know and what they are learning. The open-ended questions used in quizzes and exams often require deeper thinking from the students and can reveal more about how students understand and reason with the course material than do multiple choice questions. In this chapter, we look at how exams and quizzes can be used more effectively to evaluate what students are learning in introductory physics courses. In order to understand how they can be used to tell us more about student learning, it is useful to recap briefly some of the research on problem solving and learning.

As we saw in chapter 2, there are significant differences between novice and expert problem solvers. In addition to having more and better organized physics knowledge, expert problem solvers make heavy use of qualitative descriptions of the problem both in forming the solution and evaluating the answer.¹ This requires a good conceptual understanding and the view that physics is a coherent framework of a few key principles that can be used to understand real world phenomenon. Experts also make use of a problem solving strategy like the one shown in Table 2-1.

In contrast, at the beginning of the introductory course, many students solve physics problems by looking for equations that contain the unknown quantity and
working backwards or looking in the text for examples that are similar. During the introductory sequence, most students at some point learn to classify and approach physics problems according to their surface features. The best students develop the ability to classify and solve problems according to the physical principles involved. These students are developing a more expert view of problem solving.

It is too much to expect that the introductory physics sequence can turn students from novice problem solvers into experts. However, it is appropriate as one of our main goals for the introductory physics sequence to help students begin developing the skills and attitudes needed to become expert problem solvers. Note that these skills and attitudes that instructors would want their students to develop to achieve this goal are part of the hidden curriculum that often aren’t demonstrated, explicated, or reinforced through graded assignments to the students. For example, three of these skills and attitudes could be to have students achieve the following:

1. understand and use the principles, concepts, and laws of physics
2. develop an expert-like problem solving strategy
3. link problems to other contexts including real world applications

This chapter discusses some ways in which exams and quizzes can be used both to help students reach these goals and to learn more about how they think about physics along the way. In section II, I discuss some of the difficulties with traditional textbook problems. This is illustrated with a student’s response to a traditional problem and a qualitative problem on the same exam. Some alternative problem styles are presented as examples of how exam problems can be used to determine more of what students are thinking.
In section III, I discuss how conceptual quizzes can be used to see what students are thinking during instruction so that the instruction can be tailored to where the students are. The results of two pretests, conceptual quizzes associated with University of Washington Tutorials, from classes at University of Maryland are presented. The first pretest inquires into students’ conceptual understanding of Newton’s Laws. The second pretest looks at students’ mathematical understanding of what it means when one equation is the solution to another. The last section summarizes the chapter with a discussion on what quizzes and exams can tell us about student learning.

II. EXAM PROBLEMS

A. Traditional Textbook Problems & Qualitative Problems

One difficulty with many textbook problems is that they can be solved by students by just finding an equation that contains the unknown quantity, plugging in the given quantities, and calculating a numeric answer. These problems do not require much conceptual understanding or decision-making on the part of the students. In this dissertation, problems like this are referred to as “traditional textbook problems”.

If instructors’ goals for students go beyond solving traditional textbook problems, then using these types of problems almost exclusively on homework, quizzes, and exams may not be the best way to help them achieve the course goals. There are two reasons for this, one from a pedagogical standpoint and one from an assessment standpoint. First, from a pedagogical standpoint, traditional textbook problems encourage a mechanical, mathematical, algorithmic approach to problems solving. This in turn encourages the use of the novice strategies discussed above. As shown in the...
examples of Mazur and Hammer in chapter 2, students using this type of approach can often solve the problem and obtain an answer without understanding the underlying physics. Heller and Hollabaugh found that even when working in groups, students focus on the formulas and the calculations with little discussion and evaluation when they solve textbook problems. Tobias’ observers noted that in their physics courses, the near exclusive use of these textbook problems on homework and the use of simple versions on exams seemed to affect the course goal in the minds of many of their classmates. These students’ course goal was not to understand the physics concepts and use them to understand phenomena, but to learn to solve the textbook problems with minimal effort.

Second, from an assessment standpoint, traditional textbook problems are not always good indicators of what students know and understand. Because students can often solve these problems without much consideration of the underlying physics, the written solutions may not show much of how students think about and use physics concepts and representations. We saw examples of this in the studies by Mazur and Hammer in chapter 2. In both cases, assessment by traditional textbook problems failed to indicate significant student difficulties with the course material.

The responses of a single student to two problems (out of seven) on a final exam shown in Figures 6-1 and 6-2 are a case in point. The exam was given at the end of the first semester introductory physics sequence for engineering majors at University of Maryland. The problem shown in Figure 6-1 is a typical textbook-style problem on projectile motion. The problem shown in Figure 6-2 is a qualitative problem I wrote to look at students’ understanding of Newton’s 2\textsuperscript{nd} & 3\textsuperscript{rd} laws and velocity graphs. Note
Figure 6-1. Student solution to a traditional textbook style problem. This problem looks at student understanding of projectile motion. Notice that there is no mention of force.
Figure 6-2. Student solution to a qualitative exam problem. The problem is designed to look at student understanding of Newton’s 2nd & 3rd laws and velocity graphs.

4. Two carts, A and B (Mass$_A$ < Mass$_B$), are placed on a table then stuck together with Velcro. Using pulleys, two small blocks, C and D (mass$_C$ > mass$_D$), are connected by light strings to the carts as shown below. Initially, the carts are held in place. Ignore all friction in this problem.

At t = 0, the carts are released. At t = 3 seconds, the string attached to cart A breaks. At some later time, the carts return to their starting point.

a. Draw and label two separate free-body diagrams, one for each cart, for a time after the carts start moving but before the string breaks. (7 points)

b. Rank all the horizontal forces from both your diagrams by magnitude, from largest to smallest. Explain the reasoning that you used to rank the forces. (7 points)

All horizontal forces = $T_c + T_d + T_{cart A} + T_{cart B}$.

$T_c > T_d > T_{cart A} > T_{cart B}$

Because $T_d$ is heaviest, moves the whole system to left.

Then $T_c$ is 2nd heaviest, moves the system back to original position.

$T_{cart A}$ is 3rd, it is because cart A is lighter than cart B and because cart A is lighter, gives $T_{cart A}$ larger magnitude. And last only $T_{cart B}$ is left so it is last.

c. Briefly describe the motion of cart B from t = 0 until it returns to its starting point. On the graph below, qualitatively sketch the velocity vs. time for this time period. (6 points)
Figure 6.3. A correct solution to the qualitative problem in Figure 6-2a

We are given that \( \text{Mass}_A < \text{Mass}_B \) and \( \text{mass}_C > \text{mass}_D \) and that we can ignore frictional forces. We are told that at \( t = 0 \) the carts are released, at \( t = 3 \) seconds the string attached to cart A breaks, and at some later time the carts return to their starting point.

a. Draw and label two separate free-body diagrams, one for each cart, for a time after the carts start moving but before the string breaks.

b. Rank all the horizontal forces from both your diagrams by magnitude, from largest to smallest. Explain the reasoning that you used to rank the forces.

\( T_{\text{string} \rightarrow A} > N_{A \rightarrow B} = N_{B \rightarrow A} > T_{\text{string} \rightarrow B} \quad \text{Because m}_C > m_d, \text{ the two carts accelerate to the left after they are released. Since both carts accelerate to the left by Newton's 2nd law again, there is a net force on each cart to the left.} \Rightarrow T_{\text{string} \rightarrow A} > N_{A \rightarrow B} \text{ and } N_{B \rightarrow A} > T_{\text{string} \rightarrow B}. \quad \text{And from Newton's 3rd law we know } N_{A \rightarrow B} = N_{B \rightarrow A}. \)

c. Briefly describe the motion of cart B from \( t = 0 \) until it returns to its starting point.

On the graph below, qualitatively sketch the velocity vs. time for this time period.

The carts accelerate (constant acceleration) to the left after they are released until the string breaks. Then the carts slow down, stop, and accelerate to the right eventually reaching their starting point. After the string breaks acceleration is constant pointing to the right.
that unlike most textbook problems, there are no numbers and the motion is described in
detail qualitatively. A correct solution for the qualitative problem is shown in Figure
6-3. Let’s see how the student approached the two problems.

In the textbook problem, the student starts by drawing a picture using the
information given in the problem. Then the student writes down the three kinematic
equations as they can be found in any textbook. Then the initial values are plugged into
the general displacement equation. At this point the student did not recognize that the
initial y velocity is zero. The student solves the rest of the problem before realizing this
mistake. Then the student writes down the equations for the x and y components of
displacement and velocity. The y displacement equation is used to solve for the time of
flight. The student again proceeds to answer the remaining three parts of the problem.
Notice that once the equations are written down and the initial conditions are plugged in,
the student treats the rest of the solution as a mathematical exercise except for noting
that time is always positive. However, the student changes the sign in the y displacement
equation at the bottom of the page and ignores the minus sign in the answer for the y-
component of velocity. These indicate that the student is not using a consistent
coordinate system with respect to the y-coordinate and perhaps not relating the given
quantities to the physical situation defined by the problem. However, as far as the
problem and the grader are concerned, this student has demonstrated a good
understanding of projectile motion.

For the qualitative problem, the student draws the free-body diagrams, ranks the
forces, and draws the velocity graph. The free-body diagrams have the correct number
of forces and the forces are labeled more or less correctly by type and magnitude. Force
vector lengths do not seem to be drawn to scale. However, in ranking the forces the student ignores Newton’s laws of motion and uses two common sense beliefs. One, the system moves in the direction of the largest force and two, the heavier mass exerts the larger force. At first glance, the velocity graph looks correct but incomplete. However, the velocity goes to zero at $t = 6$ seconds and appears to stay there. Without the requested description of the motion it is hard to be sure, but other students who answered this way indicated that the second point where $v = 0$ is where cart B returns to its starting point. This would be consistent with kinematic graph difficulty discussed in chapter 2 where the students thought of velocity graphs as “pictures of the motion” and drew velocity graphs that are really position graphs. If this was the case here, the student may be indicating that the two carts move with constant velocity to the left and then move with constant velocity to the right. Note that there is no indication in part b that the carts are accelerating. An analysis of the 2-cart qualitative problem shown in Figure 6-2 for traditional and tutorial classes at University of Maryland is discussed in chapter 9.

We see that the one-dimensional qualitative problem is a better indicator of this student’s difficulty with the course material than the two-dimensional textbook-style problem. The student’s solution to the qualitative problem indicates difficulties both with velocity graphs and with Newton’s Laws while the solution to the projectile motion problem only indicates a difficulty with applying a consistent coordinate system.

However, qualitative problems are only one way to look at what students are learning. Other types of problems that can be used for this purpose including essay questions and estimation problems.
B. Essay Questions

As Tobias’ student observers commented in chapter 2, most traditional physics lecture courses do not provide much opportunity for students to discuss and debate the material. The emphasis on traditional quantitative problems also does not provide many opportunities for students to write about physics and what they are learning, even sometimes in courses where laboratory reports are required. Writing requires students to think about physics in a less mathematical way and can help students recognize the implications of what they are learning. Writing assignments also can help instructors understand how students think about specific physics concepts.

As discussed in chapter 3, one method to introduce more writing into the introductory physics class is to require the students to keep journals. Another method is to ask essay question on homework and exams. Two examples of essay exam questions used by Redish at University of Maryland are shown below:

1. Newton’s first law states an object will move with a constant velocity if nothing acts on it. This seems to contradict our everyday experience that all moving objects come to rest unless something acts on them to keep them going. Does our everyday experience contradict one of Newton’s Laws? If it does not, explain the apparent contradiction. If it does, explain why we bother to teach Newton’s first law anyway.

2. Define and discuss what we mean by an electric field.

In his instructions on exams and homework, Redish encouraged his students to cite examples and demonstrations used in class. For the first question, students needed to reconcile their real world experience with textbook physics by understanding the role of friction in everyday motion. The second question arose from concern that students in introductory physics have trouble with the concept of fields. When question 2 was given on an exam covering electric force and electric field, it brought out some student
misconceptions such as defining the field as an area or a volume and confusion over the role of the test charge.

A third example of an essay question that I wrote to look at students understanding of the nature of waves is shown with a student response (from an algebra/trig-based introductory course at a community college) in Figure 6-4. In this question, the students were asked to use what they had learned in all aspects of the class on the nature of waves including laboratories, concept-building activities, demonstrations, and class discussions to support their answer. It asks the students to think about what they know and why they believe it. Note that in the student solution shown in Figure 6-4, the student correctly states that wave speed is a property of the medium and correctly describes constructive and destructive interference but incorrectly uses interference to explain the Doppler shift of a fire engine siren when it is moving towards the observer.

In each of the three examples, the essay questions ask students to relate either different concepts or different aspects of a concept in a single question. These types of questions encourage students to think about the concepts they are learning and how they relate to one another. They are an appropriate response to the comment by Tobias’ observers (see chapter 2) that typical exam questions tend to require only the use of single concept in a simple context and do not require or encourage a good understanding of the course material.
Problem 3 (15 points)
Is sound a wave? Why do you believe your answer is correct? How would you convince a
classmate (from the day class) who disagreed with you? Use all relevant observations from class
activities and your own observations to support your answer. Does your argument convince
beyond a reasonable doubt or does it use a preponderance of evidence?

Yes, sound is a wave. I would convince a classmate
by comparing properties of sound to properties
of waves. The velocity of sound changes in different
mediums just as the velocities of waves on a
spring change when encountering different
mediums. When constructive interference occurs
sound increases due to an increase in amplitude which
resembles constructive interference using a spring.
Sound waves also encounter destructive interference
which interrupts sound as a result of the displacement
This resembles water waves or waves on a string
because the displacement of both class with destructive
interference. There are many observations that support
this. If a person is positioned appropriately between two
speakers that are spaced apart they can witness destructive
interference. A fire truck approaching someone is a good
equivalent of constructive interference because as crests meet,
crests-troughs meet troughs, the sound increases. If a person sings a note with the right pitch
of frequency, the sound waves can shatter glass due
C. Estimation Problems

Most physicists at some point in their career learn to sharpen their problem solving and thinking skills by working what are known as Fermi, order of magnitude, or estimation problems. These problems are useful because they require the solver to

- pull together knowledge from several contexts,
- think about how to estimate reasonable values for numbers, and
- relate knowledge from formal instruction to real world situations.

However, these problems are not necessarily just for physicists. Estimation problems can be written at a level appropriate to an introductory physics course and used to encourage the development of the above skills in non-physics majors. This type of problem allows the students to practice applying what they learn in new and sometimes more realistic contexts but not at a significantly greater level of difficulty. By seeing how the students apply equations and concepts out of the context in which they were learned, the instructor can get insight into how the students think about the concepts. Some examples of estimation problems are shown below in the order they might appear in an introductory physics sequence.7

1. You and a friend are planning a two-week vacation out to the West Coast for a wedding next summer. However you’re both on a tight budget. Your friend thinks it would be cheaper to drive his car than fly. A cheap plane fare from BWI to San Francisco is $350 round trip. Realistically estimate your travel expenses to drive to and from the West Coast to see if your friend is right. What would your average speed be? Assume you will have free room and board at a relative’s house once you arrive.

2. This winter, the East Coast has been hit by a number of snowstorms. Estimate the amount of work a person does shoveling the walk after a snowstorm. Among your estimates you may take the following:

- The length of a typical path from a house to the street is 10 meters.
- Assume the snow fell to a depth of 4 inches.
Assume the snow was only moderately packed so that its density was equal to 0.2 g/cm$^3$ – about one fifth that of water.

In doing this problem, you should estimate any other numbers you need to one significant figure. Be certain to state what assumptions you are making and to show clearly the logic of your calculation. (In this problem, the answer is only worth 2 points. Almost all of the credit is given for your showing correct reasoning clearly.)

3. For next year's Physics Open House the department is planning to set up a bungee jump from the top of the physics building. Assume that one end of an elastic band will be firmly attached to the top of the building and the other to the waist of a courageous participant. The participant will step off the edge of the building to be slowed and brought back up by the elastic band before hitting the ground (we hope). Estimate the length and spring constant of the elastic you would recommend using.

4. A typical television tube works by accelerating electrons through an electrostatic potential and then bending them with a magnetic field as shown in the figure at the right. If the electrostatic potential difference used to accelerate the electrons through the anodes is 10,000 Volts, estimate the maximum strength of the magnetic field needed to control the deflection of the electron beam. Use your experience with television sets to choose reasonable parameters for the distances required.

The first three have been used as exam problems; the fourth has only been used as a homework problem. Notice that the problems increase in both complexity and the amount of thought required to solve the problem as the student moves through the introductory physics sequence.

I wrote the first example problem to see how my students in the community college class would use the concepts of velocity and average velocity in a realistic scenario where they cared about the quantity in question, i.e. money. For the most part, the students made reasonable assumptions and calculations in solving the problem although some had difficulty with what was meant by average velocity.

The other three questions were written and used by Redish in the engineering physics sequence at University of Maryland. The problem on calculating the work while
shoveling snow produced some interesting student views on work and how to calculate it. Some students (10% with grades ranging from F to A) responded with how much effort or time the job would take. Some students used Work = force x distance and then calculated the work as W = mgd where d was the length of the walk. For the bungee jumper/harmonic oscillator problem, many students assumed the maximum stretch would occur when the gravitational force on the jumper is equal in magnitude to the spring force on the person rather than when the spring potential energy equals the initial gravitational potential energy of the jumper. The fourth problem proved to be especially difficult for many students because of the number of relationships required between the length of the tube, the length of the applied magnetic field, and the strength of the magnetic field.

In each of the four problems the students are asked to apply their physics knowledge to new, applied contexts. Note that the concepts and equations are used in the same way the students learned them, but now the students must make decisions as to what concepts and equations apply to the problem. Based on the work of Heller et al. reported in chapter 2, I believe the decisions the students make in terms of what knowledge they should use, what information is needed, and what quantities need to be estimated encourage the development of expert problem solving skills. Because the students draw on their own experience to visualize the problem situation and develop a sense of real world numbers, estimation problems may help the students link the classroom physics to their everyday world. As an assessment tool, they can help instructors see how students use what they know outside the familiar context of the textbook and lecture.
III. CONCEPTUAL QUIZZES

A. What is a Conceptual Quiz/Pretest?

Based on the findings of Physics Education Research discussed in chapter 2, instructors need to know what students are thinking to effect real change in the students’ views and their working knowledge. However, since exams usually follow instruction on a topic, they come too late for instructors to see what students are thinking while they are still teaching a topic. While a pre-course diagnostic can be used to this purpose to some degree, it would be useful to be able to assess student thinking to determine what students believe coming into the class and how they are interpreting what they have learned in the class before the instructor starts on new material. Weekly quizzes with textbook-style problems have attempted to fill this role in some traditional physics courses. But as discussed in the previous section, this type of problem is often more a measure of whether the students recall a similar problem rather than a measure of how students are thinking about the course material. To see how students are thinking about the course material as it is being taught, both Mazur\(^9\) and McDermott\(^{10}\) have incorporated conceptual quizzes into the research-based curricula they each developed, Peer Instruction and Tutorials, respectively. Since Mazur’s peer instruction curriculum is not a subject of my investigation, only the conceptual quizzes used in Tutorials will be discussed here. In McDermott’s tutorial curriculum, these conceptual quizzes are called pretests.

The tutorial method was developed by McDermott’s Physics Education Group at the University of Washington to address students’ difficulties with physics concepts and
representations. As a result, their pretests are designed to assess these student difficulties. In the tutorial curriculum, the traditional problem solving recitation is replaced with a cooperative group activity where the student groups go through worksheets designed to help the students confront and resolve difficulties caused by their common-sense beliefs. (The tutorial curriculum is discussed in more detail in chapter 8.) The instructor becomes a facilitator helping the students understand the implications of what they are learning by asking questions. To help the facilitators learn how the students are thinking about the material going into the tutorial, a pretest is given in lecture before the students start doing the tutorials in their recitation sections. Ideally the pretest for a specific topic is given after the students have seen the material in lecture but before they work on it in tutorial. The students receive credit for taking the pretest, but they are not graded on what they write. This allows the students to write what they think without the penalty of being marked down for it.

The pretest is typically given in the first 10 minutes of a lecture class once a week. The questions are designed, not only to help the instructors understand how students are thinking about a particular topic, but also to help the students start thinking about some of the issues that will be addressed in the coming tutorial. The tutorial instructors meet after the pretest was given and before the tutorials start for that week to go over both the student pretests and the tutorial. One of the TAs does a quick tally to identify the most common student difficulties. This identifies key issues the tutorial instructors will want to make sure their students address in their tutorial sections.

At University of Maryland, we use some of these tutorials as well of some of our own using MBL and multimedia tools. We also use some group problem solving
tutorials to help students integrate the concepts with problem solving. The pretests for problem solving tutorials often address more mathematical student difficulties. Examples of a conceptual pretest and a mathematical pretest are given below.

B. Conceptual pretest example: Newton’s third law

Newton’s third law is quite possibly the most difficult mechanics concept for students in introductory physics classes to understand and accept.\textsuperscript{12} To address this difficulty, I wrote a Newton’s third law pretest and tutorial using two carts and MBL force probes for the first semester of the introductory physics class for engineers at the University of Maryland.\textsuperscript{13} The pretest questions from the 1995 fall semester are shown in Figure 6-5 on the next page. Post-course results from this semester including the exam problem that looks at student understanding of Newton’s third law in Figure 6-2 are discussed in chapter 9. This pretest was given in the sixth week of the semester after Newton’s laws of motion had been covered in lecture. The results of a quick tally from the pretest are shown in Figure 6-6. The actual pretest is shown in Appendix D.

The quick tally of the students results from this pretest are summarized below (percentages are rounded to the nearest five percent):

- In a collision between two cars where the one on the left is stationary, only 40\% of the students correctly stated that the forces the two cars exert on one another are equal. Another 40\% clearly state the force of the moving car on the stationary car is larger because it is moving.
- Only 15\% of the students said the forces in a head-on collision between a moving van and a car moving with the same speed were equal. Almost two thirds of the students (65\%) answered that the force exerted by the van on the car would be larger. Their reasoning was that the heavier object should exert a larger force.
- Only 20\% of the students were able to correctly state Newton’s third law in their own words. Almost a third of the students (30\%) used Newton’s second law to justify the third or referred to the two forces as acting on one object.
Figure 6-5. Newton’s third law pretest used in several first semester classes of the introductory physics sequence for engineering majors at the University of Maryland from 1994-1995. This pretest was used to assess student understanding of Newton’s third law before they worked an MBL tutorial to help them better understand the concept. Note that the students are asked to rank the forces in part I-B. This allows the students to think about the relative size of the forces without calculations.

I. There is a collision between two cars of equal mass where one car is initially at rest.

A. Draw a free body diagram for each car during the collision showing all forces.

B. Rank the magnitudes of all the horizontal forces and give your reasoning.

II. Now think about a head-on collision between a moving van and a Ford Escort. Each vehicle is initially moving at the same speed. The following questions refer to what is happening during the collision.

A. Does the moving van exert a force on the Ford Escort?

B. Does the Ford Escort exert a force on the moving van?

C. If the answers to questions A and B are yes, which force is larger? Explain your answers to A, B, and C.

III. Write out a complete statement of Newton’s third law in terms of forces in your own words. (You may use a diagram if you wish.)
Figure 6-6. Quick tally of the 86 student responses from the fall 1995 semester to the Newton’s third law tutorial pretest shown in Figure 6-5.

I. There is a collision between two cars of equal mass where one car is initially at rest.  
A. Draw a free body diagram for each car during the collision showing all forces.  
   • Correct - 23 students  
   • Drawing Newton 3 pairs on the same system - 27 students  
   • Use of impetus force (force is proportional to velocity) - 15 students  
   • Drew a force due to acceleration - 9 students  
   • Other errors or blank - 10 students  

B. Rank the magnitudes of all the horizontal forces and give your reasoning. (The moving car is car a)  
   • (Force of a → b) = (Force of a → b) - 36 students  
   • (Force of a → b) > (Force of a → b) - 36 students  
     This response was justified by the car’s motion  
   • Other - 7 students  
   • Blank - 7 students

II. Now think about a head-on collision between a moving van and a Ford Escort. Each vehicle is initially moving at the same speed. The following questions refer to what is happening during the collision.  
A. Does the moving van exert a force on the Ford Escort? 2 students said no, but it is  
B. Does the Ford Escort exert a force on the moving van? not clear if they understood the questions  
C. If the answers to questions A and B are yes, which force is larger?  
   • The moving van exerts a larger force - 54 students  
   • The two forces are equal - 15 students  
   • Blank and other - 10 students

III. Write out a complete statement of Newton’s third law in terms of forces in your own words. (You may use a diagram if you wish.)  
   • Correct statement of Newton’s third law - 19 students  
   • Using Newton 2 to justify Newton 3 (the two forces act on the same object) - 26 students  
   • Action/reaction with no mention of force or equal/opposite without specifying which object - 21 students  
   • Other - 5 students  
   • Blank - 15 students
Another 25% of the students used the phrase action-reaction without mentioning force or the phrase equal and opposite without mentioning the objects. Note that the use of the phrases “equal and opposite” and “action-reaction” by students is often an indication of student difficulty with Newton’s third law.\textsuperscript{14}

C. Mathematical pretest example: A solution and the wave equation

In the 1995 fall semester, I had the opportunity to test one of my hypotheses on one aspect of students’ use of mathematics in physics classes: Do students understand what it means to say that an equation is the solution to or satisfies another equation? During the 1995 fall semester, Redish at University of Maryland spent a lot of class time in the second semester of the introductory physics sequence for engineers emphasizing the derivation, meaning, and solutions of the wave equation. In the week of the exam, while Redish was reviewing material in lecture to help students prepare for the midterm exam, Redish and I decided to implement a group problem-solving tutorial on oscillations and waves to help the students learn to apply the concepts they were learning in solving problems. (See chapter 8 for a description of the tutorial teaching method and the implementation at University of Maryland.) I wrote the tutorial including a pretest to test two aspects of the connections between the mathematical formalism and the concepts the students were learning on waves. One of the problems asked whether \( y(x,t) = A \cos (kx + \omega t) \) is a solution to the wave equation and how would the student convince a friend of their answer. The questions from the pretest are shown in Figure 6-7 on the next page.

Initially, Redish objected that this problem might be too easy since students did not have to show that \( y(x,t) = A \cos (kx + \omega t) \) could be expressed in the general form of
a solution to the wave equation \( y[x,t] = F[x - vt] + G[x + vt] \), they could just take derivatives of equation 1 and substitute into equation 2. But that was the point of the pretest. Would the students know to use one of these two approaches to show that equation 1 was a solution to equation 2?

Of the 113 students who took the pretest, only 55% explicitly stated that \( y[x,t] \) was a solution to the wave equation including 3 students who said it looks correct. One third of the students did not answer the question explicitly and only 10% of the students answered “I don’t know” or “no, it is not a solution.” Based on the students’ answers to this question, there seem to be no significant student difficulties except for not answering the question. If there was a question about what the student wrote and why in the analysis of the pretest, I assumed the student gave the correct response or used the correct reasoning.

Surprisingly, only one third of the students either described or attempted a process that would have led to a correct solution. The results can be summarized as follows:

- Approximately 25% of the students took a derivative of \( y[x,t] \). Four-fifths of these students substituted these derivatives into the wave equation and one-fifth showed that the resulting equation was true.

- Another 5% of the students used the general solution to the wave equation to support their answer. Half of these students either stated or showed that \( y[x,t] \) could be expressed as part of the general solution.

- Another 2% said that they would show that \( y[x,t] \) is a solution by showing that it satisfies the wave equation but did not take derivatives of \( y[x,t] \) or compare \( y[x,t] \) to the general solution.
SHOW ALL YOUR REASONING IN YOUR ANSWERS.

1. A friend in this class does not believe that

\[ y(x, t) = A \cos(kx + \omega t) \]  

is a solution to the wave equation for the motion of a stretched spring,

\[ \frac{\partial^2 y}{\partial x^2} = \frac{\mu}{T} \frac{\partial^2 y}{\partial t^2}. \]

Do you think it is? Explain why you think so. How would you convince your friend of your answer?

2. A transverse wave is traveling on a long spring. A segment of this spring is shown at time \( t = 0 \) in the first of the two figures below. The second figure shows what this segment looks like at time \( t = 0.01 \text{ s} \). The y-axis is in cm, while the x axis is in m.

Assuming that the wave continues in the way it is shown, write an equation that would allow you to find the displacement of the spring \( y(x, t) \) for an arbitrary position \( x \), at an arbitrary time \( t \).
The other students used a variety of approaches. It is interesting to note that other than no reason (20% of the students did not give any reasoning), the two most common incorrect responses were based on ideas that were emphasized by Redish in lecture.

- 15% of the students explained that $y[x,t]$ is a solution to the wave equation because the wave equation is really just a form of Newton’s second law $F = ma$.
- 10% of the students used dimensional analysis of $y[x,t]$ or the wave equation in their reasoning.

D. Summary

It is becoming increasingly more accepted in the physics education community that effective instruction in the introductory class must take into account what students know and how they use what they know. Instructors often have a need to know more about what their students are thinking than can be learned with diagnostic and surveys instruments while they are teaching a topic. Conceptual quizzes like pretests can help meet this need by giving the instructor indications of what the students have learned and how they are using it.

Although pretests are usually used to assess well-known student difficulties with concepts and representations, they can also be used to assess students’ understanding and use of mathematics in physics. It is widely believed by physics instructors based on anecdotal evidence that students’ “mathematical knowledge” is often less than what would be expected from the list of prerequisites. Mathematical pretests like the one shown in this section can be used by instructors to better understand the mathematics difficulties of their own students.

However, pretests also have some limitations as well. First, pretests take class time to administer and analyze. As instructors struggle to cover all the required material
in the introductory course and keep up with their hectic schedules, it is sometimes hard
to find time for conceptual quizzes. Second, sometimes students don’t reflect on what
they are doing or just don’t write explanations with their answers. Not much can be
learned on student thinking when the students don’t explain their answers. Third, the ten
minute time limit and need to use the questions to bring students’ attention to issues
relevant to the coming tutorial restricts the type of questions that can be asked.

IV. WHAT CAN EXAMS AND QUIZZES TELL US ABOUT WHAT
STUDENTS ARE LEARNING?

Many physics instructors use traditional textbook problems or minor variations to
assess student learning. There are two problems with this. First, as we saw in
chapter 2, the emphasis on this type of problem on homework and exams has been seen
to encourage undesirable student attitudes towards problem solving and the course.
These attitudes may limit what students learn. Second, and more importantly for
research purposes, problems of this type often do not reveal much about how students
think about the course material. Since students do not have to use their conceptual
understanding, their solutions often emphasize the mathematical calculation and the
numeric answer as we saw in the student solution to the projectile motion problem in
Figure 6-1. To learn more about what students are thinking and learning, it is necessary
to use problems where students explicitly show more of their reasoning.

Physics education researchers and instructors can learn more about what students
are thinking by using problems that have one or more of the following properties:
• Problems that test conceptual understanding by having students construct physics representations like velocity graphs or free-body diagrams and compare quantities like the ranking of the horizontal forces acting on a pair of carts.

• Problems that require students to express their understanding in writing such as in short essay questions.

• Problems that get students to use what they know in new and more realistic contexts such as in estimation problems.

In this chapter, I have shown examples of how problems using these properties can identify and illuminate student difficulties with such basic concepts as velocity, Newton’s laws of motion, the nature of the Doppler effect, and how to demonstrate that one equation is a solution of another. As we saw in the examples of Mazur, Hammer, and Tobias in chapter 2, few if any of these student difficulties would have been revealed by traditional quantitative textbook problems. That is not to say that quantitative problems are not useful from a research perspective, but they need to incorporate some or all of the above mentioned features and require more decision making than just finding the right equation to use.

Because students must give a solution that includes how they came up with an answer, quizzes and exams can be more informative than survey and diagnostics for showing what students are learning and thinking. However, even in quizzes and exams it is often hard to see how a student was thinking on a problem or why they answered a particular way. Here, the physics education researcher needs to resort to a method unavailable to most classroom instructors, interviews with the students.


5 Prof. E.F. Redish at the University of Maryland used these problems in the algebra/trig-based introductory course and the engineering introductory physics course.

6 The O.J. Simpson civil trial concluded during this semester and as a class we had discussed the different standards of evidence used in the criminal and civil trials.

7 For more problems and references on this topic, see the PERG estimation problem website at physics.umd.edu/rgroups/ripe/perg/fermi.html.

8 See Ref. 2.


11 See Ref. 9.


14 How student use language is often an indication of their understanding of concepts. This is a current area of physics education research and is not well understood. But many observations by the author and other members of a physics education research listserv have observed that students who use the everyday language of Newton’s
third law, i.e. phrases like “equal and opposite” or “action-reaction” often do not have a good conceptual grasp of Newton’s third law.
Chapter 7: Understanding Student Thinking through Interviews

I. OVERVIEW

In the past three chapters we have discussed methods of assessment that are relatively easy to implement in most introductory physics courses. No change in class format or additional resources like instructor time are required to use specially designed exam problems that allow students to demonstrate a richer understanding of physics than is the case with most traditional textbook-style problems. While some class time and additional resources may be required to administer and analyze conceptual quizzes and pre/post surveys like the MPEX survey and the FCI, their impact on the course and the instructor is minimal. These methods are very useful to monitor how a class is doing and can be implemented by researchers and instructors fairly easily, but there is no substitute for interviews with students for understanding what individual students are learning and thinking.

From a students’ solution to an exam problem, you can see the steps the student used to solve the problem and maybe some of the concepts that were used. In an interview, a researcher can ask the student questions to see

- why they answered the problem that way,
- what they were thinking as they solved the problem,
- how they interpret concepts and equations they used, and
- how they interpreted the question.

This type of understanding is often essential to physics education researchers to understand what individual student responses on exams, quizzes and diagnostics are telling us about students’ understanding of the course material. For example, the
understanding developed from interviews often provides a framework to interpret student exam solutions in areas where students have conceptual difficulties. In many cases, the incorrect student responses to a problem that encourages students to use their conceptual understanding are due to just a few common misconceptions. Interviews can also tell us why certain misconceptions occur, the difficulties they signify, and help researchers learn how to address them.

The use of interviews by physics education researchers led to much of the current understanding of student difficulties associated with problem solving, conceptual understanding, and expectation discussed in chapter 2. The research conducted in the early 1980s using interviews changed physics education research from studies that relied primarily on informal classroom observations of students to more detailed and systematic studies of how students think about physics. Many of these PER interviews are based on the technique developed by Piaget to study how children and adolescents reason in physical situations. (Many of Piaget’s interviews are based on common physics problems.\(^1\) Physics education researchers typically find the following three types of interviews to be the most useful: validation interviews, demonstration interviews, and problem interviews. Each type of interview serves a different research goal.

Validation interviews are used to check student interpretations of instruments. This is usually done in one of two ways. One approach is where the interviewer asks the subject about the instrument items and their responses directly. Another approach is to ask the students related questions and compare their responses with their responses to the instrument.
Demonstration interviews are used to probe students understanding and knowledge of the course material. In a demonstration interview the student is given a situation or an apparatus and asked a series of questions that relate to it. The interviewer guides the inquiry to probe the student’s understanding of the situation until they know what the student thinks will happen and why. This type of interview is often used in studies of students’ common sense beliefs and misconceptions.

Problem solving interviews are used to see how students use their knowledge and reasoning when they perform tasks like solving a physics problem. The student is given the problem or task and works on it while being observed by the interviewer. The interviewer may ask some questions, but most of the time the interviewer is a silent observer. This allows the observer to see how the student works through the task on his or her own. This type of interview was used in many of the problem solving studies described in chapter 2.

Physics education researchers generally use one of two basic interview techniques, question and answer and think aloud. In the question and answer technique, the subject responds to the interviewer’s questions. The interviewer listens to the subject’s response and may ask follow-up questions to clarify the response. In a think-aloud interview the students say whatever is going through their minds as they perform a task or answer a question. As some students find this difficult, I went over a copy of the hints shown in Table 7-1 with students participating in think-aloud interviews for this dissertation. Note that students are requested not to elaborate on what they have already said. The primary task of the interviewer is to keep the student speaking. Most
Table 7-1. Think aloud hints given to students at the beginning of demonstration or problem solving interviews. The hints come from p. 33 of the D.N. Perkins, *The Mind’s Best Work* (Harvard U.P., Cambridge, MA, 1981)

1. Say whatever is on your mind. Don’t hold back hunches, guesses, wild ideas, images, or intentions.
2. Speak as continuously as possible. Say something at least every five seconds, even if only, “I’m drawing a blank.”
3. Speak audibly. Watch out for your voice dropping as you become involved.
4. Speak as telegraphically as you please. Don’t worry about complete sentences and eloquence.
5. Don’t over explain or justify. Analyze no more than you would normally.
6. Don’t elaborate on past events. Get into the pattern of saying what you’re thinking now, not of thinking for a while and then describing your thoughts.

students have a tendency to draw within themselves when thinking about something difficult. Think-aloud interview protocols work best with individual students while question and answer interviews can be used either with individuals or groups.

In this chapter I describe the different types of interviews and protocols used in this study. In section II, I discuss the two MPEX protocols used to study student expectations. Section III goes over the interview protocol used to study students’ conceptual understanding of physics. This section also includes the results of a demonstration interview concerning the mathematical description of a propagating pulse on a string. This study is used as an example to illustrate what can be learned from this type of interview.

All the students interviews used in this dissertation were videotaped and transcribed. The student volunteers were solicited by requests made in class and required to sign a release form before being taped. A copy of the blank release form is included in Appendix E.
II. EXPECTATIONS

As discussed in chapter 5, I conducted over one hundred hours of interviews with students in my study of expectations. Two different expectation interview protocols were used, the MPEX Survey protocol and the Open MPEX protocol. Both protocols use the question and answer technique. The MPEX Survey protocol is shown in Table 7-2. It was designed primarily to validate the MPEX survey and to understand the reasoning behind the student responses by going through the survey items. The Open MPEX protocol is shown in Table 7-3. This protocol was designed primarily to probe students’ math expectations and expectations regarding the connections between physics and mathematics. The students were interviewed either singly or in groups of two, three, or in rare instances four.

A. MPEX Survey Protocol

My main goal in these interviews was to validate the survey items by listening to the students’ interpretations of the survey items and explanations why they answered the items the way they did. This told me whether the students were interpreting the survey items as I intended and if their reasoning made sense in light of the survey item. These interviews also helped me understand more about the nature of student expectations, how expectations shaped students’ view of the class, and what aspects of the class students felt helped or hindered their learning. The results from these interviews are reported in chapter 10.

Table 7-2. MPEX Survey protocol (Spring 1997 version) – This protocol is used to validate the MPEX survey and to study student expectations in the context of the survey items. The students are asked to complete a pencil and paper version of the survey and sign a release form before the interview starts.
Part 1. Background:
This is Jeff Saul. I'm here at <name of school> with <student code name> who is in <describe class> with Professor <Prof. name here>. The interview today will have three parts. I will start the interview by asking you some questions about your background. Then I will ask you some questions about your survey responses. In the last part of the interview I will ask you some questions about the class.

Background questions
• So, first off, tell me a little bit about yourself. Where'd you grow up?
• What math classes did you have in high school?
• What science classes did you have in high school? Did you have physics in high school? Did your physics class use calculus?
• What made you decide to come to college here?
• Is the class what you expected to be?

Part 2. MPEX survey items:
[In this section the students are asked to go over 15-30 of the MPEX survey items depending on the length of the students’ responses and the available time for a particular interview. Items 1, 2, 5, 9, 12, 13, 14, 19, 21, 22, and 34 were asked in most interviews.] The students are asked to read the survey item aloud, read their response, and explain why they answered that way. If students change their minds because they misread the question, they are instructed to cross out their first response with a single line and circle their new response. If students change their minds after thinking about their reasoning, they are instructed to cross out the old number completely and circle their new response.

Part 3. Comments on instruction:
Questions
• What did you like about the class? What did you dislike about the class?
• What would you say is the most interesting or significant thing you've learned this year in physics class?
• Of all the activities you did as part of this class, which were most useful for learning physics? Which were least useful?
• Did you work with other people outside of class? On a regular basis?
• If you had to make one change to the class to improve it, what would it be?
Table 7-3. Open MPEX protocol – this protocol is used to study student expectations with an emphasis on the connections between math and physics.

1. How would you describe yourself (as a student)?
2. What is your major?
3. What made you decide to be a ___________ major?
4. What is math?
5. What do you like about math and what do you dislike about it?
6. Do you think math is useful? Give an example.
7. What do you consider to be some interesting or significant things you have learned in your math classes?
8. What has stayed with you from your school experiences in math and (physical) science classes?
9. Do you recall any topics or issues from one of your “early” math classes that you found particularly difficult or troubling?
10. Do you use what you have learned in mathematics classes in other classes or in your everyday life? Give an example.
11. What is an equation? How can you use one in physics?
12. How do you take a physics problem and convert it to mathematics?
13. How do you decide which equation to use in a particular problem?
14. What are the steps you use in solving a physics problem? / How do you go about solving a problem?
15. When you have obtained an answer to a problem, what do you do next? Is this different for solving an exam problem vs. a homework problem?
16. When you have solved a physics problem do you have an intuitive feel for the correctness of your answer or do you need to look it up in the back of the book?
17. Do you find derivations done either in the text or in the lecture to be useful? Do you ever do any yourself?
18. Is showing that something is true in math different from the way you would show something is true in physics?
19. When you do an experiment in physics class and you don't get the expected results, what have you shown?
20. Do you feel that the physics you learn in class is connected to the real world?
21. Do you ever find it useful to think about your own experiences and things that you've seen and things that you've done when you're trying to learn physics either in class or when you're studying physics on your own?
These interviews were conducted in three parts: background, survey items, and class comments. If there was sufficient time, a student was asked to work a problem individually as a think aloud. The two-rock problem interview, discussed later in this chapter, was the one most commonly used, particularly at the end of the first term. In the background section, I ask the students about where they grew up, the math and science classes they took previously, and some general expectation questions such as “What did you expect this class to be like?” and “Will learning physics be useful to you in your career?”

For the part of the interview on the survey items, I selected 15-30 of the MPEX survey items to go over with the student(s). In the first two years of the study, I asked the students to read the item aloud, mark their response, and then tell me their response and why they answered that way. When the student finished explaining their response, I would ask a follow-up question if their reasoning was not clear or if they did not give an example. When we were finished with that item, I would tell the students which item I wanted them to discuss next.

This format worked well, but because the students were going through the items one at a time there was a possibility that the discussion of the earlier survey items in the interview could be affecting the students’ response and discussion of the latter items. The protocol was changed so that students filled out a special version of the MPEX survey form with room for comments before the interview. I still had the students discuss their responses to selected survey items one at a time, but this allowed me to see which student responses were being changed during the interview and to ask why they were being changed.
In the last part of the interview, I asked the students questions about the class. These included questions like “Which activities were most helpful for helping you learn physics?” and “What would you do to improve the course?” I concluded this protocol by asking the students if they had any additional comments about the class or on what they had learned.

The validation results from interviews using this protocol are discussed in chapter 5. An analysis of expectation interviews with students is discussed in chapter 10.

B. Open MPEX Protocol

The open MPEX protocol is shown in Table 7-3. This protocol was designed to look at issues related to the MPEX survey with a heavy emphasis on math expectations and the connections between math and physics. Some of the questions used in this protocol are follow-up questions for the MPEX survey protocol. This protocol was developed to ask questions that would not be as leading to the students as the survey items. The open format with its more open-ended questions tended to produce student responses that were less directed and less focused than the MPEX survey interviews.

Because the this format only addresses some of the issues brought up by the MPEX survey, results from these interviews are discussed in only limited detail in chapter 10. A more detailed analysis of these interviews will be continued after the completion of this dissertation.
III. CONCEPTS

While the MPEX interview protocols use the question and answer technique, both demonstration and problem interviews use the think aloud technique to probe how students think about physics. The two interview studies discussed in this section each play a role in interpreting the study results presented in chapters 9 and 10. The first interview study, the waves-math interview, is a good example of a demonstration interview. It also illustrates an important effect of the fragmentation of student knowledge discussed in chapter 2; namely, that students can simultaneously hold conflicting views and their response to a question may depend on the specific trigger. The second interview, the two-rock problem interview, was used during site visits at three of the ten schools participating in this study as part of the evaluation of both students’ conceptual understanding and expectations. Because of students’ difficulty with the two-rock problem and the interview focus on issues of conceptual understanding and expectations, the interview protocol calls for more interaction from the interviewer than is usually the case for a problem solving interview both to help the students and to clarify what the student is doing and why.

A. Waves-Math Demonstration Interview

From the 1996 spring semester to the 1997 spring semester, Michael Wittmann and I conducted interviews with the waves-math think-aloud protocol shown in Table 7-4. We designed this protocol to investigate an unusual result from the part one of the pretest shown in Figure 7-1 (a description of pretests and how they are used may be found in chapter 6). The students were shown the shape of a single propagating pulse
1. Consider a pulse propagating along a long taught string in the \( +x \)-direction. The diagram below shows the shape of the pulse at \( t=0 \). Suppose the displacement of the string at various values of \( x \) is given by \( y(x) = Ae^{-\left(\frac{x}{a}\right)^2} \).

![Diagram of a pulse]

A. On the diagram above, sketch the shape of the string after it has traveled a distance \( x_0 \), where \( x_0 \) is shown in the figure.

For the instant of time that you have sketched, find the displacement of the string as a function of \( x \). Explain how you determined your answer.

B. Consider a small piece of tape attached to the string at \( x_0 \). Describe the motion of the tape during the entire motion of the pulse.

2. The experiment described in question 1 is repeated, except that at \( t=0 \),

\[
y(x) = 2Ae^{-\left(\frac{x}{a}\right)^2}.
\]

A. Compare the motion of the pulse in question 1 with the motion of the pulse in this experiment. Explain.

B. Compare the motion of the piece of tape in question 1 with the motion of the tape in this experiment. Explain your reasoning.
on a long string at time $t = 0$ and given a Gaussian equation $y(x)$ that described the shape of the wave at that time. Then they were asked to draw the shape of the string after the pulses had traveled a distance $x_0$. Approximately one third of the students drew a pulse that had the same shape but was definitely smaller. This would be a correct response if a student drew the pulse smaller to account for the dissipation that would be observed in a real pulse. However, most of the students who gave a reason explained that the decreased amplitude was due to exponential decay. Wittmann observed a similar result when the pretest was given the following semester.

We believed that the exponential expression $e^{-\left(\frac{x}{b}\right)^2}$ in the equation for $y(x, t = 0)$ triggered the students to automatically sketch a smaller pulse. To test our hypothesis, we interviewed student volunteers from a traditional lecture class that had already covered waves. We found that the students had two different ways to think about what was happening to the pulse, either in terms of physical waves or equations.

The two ways of thinking led to two different answers. If a student answered the question by thinking of the pulse as a wave, they said the pulse shape should remain the same. If a student answered the question in terms of the equation $y(x)$ at $t = 0$, they said the wave would get smaller because of the negative exponent in the exponential term. Two of the students asked if they should consider the problem in terms of waves or the equation and gave the appropriate answers for each view. Another student (code named Ferno) changed his answer as he thought about the problem as shown on page 7-15.\(^4\) (Italics are used here to represent direct quotes from Ferno. Italics within quotation
PART I.

STUDENTS ARE GIVEN THE FOLLOWING ON A SEPARATE SHEET:

Consider a pulse propagating along a long taught string in the +x-direction. The string is lying on a table, and there is no friction between the string and the table. The diagram below shows the shape of the pulse at \( t = 0 \) s. Suppose the displacement of the string at various values of \( x \) is given by \( y(x) = Ae^{-\left(\frac{x}{b}\right)^2} \).

ASK THE FOLLOWING QUESTIONS:

- On the diagram you are given, sketch the shape of the string after the pulse has traveled a distance \( x_0 \).
- Explain why you drew the string in the shape you did. (How do you account for the shape that you have drawn?)
- Write an equation that describes the displacement of the string after the pulse has moved a distance \( x_0 \).
- For the instant you have sketched, write an equation that describes the displacement of the string for all \( x \) and all \( t \).
- Imagine you are holding a string attached to a distant wall. How would you create a shape like this on a string?
- Imagine that it takes the pulse 1 second to move from \( x = 0 \) to \( x = x_0 \). How would you shorten this time? (possible follow-up: how would you increase this time?)
PART II.

A BLANK GRAPH IS GIVEN TO THE STUDENTS. ASK THE FOLLOWING QUESTIONS:

The previous experiment is repeated, except that now you move your hand twice as far to the side in the same amount of time as before.

ASK THE FOLLOWING QUESTIONS:

- Sketch the shape of the string at $t = 0$? (how is this shape different from the shape of the string in the previous experiment?)
- Could you write an equation for the shape of the string at time $t = 0$ sec?
- How would the motion of the pulse in this situation compare to the motion of the pulse in the previous experiment?
- Could you sketch the shape of the string at $t = 1$ sec?
- Write an equation that describes the displacement of the string for all $x$ and all $t$. 

FERNO: [Reading the question]"... A on the diagram above, sketch the shape of the spring after the pulse has traveled a distance $X$, zero $[x_0]$, where $X$, zero is shown in the figure. Explain why you sketched the shape as you did."

Okay. Umm ...Let’s see. "Sketch the shape of the spring after the pulse has traveled” (MUMBILING AS HE REREADS THE PROBLEM). ... Okay. Over a long, taut spring, the friction or the loss of energy should not be significant; so the wave should be pretty much the exact, same height, distance -- everything. So, it should be about the same wave. If I could draw it the same. So, it’s got the same height, just at a different $X$ value.

No, wait. Okay. "... the displacement of (MORE MUMBLING, QUICK READING) ... is given by" -- B, I guess, is a constant; so -- It doesn't say that $Y$ varies with time, but it does say it varies with $X$. So -- That was my first intuition--but, then, looking at the function of $Y$ ... Let’s see, that -- It's actually going to be -- I guess it'll be a lot smaller than the wave I drew, because the first time -- $X$ is zero, which means $A$ must be equal to whatever that value is; because $E$ raised to the zero's going to be 1. So, that's what $A$ is equal to. And then as $X$ increases, this value, $E$ raised to the negative, is going to get bigger as we go up. So, kind of depending on what $V$ is ... Okay. So, if $X$ keeps on getting bigger, $E$ raised to the negative of that is going to keep on getting smaller. So the -- So the actual function's going to be a lot smaller. So, it should be about the same length, just a lot shorter in length.

Ferno began by thinking of the pulse as a wave in the second paragraph, then he begins to think about the equation $y(x, t = 0)$, and then changed his answer because he saw that $y$ was going to get smaller as $x$ gets bigger. When asked what the equation describes, he replies,

FERNO: The equation is just exactly what it says. It's telling you the displacement of the spring, which is going to be in the $Y$ direction. As you pick your $X$ value, that's going to determine your $Y$ value. So, that equation just -- You pick your $XO \ [x_0]$ value, wherever it may be; and that's going to dictate where your $Y$ is going to be.

From the way Ferno and some of the other students explained what the equation described, it seems that the smaller pulse response is not an automatic response to the
sight of the decaying exponential term but instead is due to students thinking that $y(x)$ represents the maximum displacement of the pulse, the amplitude, as a function of position rather than describing the shape of the pulse at time $t = 0$. The following semester I tested this hypothesis by using the same protocol with two more student volunteers, but the equation describing the pulse was changed to $y(x) = \frac{A}{\left(\frac{y}{b}\right)^2 + 1}$. One of the students used a wave description and said the pulse would retain its shape and stay the same size. The second student initially answered that the pulse would get smaller because $y$ was getting smaller as $x$ got larger. Then recognizing that $A$ would be maximum displacement or amplitude, she became confused until she realized that the displacement $y$ was not the same as the amplitude $A$. While trying to resolve $y$ and the amplitude, she clearly indicated that she had thought $y$ described the amplitude of the pulse as a function of the position of the pulse peak.

In summary, by using the pretest results, Wittmann and I were able to design an interview protocol to investigate a student difficulty and understand the student reasoning behind it. From the interview results, we reached the following conclusions:

1. Students are misinterpreting the meaning of the equation and assume a special significance for the pulse peak.
2. Students answer the question differently depending on whether they use their knowledge of waves or their understanding of the equation. Some students hold both perspectives and give both answers simultaneously.

Note that while the first result is consistent with other observations of student difficulties with waves and mathematics, it was only possible to see this connection by using the interviews to see what the students were thinking. Likewise, the second result would likely be unobtainable without the information from the interview. Here we have a direct
indication of the fragmentation of student knowledge discussed in chapter 2 and indications of how it is triggered. The two students who held both views simultaneously were aware of the contradiction, but made no effort to resolve it. This issue of simultaneous conflicting concepts and triggering will be needed in interpreting the study results in chapters 9 and 10.

B. Two Rock Problem Interview

The two-rock problem interview is useful for looking at several aspects of student learning including conceptual understanding and expectations. For this reason Hammer used this problem in his interviews with students as part of his dissertation research on cognitive beliefs described in chapter 2.\(^6\) He found it particularly useful for looking at students understanding of the equation \(v(t) = v_0 + at\). The problem itself is stated below:

\[
\text{TWO ROCK PROBLEM: Two rocks are thrown with a speed of } v_0 \text{ from a cliff of height } h. \text{ One is thrown horizontally and the other straight down. Which one hits the ground first and which one hits with greater speed?}
\]

Although it does not look intimidating to students, many of them find this problem very difficult. One reason is that the wording suggests to them that the problem can be solved easily with the kinematics equations. But without using numbers, students cannot easily compare the times of flights or the final speeds using only the kinematic equations. They are forced to either make up numbers to substitute into the kinematic equations or use their creativity and knowledge to find another way to solve the problem. This can lead to discussions of what the students knows about several of the
basic topics in mechanics including the kinematic equations, vectors and vector components, forces, and conservation of energy as well as why the students believe what they know. Hammer used this problem to get students to explain where they thought the equation \( \mathbf{v} = \mathbf{v}_0 + \mathbf{a}t \) came from, how they knew it was true, and how they might explain it to someone else. He chose to use this problem for his dissertation study because he found Liza and Ellen’s interview solutions (in the preliminary study discussed in chapter 2) to the two-rock problem were indicative of their different expectations for the class.

Based on Hammer’s work, I decided to use this problem as a follow-up question in many of the MPEX survey interviews to see how the students approached a non-trivial physics problem and to study their knowledge of physics and reasoning in the context of a mechanics problem. The protocol is shown in Table 7-5. This protocol is less detailed than the others described in this chapter because of the more open problem format. The interviewer needs to tailor the questions based on how the student chooses to solve the problem. The questions are guided by the need to clarify the students’ approach, their conceptual understanding, and their expectations. One interesting observation from using this protocol is that if the students do not choose to make up and substitute values for the problem, they will usually get stuck. What they do when they get stuck often demonstrates the accessibility and extent of their physics knowledge.

Table 7-5. Two-Rock Problem protocol: This problem was used by Hammer in his dissertation research discussed in chapter 2.

<table>
<thead>
<tr>
<th>Table 7-5. Two-Rock Problem protocol: This problem was used by Hammer in his dissertation research discussed in chapter 2.</th>
</tr>
</thead>
</table>

Explain what a think aloud interview is and go over the handout with the student.

**TWO ROCK PROBLEM:** Two rocks are thrown with a speed of \( v_0 \) from a cliff of height \( h \). One is thrown horizontally and the other straight down. Which one hits the ground first and which one hits with greater speed?
Give the above problem to the students and ask the student to read the problem aloud. Then ask the student what he thinks the answers are to the two parts of the problem and why. After the student has answered, then ask the student to demonstrate that their answer is correct by writing a solution on paper or on the board.

Questions
Which questions are asked depends on how the student goes about solving the problem. Ask questions that clarify how the student is thinking about the concepts and equations that are being used in the solution. The questions asked during the interviews include:

- What is the difference between speed and velocity?
- Are work and energy vectors?
- Is that the velocity vector or a component of the velocity?
- What does conservation of energy mean and how is useful?
- Where do the kinematic equations come from? Can you derive them?

[Some students may be uncomfortable working with symbolic equations. Student may substitute numbers in their solution at any point, but they should be asked how their answers would change if one of the numbers was doubled or halved.]

Interview continues until the student solves the problem or the interviewer calls time.
The two-rock problem protocol is an interesting interview task because it is simple to visualize yet difficult to solve for the students and involves many areas of mechanics. Since I am interested in more than just the student’s problem solving skills, this protocol combines features from both demonstration- and problem-style interviews to study conceptual understand and expectations. The results from this protocol are discussed in chapters 9 & 10.

IV. LIMITATIONS

As stated in the beginning of the chapter, interviews are the most effective tool for studying students thinking and reasoning. In interviews one can observe and probe individual students to understand what they are doing and why they are doing it. As we saw in the case of the waves-math study, interviews can provide researchers and instructors with the deeper understanding of what students are doing that is needed to correctly interpret the results of surveys, diagnostics, quizzes, and exams.

However, interviews also have limitations as a research tool. Since this method of evaluation requires extensive resources (mainly time) to conduct, transcribe, and analyze interviews, this method is typically limited to only a small sample (< 10) of students from a given class. This means that means that research studies, that rely solely on interviews, cannot determine how prevalent their findings are for a given class; although, since many of the students share the same difficulties and look at physics in similar ways, a reasonable sample of students will usually give a range of interview responses that are representative of a class.
Another limitation is that interviews are more likely than the other PER assessment methods to cause measurement effects. That is, the interview may cause the subjects to reflect on what they did or even to learn new things. This may lead to changes in the way the subjects learn, think, or reason with regard to physics. This is not usually a problem for single interview, but can be a problem when the interview subjects are interviewed more than once. The interviews can also influence how the subjects respond to class-wide measures such as quizzes, exams, surveys, and diagnostics. One needs to be aware of these potential problems when using interviews in conjunction with other methods of assessment. Effects of this nature can be minimized by restricting the protocol to activities similar to what the subjects already do either in class or on their own.

Another consideration for interviews is the sample of students who participate in interviews. Ideally, one would like a random sample whose views are representative of the class. For most of the interviews conducted as part of this study, we recruited students by just asking for volunteers in class. There is a tendency with this method to recruit the better students from a class because they are more likely to volunteer. In the most recent site visits I have asked instructors specifically to recruit students from the top third, middle third, and bottom third of the class based on their judgment. This provides a more representative student sample. The sample size is also an important factor in determining how representative it is. In this study, sample size varied from 5 to 10 students depending on class size and the availability of the students.

---


Michael Wittmann is another senior graduate student in the Physics Education Research Group at the University of Maryland. His dissertation research is on students’ conceptual and mathematical understanding of waves. He expects to defend in summer 1998.

It should be noted that Ferno was one of the few students who was able to work through and answer the rest of the protocol questions correctly.


See Ref. 6.
PART III. EVALUATION OF RESEARCH-BASED TEACHING METHODS

Chapter 8. Courses, Teaching Methods, and Schools

I. OVERVIEW

While many of the results of student difficulties in introductory physics can be generalized to students in similar contexts, the results of modified instruction and interventions are often dependent on the nature of the institution, the nature of the students, and the specific details of the implementation. This is one reason why when an institution adopts a curriculum or teaching method developed at another school, the results of the adopting institution are often not as impressive as they are at the developing institution. Each of the three research-based teaching methods discussed in this chapter have several factors that can contribute to its success or failure.

For example, the best result on a mechanics concept test using tutorials at University of Maryland was obtained by the instructor who best integrated the ideas of the tutorial into his lectures. Other professors who did not adapt their lectures as much improved their students' scores on conceptual tests but not as much. Any teaching method beyond a single brief intervention has many factors that must be considered. Among them are the following:

1. How well integrated is the course as a whole? Do the various components of the course build on one another? Does the course make use of underlying themes?
2. Are the course goals explicitly stated and adequately reinforced through student assignments and grading?
In order to understand what the research is telling us about what students are learning in the introductory class, it is important to put the results in the context of the school, the teaching method, the details of implementation, and the student population. The purpose of this chapter is to provide these details. To help the reader interpret and make sense of these results in later chapters, a brief description of the schools, teaching methods, and implementations follows. Information on the schools is briefly summarized in Table 8-1 in the next section. A summary of the teaching methods and details on the implementations for each course can be found in Table 8-2. Since not all introductory physics courses cover the same material, topic coverage for each of the courses is summarized in Table 8-3. Table 8-4 offers a quick reference as to what data was collected at each of the ten schools. Note that not all types of data were collected at all schools.

II. SCHOOLS AND STUDENT POPULATIONS

Three types of institutions were used in this study: (a) large public universities, (b) small liberal arts colleges and universities, and (c) community colleges. In this dissertation, I discuss results obtained from the following schools:

1. UMD - University of Maryland (a)  
2. MIN - University of Minnesota (a)  
3. OSU - Ohio State University (a)  
4. CAR - Carroll College (b)  
5. DCK - Dickinson College (b)  
6. DRY - Drury College (b)  
7. MSU - Moorhead State University (b)  
8. NWU - Nebraska Wesleyan University (b)  
9. SKD - Skidmore College (b)
10. PGCC - Prince Georges Community College (c)

A brief description of each school including total number of full-time undergraduate students, student SAT/ACT scores, and selectivity is given below in Table 8-1. Note that the most of the schools are easily typed by the size of the undergraduate student population with the exception of Moorhead State University which is a moderately sized state school. However, because of their small class sizes (see table 8-2) and the lack of an engineering major program, for the purposes of this study they are considered a small liberal arts university.

The three research-based teaching methods discussed in the next session were developed at University of Washington (Tutorials), University of Minnesota (Group Problem Solving), and Dickinson College (Workshop Physics). These schools will be referred to as developing schools. The descriptions of the three research-based curricula describe the implementations at the developing schools. (Note, however, that University of Washington is not one of the schools included in this study.) The other schools listed below using the research-based teaching methods developed at one of these three schools are referred to as adopting schools. These adopting schools include University of Maryland (Tutorials), Ohio State University (Group problem solving), Drury College (Workshop Physics), Moorhead State University (Workshop Physics), Nebraska Wesleyan University (Workshop Physics), and Skidmore College (Workshop Physics). Traditional classes were studied at the University of Maryland, Carroll College, and Prince Georges Community College.
Table 8-1: A description of schools participating in this study

<table>
<thead>
<tr>
<th>School</th>
<th>Type</th>
<th># of Students</th>
<th>SAT Ver/Mth</th>
<th>Selectivity³</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Maryland (UMD)</td>
<td>a</td>
<td>20,344</td>
<td>500/590 M</td>
<td>72</td>
</tr>
<tr>
<td>University of Minnesota (MIN)</td>
<td>a</td>
<td>16,457</td>
<td>570/590 M</td>
<td>77</td>
</tr>
<tr>
<td>Ohio State University (OSU)</td>
<td>a</td>
<td>30,500</td>
<td>540/550 M</td>
<td>66</td>
</tr>
<tr>
<td>Carroll College (CAR)</td>
<td>b</td>
<td>1,542</td>
<td>na</td>
<td>63</td>
</tr>
<tr>
<td>Dickinson College (DCK)</td>
<td>b</td>
<td>1,789</td>
<td>600/590 M</td>
<td>74</td>
</tr>
<tr>
<td>Drury College (DRY)</td>
<td>b</td>
<td>1,221</td>
<td>ACT 25</td>
<td>74</td>
</tr>
<tr>
<td>Moorhead State University (MSU)</td>
<td>b</td>
<td>6,252</td>
<td>ACT 22</td>
<td>67</td>
</tr>
<tr>
<td>Nebraska Wesleyan University (NWU)</td>
<td>b</td>
<td>1,378</td>
<td>ACT 24</td>
<td>72</td>
</tr>
<tr>
<td>Skidmore College (SKD)</td>
<td>b</td>
<td>2,150</td>
<td>610/600 M</td>
<td>76</td>
</tr>
<tr>
<td>Prince Georges Community College (PGCC)</td>
<td>c</td>
<td>34,000</td>
<td>na</td>
<td>na</td>
</tr>
</tbody>
</table>

Types of Schools: (a) Large public research university, (b) small liberal arts college or university, and (c) community college

# of Students: This indicates the total number of full-time undergraduates as of 1995

SAT: Verbal/Math (the M designates the math SAT score)

Selectivity: The selectivity ranking was determined by the 1996 Princeton Review. The ranking is determined by a formula that considers the college’s acceptance rate and the percentage of acceptees who actually enroll as well as the class rank and average test scores of the entering freshmen. Selectivity rankings range from 56 to 100. Selectivity scale as follows:

- 56 - 69  Not Selective
- 70 - 79  Selective
- 80 - 89  Very Selective
- 90 - 100 Mega Selective

There is no selectivity score or college entrance exam scores for the community college. Any student with a high school diploma or equivalent may apply and attend.

Selectivity Average for the nine schools = 70.6 ± 4.8
III. TEACHING METHODS

In this project we studied courses that made use of four different teaching methods:

1. Traditional (TRD)
2. Tutorial (TUT)
3. Group Problem Solving and Problem Solving Labs (GPS)
4. Workshop Physics (WP)

The last 3 teaching methods are research-based curricula that make use of active-learning cooperative-group activities. They vary in both the amount of time spent on group activities (from a minimum of one hour per week in tutorials to a maximum of five hours each week in Workshop Physics) and in the type of group activities (described in detail later in this section). Since we are studying the effects of research-based instructional methods, the traditional teaching method is used as a control. All three of the research-based teaching methods are based on the same philosophy, namely, that students learn more effectively when they are actively engaged: physically, mentally, and socially. In each of the three research methods, the students are given the opportunity to interact with each other in cooperative groups where they discuss and dispute each others ideas. Each research-based method also encourages the “less is more” approach to content coverage currently advocated by many physics educators. All three research-based teaching methods have been in use at the developers’ home institution since 1989 and have been adopted by other schools. All three are described by the developers in detail as sample classes in The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education.\(^5\)

257
Each researched-based teaching method is described in this section as implemented by the developers and includes the motivation for development of the curriculum, details of the implementation, problems and difficulties with the implementation, and a summary of whatever evaluation was performed by the developers. The details of the implementations at each of the schools participating in the study are discussed in the next section.

A. TRADITIONAL INSTRUCTION

1. Motivation

The traditional lecture method is in widespread use for two reasons, tradition and cost effectiveness. This method has been used to teach physics for over one hundred years and almost every college and university physics instructor learned physics this way. It is relatively inexpensive in that one professor can lecture to hundreds of students while the smaller recitation and laboratory sections are taught by multiple, often less expensive and less experienced instructors such as TAs.

2. Implementation

Here, a course using traditional teaching methods means a traditional lecture course with recitation and sometimes laboratory sections. In lecture, the instructor presents material to the students sometimes making use of demonstrations and multimedia. The students are mostly sitting passively, listening (hopefully), taking notes, and occasionally asking questions. In the recitation, the students typically ask questions about their homework assignment which the TA will work through on the chalkboard. In the laboratory, the students perform structured recipe-like experiments to verify
principles given in lecture; then they write up the activity in a lab report. Total class time per week is six to eight hours consisting of three hours of lecture, one to two hours of recitation, and two to three hours of laboratory per week. This method is used at most large universities, many small colleges, and even many community colleges in the United States.

3. Difficulties

For most students, lecture and often recitation are very passive. There is little opportunity for the students to discuss ideas and ask questions with the instructor or anyone else in class. Even in recitation, only a few minutes can be spent on a particular student’s questions. However, students often don’t make use of office hours, their one opportunity to work one on one with an instructor. In one sense, this is fortunate because there usually aren’t enough office hours to work with a large fraction of the students.

These characteristics present a problem because many students need help to build their understanding of introductory physics. In addition, when instructors go over the course material and problem solutions in this method, they often don’t model how to build an expert-like understanding or expert-like problem solving skills. While physics faculty often successfully master their field with classes taught in this style, there are two points to consider. One, most of them had three iterations of basic physics courses. And two, they did not learn physics by just repeating what was lectured. At some point in their undergraduate or graduate careers, they sat down with the material and made sense of it for themselves.
The traditional lecture style of instruction also encourages the view that learning science is learning the facts that experts know about the world rather than the view that learning science is the process of trying to make sense of the world.

4. Evaluation

Anecdotal evidence from observations, conversations with faculty, and the research literature strongly suggest that less than half (perhaps only 20%) of the students are effectively achieving the main goal of a useful, functional knowledge of physics in calculus-based introductory classes taught by this method. Some students succeed by memorizing and solving problems by algorithmic pattern matching. The evidence for this is discussed in more detail in chapter 2. The three most general results from PER with implications for teaching are the following:

1. Students bring their own ideas of physics, learning, mathematics, and problem solving into the introductory class based on their own experiences. This can have a strong and often negative effect on what students learn in traditional introductory physics classes.

2. Curriculum carefully designed to take students from where they are to where instructors would like them to be can be effective for addressing these student ideas and making progress towards achieving our primary goal as stated above.

3. Since most students do not have the mental tools to construct their own understanding from the traditional course structure, structured active learning (mentally active) activities with peer interaction in class can be effective in helping students build a better functional understanding of physics. This is a common element of most research-based physics curricula.

Since in most large universities it is impractical to change the course format to small classes, some research-based curricula, including tutorials and group problem solving, are variations on the lecture format to produce more active-learning in large classes.
B. Tutorials

1. Motivation

Many different schools use the word tutorials to describe certain course activities. The tutorial method described below was developed by Lillian C. McDermott and the Physics Education Group at the University of Washington (UW) to improve student understanding of fundamental physics concepts in a cost-effective manner within the traditional large lecture structure. The tutorials grew out of the group’s research on students’ common sense beliefs and their work developing the Physics by Inquiry curriculum for pre-service and in-service K-12 teachers.

2. Implementation

These UW tutorials have the following components:

1. A 10 minute ungraded "pretest" is given in lecture once a week. This test asks qualitative conceptual questions about the subject to be covered in tutorial the following week. Often the material covered in the pretest has already been covered in lectures and homework assignments. Students receive points for taking the pretest but not for the correctness of their responses. The pretests play two roles. One, they help focus the students’ attention on issues that will be discussed in tutorial the following week. And two, the pretests give an indication of student thinking and difficulties before the tutorial.

2. The teaching assistants and faculty involved participate in a 1.5 hour weekly training session. In the training session they take the pretest, go over both the student responses to the pretests, and then go over the tutorial to be used in the coming week. The emphasis of the discussion on the tutorial is on developing appropriate questions to ask the students to illuminate their thinking and lead them towards a physics point of view.

3. A one hour (50 minute) tutorial session replaces the traditional problem-solving recitation section. Students work together in groups of three or four and answer questions on a worksheet that guides them through building qualitative reasoning on a fundamental concept. At least two teaching assistants serve as facilitators in each tutorial section, asking leading questions in a semi-Socratic dialog to help the students work through their difficulties by encouraging them to think. The students’ worksheets are not collected. The students select their own group with little or no intervention by the TAs.
4. Students have a brief qualitative homework assignment in which they explain their reasoning. This is a part of their weekly homework which also includes problems assigned from the text. No solutions of tutorial homework are made available to the students.

5. At least one question emphasizing material from tutorials is asked on each examination.

At the University of Washington, tutorial worksheets are developed over a period of many years through an iterative cycle of research/curriculum-development/instruction. The tutorials often make use of "cognitive conflict." In this approach, situations are presented which trigger the common student conceptual difficulties revealed by research. After the student difficulty is triggered, a situation is presented where the difficulty brings about a contradiction with what the students have been taught. The facilitators then help those students who show the predicted difficulties work through their ideas themselves. McDermott refers to this process as elicit/confront/resolve. The facilitators are mostly graduate and undergraduate TAs who receive no special training prior to their assignment to teach tutorials. The tutorial program is administered by the Physics Education Group. Note that lecturers may choose not to be facilitators or to participate in the weekly training meeting. If so, the tutorials have no adverse impact on instructor time outside of the weekly ten minute pre-test during lecture. However, the instructors are required to include at least one tutorial problem written by the Physics Education Group on each exam.

To address more difficult concepts in the introductory sequence, cover more concepts, and make use of recurring themes, at University of Washington tutorials are now taught in both recitation and the first hour of laboratory every week of the quarter.
The laboratory tutorials provide a qualitative introduction to the experiment performed later in the class. These lab tutorials do not have homework assigned.

3. Difficulties

The key to effective tutorials is in helping the students form effective groups. An effective group is one where the issues and difficulties are discussed and resolved but the group continues to progress through the tutorial. Some groups go through the tutorial too quickly and don’t develop a good grasp of the tutorial issues. Other groups have the opposite problem. They often get stuck and have difficulty resolving the contradictions that arise. These students can become frustrated when they can not finish the tutorials. One difficulty that has come up is that many students’ have the perception of conceptual understanding and quantitative problem solving as separate dimensions of the introductory course. They do not always see how the concepts learned in tutorials relate to the problems they are being asked to solve on exams and homework. For the facilitators, typically the hardest part of teaching tutorials is learning to listen to what the students are saying, to ask leading questions, to not tell them the answers, and to know when to just listen and leave well enough alone.

4. Internal evaluation

At University of Washington, individual tutorials are evaluated through classroom observations, tests, course examinations, and post-tests administered one or more quarters after the relevant course was completed. Before full implementation of tutorials, course examinations and post-tests were administered to calculus-based classes with and without tutorials taught in parallel. The tutorial students did “markedly better”
on qualitative and quantitative problems on course exams.\textsuperscript{11} They also did better on post-tests on circuits and Atwood’s machines. In several cases the tutorial students did as well as graduate students who were given the qualitative problems on qualifier exams.

After going to full implementation, new tutorials are evaluated by exam problems written specifically to test concepts emphasized by that tutorial. The problems are given to classes with and without that tutorial. After several cycles of development for a given tutorial, students who take the introductory course with tutorials do significantly better on the exam problems.\textsuperscript{12} A sample of students are interviewed to probe more deeply into their understanding of the key concepts and validate the problems.

C. Group Problem Solving and Problem Solving Labs

1. Motivation

The Group Problem Solving approach was developed by Patricia Heller and the Physics Education Group at University of Minnesota.\textsuperscript{13} Since a primary goal of instruction in the introductory physics course is to help students build a good functional understanding of physics that they can use to solve problems in new contexts, the Minnesota group focused on problem solving instead of conceptual understanding. Their approach is to use cooperative group activities that work explicitly on building expert problem solving skills. This addresses one of the difficulties in traditional physics instruction and in tutorials; namely, that students in introductory courses often consider problem solving and conceptual understanding to be independent. (See chapter two for more detail).\textsuperscript{14}
They developed this approach for two reasons. One, cooperative group problem solving has been shown to be an effective technique for developing expert problem solving skills, and two, of the recommended techniques for helping students become better problem solvers, cooperative grouping places the least demand on the instructors, in this case graduate teaching assistants who require and receive minimal training to implement this approach.

Although the group problem solving method uses the same format as the traditional large lecture course, the curriculum changes all three aspects of the course at University of Minnesota: lecture, recitation, and lab. The course goals, as established by surveying the departments served by the introductory physics sequences, are for students to learn:

1. the fundamental principles of physics,
2. general quantitative and qualitative problem solving skills that they can apply to new situations, and
3. to overcome their misconceptions (common sense beliefs) about the behavior of the physical world.

Note that these goals are very similar to the main course goals discussed in chapter 1.

2. Implementation

The three components of the course are coordinated to cover the material coherently by a course team consisting of the lecturer and the TAs teaching the associated labs and recitations. The course team meets biweekly to brief the TAs on the direction of the lectures, to give feedback to the lecturer, to decide on problems and course emphasis for the next two weeks, and to discuss student performance. In addition, all three aspects use and/or support the following strategies:
• use of a story line to determine specific content,
• modeling the construction of knowledge,
• use of multiple contexts for each concept,
• focus on the fundamental principles and concepts,
• use of an explicit problem solving strategy,
• use of realistic context-rich problems, and
• use of testing and grading practices to reinforce desired student behavior.

One of the key elements of the course is that the students are taught an explicit problem solving strategy based on expert problem solving strategies. The problem solving strategy they use was strongly influenced by the work of Frederick Reif and Joan Heller\textsuperscript{19} as well as Alan Shoenfeld’s framework for mathematics problem solving.\textsuperscript{20} The five steps of the prescribed problem solving strategy are listed below:\textsuperscript{21} (A more detailed description of the five step strategy is presented in Table 2-1.)

1. Visualize the problem
2. Describe the physics of the situation (Qualitative physics description)
3. Plan a solution
4. Execute the plan
5. Check and evaluate

The student groups apply this strategy in solving problems in both the recitation and the laboratory. In order for the groups to function properly, the choice of problems is crucial.\textsuperscript{22} The problems need several characteristics to encourage the students to work together to solve the problem.\textsuperscript{23} Namely,

• They need to be challenging enough that a single student cannot solve it, but not so challenging that a group cannot solve it.
• They need to be structured so that the groups can make decisions on how to proceed with the solution.
• They should be relevant to the lives of the students.
They cannot depend on students knowing a trick nor can they be mathematically tedious.

Most ordinary textbook problems are inadequate for this task. Heller *et al.* designed their own complex problems incorporating these characteristics which they call context-rich problems. They are designed to focus students’ attention on the need to use their conceptual knowledge of physics to qualitatively analyze a problem before they begin to manipulate equations. They are essentially short stories that include a reason for calculating some quantity about a real object or event. In addition, context-rich problems may have one or more of the following real world characteristics:

1. The problem statement may not explicitly identify the unknown variable,
2. There may be more information available than is needed to solve the problem,
3. Some information may be missing from the problem statement but may be easily estimated, and
4. Reasonable assumptions may be needed to solve the problem.

An example of a Context Rich Problem is shown in Table 2-2.

Students have the same TA and work in the same groups for both recitation and laboratory. This helps the course coherence as well giving the students more practice in group work. Each section of 18 students is broken into groups of three based on their ranking in the class. Each group has one student each from the top third, the middle third, and the bottom third of the class. The students are reassigned into new groups after each exam, two to three times a quarter.

The operation of the lecture, recitation, and laboratory components of the course are described below. (The following is paraphrased from the University of Minnesota Web pages describing the method and philosophy of their group problem solving approach.)
Lecture: The majority of the lecture time is spent in the traditional manner. However, parts of the presentation explicitly model the construction of scientific knowledge and the prescribed problem solving method. Also, some cooperative group work (using groups of 2-3 students sitting near each other) is used to help the students develop concepts. This is sometimes followed by a short question and answer session. Occasionally, group predictions or answers to questions are written down and collected for grading. This lets the students know that their active involvement is an important part of the class.

Recitation: A typical 50 minute recitation section has three parts: introduction, task, and closure. First, the TA briefly goes over the learning goals for the session. Then the TA passes out the assigned context rich problem and assigns the roles of Manager, Recorder/Checker, and Skeptic to the three members of each group. The students have 30 minutes to complete the problem in their groups. The TA observes the groups and intervenes only when no progress is being made by a group or when the students have drifted from their roles. At the end of the session, the TA begins a class-wide discussion on the problem by randomly calling on one member from each group to write their solution on the board. The similarities and differences of the solutions are then discussed. Then the students are given five minutes to evaluate how they worked together and what they could do to improve next time. Students are given a complete written solution to the class problem at the end of the session. Part of each exam is a group problem that is worked in the recitation section.

Laboratories: The laboratories are coordinated with the other parts of the course to address the same content at the same time. The labs are not cookbook, verification labs. The laboratory problems are designed to allow students to apply the problem solving strategy to concrete situations and to help them confront their common sense beliefs. The learning process of the labs can be described as predict/explore/measure/explain. The lab manual is divided into 4 two to three week units, an equipment appendix, and five technique appendices. Each unit is comprised of an introduction page and several related problems. The lab manual contains no theory or background information on the experiments and few specific directions. This is intentional to emphasize that the laboratory is an integral part of the entire course. The write up for each problem refers to the relevant sections in the textbook. A computer check out is used to make sure that each students has a basic understanding of the necessary theory before coming to class.

To focus students’ group discussions on the physics of the situation, the students are required to qualitatively analyze the situation and make group predictions about all measurements before they begin data collection and quantitative analysis. The student groups must decide what data to collect, how the data should be collected, and how the data should be analyzed to solve the experimental problem. The purpose of this is to get the students to make an
intellectual commitment to the lab, not to make sure the students know the right answers at this point.

The lab format is similar to that of the discussion section: introduction, task, and closure. The main difference is that there is no set number of problems to complete; although, the goal is for each group to complete at least 2 problems in the two hour period. The students have the opportunity to return to a problem if their measurements conflict with their predictions. The role of the TAs is to coach the student groups through difficulties and weaknesses.

There are two kinds of lab problems, quantitative and qualitative. The quantitative problems require students to create a mathematical expression that they feel describes the system being investigated. The qualitative or ‘exploratory’ problems require students to use their intuition to predict how the system being investigated behaves. The labs do not currently use any computer data acquisition or analysis; however, the Physics Education Group at University of Minnesota has recently begun developing problem solving labs that use MBL tools and video analysis. They were scheduled to begin experimental implementation in the 1997 Summer quarter.

3. Difficulties

There are three major difficulties in implementing the Group Problem Solving method. First, the method demands additional time from the lecturers to manage, coordinate, and observe the TAs. Second, the TAs must be educated in the story line of the course, students’ common sense belief and everyday use of physics language, the problem solving strategy, cooperative group learning and their role as coaches, and constructive grading practices. Thirty hours of pre-course training are needed for new TAs to be effective and comfortable in their role. In addition, each new TA is assigned a mentor TA who observes them in class and gives feedback. Third, as with the other research-based teaching methods, both the lecturers and the TAs must break the cycle of teaching-as taught. They must be aware of the course structure and strategy as well as the student difficulties while preparing to teach this way. In particular, the lecturer must use this awareness in modeling the construction of knowledge through a story line and
modeling problem solving in lecture. Also, as with tutorials the TAs must learn to guide and coach in a semi-Socratic manner similar to that used in tutorials and not just tell the students how to do it right.

4. **Internal evaluation**

An evaluation of this curriculum by the Minnesota group is in preparation.\(^{25}\) However, Heller, Keith, and Anderson investigated the effects of this curriculum on the problem solving performance of students in an experimental section of the two-quarter algebra/trig based introductory course at University of Minnesota.\(^{26}\) First, they developed and validated a rating scheme for problems to determine relative difficulty. Then they developed a scoring scheme to evaluate the student solutions. They defined ‘better’ student solutions as those exhibiting following six expert solution characteristics: evidence of conceptual understanding, usefulness of physics description, match of equations with physics description, reasonableness of plan, logical presentation, and appropriate mathematics.

They studied the student exam problem solutions for a single two-quarter sequence. Each exam had two parts: first, a context-rich problem to be solved in cooperative groups in the recitation, and second, a short qualitative problem and two context-rich problems to be solved individually in the lecture period the following day. The students received a solution to the group problem at the end of the group exam. In the first part of the study they studied individual and group student solutions of problems where the individual problem on the exam was of equal or of slightly less difficulty than the group problem on the same exam.
Using their expert solution scoring scheme, they found that that average score on the solutions to the group problems was more than three $\sigma$’s better than the individual solutions of the best student in each group as determined by exam grade. When the scores were broken down by category, the biggest differences were in the categories of evidence of conceptual understanding, usefulness of physics description, and the matching of equations with the physics description. By analyzing student solution scores over time, they also found evidence that top third, middle third, and bottom third of students improved at roughly the same rate in all categories except evidence of correct conceptual understanding.

Heller et al. also compared the problem solving skill of the experimental section with a traditional section. Questionnaire results indicated that the students in the two sections had similar backgrounds and characteristics. Since even the easy context-rich problems were judged by the instructors of the traditional section to be too difficult for their students, two standard problems from the traditional section’s final were used in the experimental section’s final exam. The student solutions from both classes were evaluated using the expert characteristic scoring scheme described above. The students in the experimental class taught with group problem solving had an average score more than three standard deviations above that of the traditional section on both problems. Note that this does not reflect the numbers of students in either class who got the problems right; only that the students who were taught an explicit problem solving strategy as described wrote solutions that had more characteristics of expert problem solvers than the traditional class.
D. Workshop Physics

1. Motivation

The first two research-based methods improve instruction by adapting the structure of the traditional large lecture class to make use of cooperative group activities while keeping the large lecture. But is the large lecture format, even with the modifications described above, the best way to teach physics? Priscilla Laws et al. decided to try another way. She and her colleagues at Dickinson College developed Workshop Physics, an activity-based laboratory curriculum. The Workshop Physics materials were developed from the outcomes of the available research in science education. The primary goal of the designers for this curriculum was to help students acquire transferable skills of scientific inquiry based on real experiences. More specifically as stated in her recent project for disseminating Workshop Physics, the goal of Workshop Physics is

“… to enable students to:

- construct conceptual models of phenomena and relate these to mathematical models;
- learn enough scientific literacy to learn without formal instruction;
- develop proficiency with computers and other research tools;
- appreciate science and want to learn more; and
- engage in the further study of science.”

2. Implementation

The WP curriculum makes heavy use of spreadsheet, MBL, digital video analysis, and other integrated computer tools as well as devices that allow students to experience motions and forces with their own bodies (kinesthetic physics). The use of MBL and
digitized video allows the students to see graphical representations of physical systems in real time and to see how changing the conditions of an experiment affect the graph. Spreadsheets are used to create mathematical models that can be compared with the digitized data from experiments. The students meet for six hours a week in a laboratory classroom. The instructors still lecture at the beginning of each class, approximately one hour out of six per week, but the bulk of the time in class is spent performing and analyzing guided-discovery experiments working in groups of two to four students each. Part of the lecture time is spent going over homework problems. The course material is broken up into weekly units that have four parts:

1. exploration of the students preconceptions,
2. qualitative observations,
3. development of definitions and mathematical models, and
4. quantitative experiments centered on the mathematical models.

In addition to their in-class laboratory activities, the students also do homework problems out of a traditional text. Textbook-style problems are included as part of each exam. Students are allowed to use their activity guide on exams. The activity guide is a combination textbook, laboratory manual, and notebook. In addition to their weekly homework and classroom activities, the students are required to do a term physics project involving video analysis each semester. Past student projects have included the physics of Michael Jordan’s lay-up and an analysis of cartoon motion.

Because of the additional time needed to cover material with this method, it is not possible to cover the same amount of material that a traditional course would cover. The Workshop Physics course\textsuperscript{31} covers about 25% less material than was covered previously in the traditional one year calculus-based introductory physics course at
Dickinson College. Each Workshop Physics institution makes its own decisions as to what material is covered and what is left out. At Dickinson College for example, they removed waves, AC circuits, optics, relativity, and quantum from their introductory course. However, they did include some contemporary physics topics like chaos, radon monitoring, and digital electronics.

The classroom is specially designed to be conducive to group activities, classroom discussion, and demonstrations. A typical class size is 24 students. Each class has an instructor and an undergraduate TA (UTA) available during group work to listen and to help but not to give answers. The UTAs and the Workshop Physics classroom are available to the students every weekday evening. The UTAs are selected from students who have previously completed the Workshop Physics course. They receive no other training. They wander around the room while students are doing activities and give hints or suggestions if the students are very frustrated. The UTAs also note completion of activities from the Activity Guide. Other undergraduates are hired to grade for the course.

In the past two years at Dickinson College, they have begun assigning specially designed homework problems that incorporate the ideas and methods developed in class. Traditional textbook problems were not providing good reinforcement of class activities. According to Laws,\textsuperscript{32}

The text problems are too narrow in scope. Students don’t see how they relate to activities in class and many of them [text problems] allow for thoughtless plug and chug. We like to assign extended problems that either have a context that extends in-class activities or something interesting and usually real world. Often the assignments are many part problems that include conceptual questions and involve mathematical
analysis at the end [of the problem] after the students have thought about
the probable behavior of the system from a qualitative perspective.

3. Difficulties

The Workshop Physics method at Dickinson College is very resource intensive in
terms of equipment and instructors. The curriculum makes heavy use of digitized video,
computers and sensor probes in addition to the standard laboratory equipment. There
are at least two facilitators, the instructor and an undergraduate TA, in class at all times.
Since the course is laboratory-based instead of lecture-based, class size is very limited.
(Although, note that Jack Wilson at RPI has developed a similar introductory course
called Studio Physics that can accommodate 50 students per class. 33)

The role of the instructor is important to the success of the course. The
instructor needs to stay out of the way of the materials and not tell the students the
answers. Poor scores on a concept test from a Workshop Physics class were traced to
an instructor who basically used this format to lecture and teach a more traditional
course. The instructor’s role in WP is to be more of a mentor, coach, and intellectual
manager and much less of a traditional lecturer. 34

4. Internal evaluation

The evaluation of Workshop Physics at Dickinson College uses several concept
tests, exam results, and a survey of student attitudes 35 towards the introductory course.
Students from the introductory course were tested before and after the Workshop
Physics program was implemented at Dickinson. The preliminary results from a 1991
article can be summarized as follows (Note that no follow up report has yet been
written). 36
1. Based on written evaluations, two-thirds of the students in the calculus-based introductory class prefer the workshop approach over what they imagine a lecture approach to be.

2. Based on the results of “conceptual questions,” a greater percentage of students master concepts that are considered difficult to teach because they involve classic student misconceptions.

[These conceptual questions range from multiple choice questions from concept tests such as the FMCE to qualitative exam problems from University of Washington similar to those described in the description of tutorials above.]

3. Performance of Workshop Physics students in upper-level physics classes and in solving traditional textbook problems is as good as that of students who had the traditional lecture course. Here, student performance is judged by scores, grades, and instructor impressions.

4. From observations by instructors and off-campus observers, the students who complete Workshop Physics are more comfortable working in a laboratory setting and more comfortable working with computers.

5. However, some students complain that Workshop Physics is too complex and demands too much time. It should be noted that the Workshop Physics students in this study worked an average of seven hours a week outside of class on physics. However, in a poll conducted by Laws the average time spent outside of class for physics at 16 other colleges was 6.5 hours.

6. A small percentage of students thoroughly dislike the Workshop Physics approach. They tend to be juniors and seniors who have been successful with more traditional instruction.37

[From my own observations and conversations with peers in my field, it seems that whenever an active learning format is implemented, it takes students anywhere from weeks to months to adapt and appreciate it.38 About 10-20% of the students never seem to make this transition.]

One additional note is that of the three research-based curricula in this dissertation, Workshop Physics has been the most widely adopted. In fact, the chemistry and mathematics departments at Dickinson College have adopted this approach for introductory courses in chemistry, statistics, and calculus. However, few of the other schools implementing Workshop Physics have achieved the same degree of success as Dickinson College. Details on differences in implementations for adopting schools are described in the next section.
IV. COURSES AND IMPLEMENTATIONS

Even if two schools use the same curriculum, the implementation may differ significantly due to differences in philosophy, resources, or student populations. Even two traditional courses at different schools may have subtle differences that would affect a study of this nature. The previous section described the four teaching methods used by introductory classes participating in this investigation. This section will review the implementation of each curriculum at each institution. A summary of the ten classes involved in this study is given in Table 8-2. A list of topics covered by each class broken down by terms is given in Table 8-3.

Special attention is paid to how the first-term mechanics section of introductory physics is taught since the concept tests described in chapter 4 and reported on in chapter 9 look primarily at students’ understanding of Newtonian ideas of force.

A. Traditional

1. University of Maryland (UMD)

The introductory course at Maryland for engineering students is typical of such courses offered across the country with a few special characteristics. It is a three-semester sequence with no laboratory in the first semester. The second semester laboratory begins with mechanics experiments based on material from the first semester course. Note that both physics majors and pre-medicine majors are encouraged to take other introductory sequences. In the Fall 1996 semester the laboratory in the second and third semesters was increased from two hours to three hours in a move to de-emphasize
the formal lab report and give students class time to finish their lab write-ups in class.

The weekly course format for each semester is as follows:

<table>
<thead>
<tr>
<th></th>
<th>1&lt;sup&gt;st&lt;/sup&gt; Semester</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; Semester</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; Semester</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lectures/wk</td>
<td>2-3</td>
<td>2-3</td>
<td>2-3</td>
</tr>
<tr>
<td>150 min/wk</td>
<td>150 min/wk</td>
<td>150 min/wk</td>
<td>150 min/wk</td>
</tr>
<tr>
<td>1 Recitation/wk</td>
<td>1 Recitation/wk</td>
<td>1 Recitation/wk</td>
<td>1 Recitation/wk</td>
</tr>
<tr>
<td>50 min/wk</td>
<td>50 min/wk</td>
<td>50 min/wk</td>
<td>50 min/wk</td>
</tr>
<tr>
<td>no laboratory</td>
<td>1 laboratory/wk</td>
<td>1 laboratory/wk</td>
<td>1 laboratory/wk</td>
</tr>
<tr>
<td>110-170 min/wk</td>
<td>110-170 min/wk</td>
<td>110-170 min/wk</td>
<td>110-170 min/wk</td>
</tr>
</tbody>
</table>

The size of a lecture varies from about 40 to 170 students depending largely on whether the class is offered on or off sequence and at what time the class is offered. Evening classes were specifically excluded from the study to control for population variation, i.e. a large number of non-traditional students. Two or three midterm exams plus a final exam are given each semester. Students are often given key equations on exams or allowed to bring a reference sheet into the exam. The course textbook is selected by a faculty committee. Homework from the lecture course usually consists of a reading assignment and 10-14 end-of-chapter problems each week. In some semesters the lecturers for a course may confer and assign common homework problems. Homework solutions are usually posted shortly after the assignment is collected.

Recitation and laboratory section size varies from 10-30 students per section. Sections of 20-25 students are typical. The laboratory sections are taught as a separate course. They are the responsibility of a separate laboratory instructor and they are taught by separate laboratory TAs. Students can take any laboratory associated with the course but must take a recitation section associated with a particular lecture instructor.
The recitations are taught by recitation TAs who report to the lecture instructors. Most weeks there is a ten-minute quiz in the recitation section. Attendance is usually poor in recitation sections if quizzes are not given. Outside of class, students can take their questions from the reading, lecture, or homework to their lecture instructor’s office hours or their TA’s office hours. Additional help on homework is available from the Slawsky Clinic. The Slawsky Clinic is run by two retired physicists and offers help to students on specific homework problems.

We studied this sequence from the Fall 1991 semester through the Spring 1997 semester. Different aspects of the sequence were studied at different times during this period. Details on the data collected are given in table 8-4. MPEX and demonstration interviews were held with many students from Spring 1994 through Fall 1996.

2. Prince Georges Community College (PGCC)

The calculus-based introductory sequence at Prince Georges is a three-semester sequence designed to duplicate the introductory physics for engineers’ sequence at Maryland with three exceptions. First, the class size is only 10-25 students. Second, all parts of the course are taught and graded by a single instructor. There are no TAs. Third, although the content coverage of the two courses is the same, the order of the topics is different as shown in table 8-3. For example: The second semester of the traditional course at Prince Georges covers electricity and magnetism while the Maryland course cover vibrations, waves, sound, fluids, heat & temperature, and electricity with magnetism covered in the third semester. Also, they have not yet followed Maryland’s increase in lab time from 110 to 170 minutes per week in the second and third semesters.
There were no MBL activities or computer-assisted demonstrations at PGCC during this study.

Classes from Prince Georges Community College participated in this study from the 1994 fall semester through the 1996 spring semester. Because of the small class sizes, both day and evening students were studied. I interviewed two students in the 1994-5 academic year.

3. Carroll College (CAR)

The calculus-based introductory sequence at Carroll College is a typical, traditional two-semester course primarily designed for science majors. The course is unusual in that it meets for four hours of lecture and three hours of laboratory each week. There is no recitation section *per se*. The lecture is taught in a traditional format with demonstrations occurring on a regular basis. The instructor makes regular use of the computer for presentations, simulations, and demonstrations.

Classes from Carroll College participated in the study during the 1995-96 and 1996-97 school years. The instructors believe that the students in these classes were typical for this sequence. No interviews were conducted.

B. Tutorials - University of Maryland (UMD)

In the tutorial method as implemented at University of Maryland, the main difference with traditional lecture classes is that tutorials replace the traditional recitation. Since the fall semester of 1993, Maryland has been implementing tutorials in one or more of the classes in the engineering sequence each semester. The tutorials are run by the Physics Education Research Group (PERG) under Redish and Steinberg.
Although initially the tutorials were implemented in very much the same style as described in the previous section at University of Washington, the current implementation at Maryland differs in some ways from the implementation at University of Washington.

To begin with, Maryland has no laboratory tutorials. Instructors are encouraged to participate in the weekly training meeting as well as in the tutorials themselves. Undergraduate TAs are sometimes used as facilitators to make sure that there are at least two facilitators in each section. In addition, the PER group at Maryland has developed tutorials that make use of microcomputer based laboratory (MBL) equipment, digitized video (for waves and sound), and computer simulations (primarily using M.U.P.P.E.T\textsuperscript{39} and EM Fields\textsuperscript{40}) as well as tutorials that focus on mathematical ideas. In the first semester course, which covers mechanics, the University of Washington tutorials are supplemented with 3 MBL tutorials.

These three MBL tutorials assist students with the concepts of instantaneous velocity, Newton’s 2\textsuperscript{nd} Law, and Newton’s 3\textsuperscript{rd} Law. All three tutorials were created and implemented by us in the 1994-95 school year. Like the University of Washington Tutorials, each has undergone successive refinement. Our MBL equipment is a computer connected to a universal laboratory interface box (ULI) with a sonic ranger and two force probes.\textsuperscript{41} These tutorials are included in Appendix D.

Redish and I wrote the instantaneous velocity tutorial based directly on the MBL activities developed by Thornton and Sokoloff labs in \textit{Tools for Scientific Thinking}.\textsuperscript{42} We extracted from their velocity labs what we considered the essential elements, following the guidance in their paper.\textsuperscript{43} In the tutorial, students walk in front of a sonic
ranger which provides immediate feedback and reduces data-collection drudgery. In the tutorial, students use their own bodies to

1. familiarize themselves with the equipment by creating a series of position graphs;
2. create a series of simple velocity graphs;
3. match a given complex velocity graph.\textsuperscript{44}

In each case, the students work together in groups of three or four. They discuss and make predictions of what the graph will look like or how they have to move in order to produce the desired result and they write these predictions on their worksheets. The entire activity is easily completed in one fifty-minute period.

I wrote the Newton’s second law tutorial based on activities developed by Morse,\textsuperscript{45} Laws, Thornton, and Sokoloff.\textsuperscript{34} Again, I extracted from their MBL labs what I considered the essential elements. In this tutorial, the students use the motion detector to analyze velocity-time graphs of the motion of a fan cart to determine its acceleration in various situations increasing in complexity. The students then compare the measured acceleration with predictions from an analysis of the forces from free-body diagrams.

I wrote the Newton’s third law tutorial based on suggestions of Laws, Thornton and Sokoloff.\textsuperscript{46} Newton's third law is explored by having students connect the force probes to two low-friction carts and observe the result of their interaction. The apparatus is sketched in Fig. 8-1.\textsuperscript{47}

In the tutorial, the students do the following:

1. first psychologically calibrate the force probe by pushing and pulling on it and watching the result on the computer screen;
2. predict the relative size of forces for a light car pushing a heavy truck;
3. predict and observe the forces two identical carts exert on each other when one pushes the other;

4. predict and observe the forces two carts exert on each other when one is weighted with iron blocks;

5. predict and observe the forces two identical carts exert on each other when one collides with the other;

6. predict and observe the forces two carts exert on each other when one collides with a second weighted with iron blocks.

In addition, the students are asked to draw free-body diagrams and use them in their predictions. Again, this activity is easily completed in one fifty-minute period. This sequence taught with tutorials was studied from the Fall 1993 semester through the Spring 1997 semester. However, like the University of Maryland traditional classes, different aspects of the sequence were studied at different times during this period. Details on the collected data are available in table 8-4. MPEX and demonstration interviews were held with many students from Spring 1994 through Fall 1996.
C. Group Problem Solving and Problem Solving Labs

1. University of Minnesota (MIN)

The University of Minnesota has developed and implemented the GPS curriculum in their calculus-based introductory sequence. The GPS recitations and Problem Solving Labs were run by the physics education group at Minnesota from fall 1993 until spring 1996. In the fall 1996 semester responsibility for the course passed to the physics department. Unfortunately, several of the professors were new to the project and did not adequately supervise the TAs. As a result, the students were not encouraged to work in cooperative groups in recitations or labs, and the problems assigned did not follow the guidelines for context rich problems. Also, the mentor TAs afforded less time to supervise the new TAs. Based on their own evaluation, the physics education group considers the Fall 1996 to Spring 1997 sequence to have been essentially taught with traditional instruction.

The sequence was studied using pre/post diagnostic tests from Fall quarter 1994 through Spring 1997. Details on the data collected can be found in table 8-4. Note that only FCI and MPEX was collected from the classes at the University of Minnesota.

2. The Ohio State University (OSU)

The engineering sequence at OSU is a three-quarter sequence with mechanics in the first quarter, electricity and magnetism in the second quarter, and oscillations, waves, optics & modern physics in the third quarter. In the 1994-95 school year, Allen Van Heuvelen began implementing the GPS approach into this formerly traditional introductory physics sequence. Van Heuvelen is a member of the physics education
group (PEG) at OSU and is well known for his own efforts in curriculum development with introductory physics courses. Because the enrollment in the sequence is approximately 900 students a quarter and because the physics education group was just starting at that time, these changes were implemented by the physics department with regular graduate TAs. There was no special effort to put PEG graduate students as TAs for this course, but there was extensive TA preparation for the graduate students in the summer training program. A large portion of this program is given to surveying the literature on physics education research and physics curriculum development as well as providing TA training. The GPS implementation at OSU is based heavily on the implementation at MIN; however, there are significant differences in all three components of the course: lecture, recitation and lab.

**Lecture:** (3 hours per week) The lecturers teach in their own way. Some instructors give very traditional lectures; others make substantial use of interactive lecture demonstrations (discussed previously in chapter 2). No attempt is made to model problem solving in lecture nor is any explicit problem solving strategy discussed. Homework is graded on how much of the assignment is attempted. Solutions are posted shortly after the homework is turned in. There are two midterm exams and a final exam every quarter.

**Recitation:** (two 48 minute recitations per week) This is the only sequence in the study that has two recitation sections per week. The first weekly recitation is used for group problem solving. Students are assigned into groups at random and the student groups stay the same though out the quarter. Group roles are not assigned and there is no group evaluation of how to improve the group dynamics. The problems come from either ALPS Kits or context rich problems from Minnesota and OSU. The TAs try to go over the group problems at the end of the class but if there is not enough time the problem will be discussed in the second weekly recitation. Students do a group exam problem in recitation either before or after each midterm and final exam.

The second weekly recitation is mainly for going over homework interactively. The TA uses semi-Socratic dialog to get the students to discuss the problem and work through it as a class. The last 18 minutes of the recitation are used for a quiz based on the homework. The TAs grade student homework during the quiz by checking to see if two problems on the assignment were done
before passing the homework back at the end of the recitation. The students do not receive any feedback on the homework.

Each recitation section is associated with a particular lecture class. The recitation TAs meet weekly with their respective lecture instructor for an hour and a half to discuss how the class is going, where the students are having difficulties, and what problems will be used for group work and the quiz that week. The TAs spend one of their two office hours each week in the tutoring room which is open to students forty hours a week. While there is no discussion of problem solving strategies in the recitation, some TAs do go over strategies in recitation. Also, some of the ALPs problems lead the student through a strategy that emphasizes physics representations and an evaluation of the solution.

Laboratory: (1 two hour lab per week) The laboratory is conducted with cooperative groups performing problem solving laboratories similar to those in at Minnesota and guided investigative labs similar to Workshop Physics activities. Like Minnesota, in the PSL activities the OSU students are asked to design an experiment and make a prediction before they do the experiment. The student groups do the labs on worksheets and are checked out before they leave. The student groups are not the same as those in recitation and students from any lecture section may take any laboratory section. There is no effort made to coordinate the lab with a particular lecture section. Some laboratories are run by undergraduate TAs (UTAs) who have not gone through the 10 week summer program. The UTAs are mainly trained by a graduate super TA, who runs the laboratory portion of the course.

Occasionally experimental classes are taught by faculty from the physics education group to try out new ideas. One area being re-evaluated is the use of the GIL activities in laboratory. Van Heuvelen comments that they do not seem to work as well as they should. The students do the activities but they do not seem to think deeply on what is learned. This parallels my own experience with similar materials at PGC. While Van Heuvelen is satisfied with the active-engagement activities for the first two quarters of the sequence, they are still developing the activities for the third quarter of the sequence.

A physics education graduate student at Ohio State commented that not all lecture instructors or TAs are equivalent. Some differ significantly in their attitudes and skills concerning cooperative group activities. Some faculty and TAs use the
cooperative group activities very effectively; others do not. Roughly three-fourths of the lecture faculty seem to want the cooperative group activities to work and support it with group work in lecture. The remainder do minimal or no group work in lecture. In fact, some faculty used the group work in recitation and laboratory only because it was required. In these cases, the group work did not fit well with the rest of the class. This negatively affected both the students’ and TAs’ outlook towards the recitation and the lab.

We collected diagnostic test data at various times from the engineering sequence beginning in the 1995 winter and spring quarters and continuing through the 1996-1997 academic year. I conducted interviews with twenty-five students from all three quarters of this introductory sequence at the end of the winter 1995 quarter.

D. Workshop Physics

1. Dickinson College (DCK)

The calculus-based Workshop Physics course at Dickinson College is a two-semester sequence. Some topics normally covered in a traditional introductory course are not included in the first year course. Waves and optics, for example, are covered in the 2nd year physics course for majors. The details of the implementation of Workshop Physics at Dickinson are described in the previous section. The mechanics modules make heavy use of MBL and spreadsheets.

As this is the only calculus-based introductory physics course offered, only a few of the students in the course are physics majors. The rest of the students are mainly liberal arts, education, and pre-med majors.
We collected pre/post diagnostic test data from the introductory sequence at Dickinson College from the 1994 fall semester through the 1997 spring semester. I interviewed eight students at Dickinson at the end of the 1994 fall semester, eight students at the beginning of the 1995 fall semester, and six students the end of the spring 1996 semester.

2. Drury College (DRY)

During our study, Drury College was is in transition from a traditional lecture format to the Workshop Physics format. They began implementing Workshop Physics in Fall 1995. During their participation in our study, one algebra/trig based class and their only calculus based class were taught in the Workshop format. Two additional algebra/trig courses were taught in parallel in the traditional lecture format. The calculus-based class used the activity guide from Dickinson College with the addition of a unit on optics from a sophomore level physics course at Dickinson.

While the program and equipment are similar to those at Dickinson, there are two major logistical differences. There is only one instructor in the room and the room itself is a traditional lab room with three long benches that is not conducive to cooperative group work or class discussions. The students in the calculus-based course are mainly science and pre-med majors. (However, most pre-meds take the algebra/trig-based course.)

I made a site visit at the end of the Spring 1997 semester to observe Drury’s WP classes and interview several student volunteers. Because of the small size of the calculus-based class (8 students), having only one instructor is not a problem and despite the room layout, the three groups of 3-4 students each appeared to be working well. In
the class I observed, there was a short lecture on the day’s topics and then the students went to work on their activities. At the end of class there was a discussion on what the students had found.

We collected diagnostic test data from one complete sequence during the 1996-97 academic year. Four students from the calculus-based class were interviewed during my site visit.

3. Moorhead State University (MSU)

Moorhead State University had been running Workshop Physics classes for two years before participating in this study during the 1995-6 school year. The implementation was as close to the Dickinson implementation at they could make it. However, there were some differences. Because MSU does not use undergraduate TAs, there was only one instructor was in the classroom at all times. Also, MSU found that they need lab technician support for setup and repair of the lab equipment and the computer network to keep the WP laboratory running smoothly. There were also significant differences between the student population at MSU and the Dickinson students. The students in the MSU class were mostly engineering and chemistry majors and they tend to be older and have more non-academic commitments. In addition, many of them work part-time. Also, Dickinson is selective in its admissions while MSU is not.

We collected diagnostic test data from the full sequence during the 1995-96 academic year. No site visits or interviews were conducted at MSU.
4. Nebraska Wesleyan University (NWU)

NWU has been teaching the Workshop Physics format since fall 1995. They began implementation of Workshop Physics with support from a FIPSE grant headed by Priscilla Laws at Dickinson College to help the implement Workshop Physics at six schools across the country. NWU teaches one introductory sequence that is a hybrid algebra/trig- and calculus-based course. They use the calculus-based Workshop Physics Activity Guide and an algebra/trig-level physics text. Several units from the activity guide were adapted from the algebra-based WP activity guide. Some of these units are still being refined. Calculus is not required for this course. Some topics are covered with dual approaches: “For those of you with calculus, do this, and for those of you who haven’t, do that.” At the end of one of these sections, the class discusses what was learned by both types of groups.

I attended one session of both WP classes during a site visit to NWU at the end of the 1997 spring semester. The room is quite suitable for group work with lab tables on the side and open space in the middle. The students work in groups of three to four. Like Drury, there is only one instructor in the room, but here there are 25 to 30 students per class. This is a problem when several of the groups simultaneously need guidance to proceed. As part of the FIPSE project, Workshop Physics developers including Laws also made site visits to NWU. While they report that the implementation is going well, they did observe some problems. The two classes Laws observed in the 1997 Spring semester seemed to flow well. In her words,

The students worked at a steady pace and engaged in thoughtful collaborative interchange. They identified good questions to discuss with each other as they moved through the materials. However, the students...
seemed less intense than those we have at Dickinson. The benefit of the slower pace was that the students worked through ideas more carefully and enjoyed each other’s company. On the other hand, it has been hard for the faculty here to cover as much material as we do at Dickinson.

Because of the broader than normal background and interests of the students, there has been some difficulty in deciding what topics to include and how much coverage for each topic. Also, while every group has a computer with a ULI, there seemed to be insufficient lab equipment for each group to do some of the activities themselves. In addition, the faculty felt they needed lab technician support (not available at the time) for equipment repair and setup. The equipment problems should be resolved next year, thanks to additional grants.

We collected MPEX and FCI data from the two complete introductory sequences taught between fall 1995 and spring 1996. I interviewed 10 student volunteers combined from both classes during my site visit.

5. Skidmore College (SKD)

Skidmore College is a liberal arts college in upstate New York. They began implementing Workshop Physics in the fall 1996 semester in a new specially designed laboratory space. Skidmore is also one of the six school funded by FIPSE grant for the dissemination of Workshop Physics. The course is designed primarily for science majors, primarily biology, chemistry, and mathematics. The implementation is similar to Dickinson’s with the following exceptions:

1. As at DRY, MSU, and NWU, no undergraduate TAs are used at Skidmore. However, a department staff member set up the equipment and serves as a TA so two instructors were available in both classes of 20-25 students each.

2. They use the Dickinson activity guide but they cover only 70% of the content. Laws determined that this is due to a combination of Skidmore’s
semester being shorter, the students working at a slower pace, and fewer home problems being assigned each week. NWU has similar difficulties with covering material.

3. SKD did not have a lot of the equipment they needed and what they had often did not work in class.

The Skidmore instructors commented that some of the upper-division students, especially pre-meds, were very resistant to the Workshop Physics approach. They found it too slow and too frustrating. Laws has observed this resistance of upper-
Table 8-2: Description of Introductory Calculus-Based Classes Studied  (na indicates this course does not have lecture or recitation)

<table>
<thead>
<tr>
<th>Institution</th>
<th>Teaching Method</th>
<th>Class size Lecture (rec / lab)</th>
<th>Lecture</th>
<th>Recitation</th>
<th>Laboratory</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Maryland College Park, MD</td>
<td>Traditional</td>
<td>50-150 (25 / 25)</td>
<td>3 hrs/wk by faculty</td>
<td>1 hr/wk by 1 TA</td>
<td>2-3 hrs/wk by TA</td>
<td>3 semester sequence with no lab in the first semester</td>
</tr>
<tr>
<td>University of Maryland College Park, MD</td>
<td>Tutorials</td>
<td>50-150 (25 / 25)</td>
<td>3 hrs/wk by faculty</td>
<td>1 hr/wk by 2 TAs</td>
<td>2-3 hrs/wk by 1 TA</td>
<td>Similar to UMD traditional with tutorials replacing recitation</td>
</tr>
<tr>
<td>University of Minnesota Minneapolis, MN</td>
<td>Group Problem Solving</td>
<td>150-200 (18 / 18)</td>
<td>3 hrs/wk by faculty</td>
<td>1 hr/wk by 1 TA</td>
<td>2 hrs/wk by 1 TA</td>
<td>3 quarter sequence / Integrated themes / Large Lecture with GPS in recitation and PSL in labs</td>
</tr>
<tr>
<td>Ohio State University, Columbus, OH</td>
<td>Group Problem Solving</td>
<td>150-180 (25 / 25)</td>
<td>3 hrs/wk by faculty</td>
<td>2 hrs/wk by 1 TA</td>
<td>2 hrs/wk by 1 TA</td>
<td>3 quarter sequence / Large Lecture with interactive demos/ GPS in recitation and PSL in labs</td>
</tr>
<tr>
<td>Carroll College Waukesha, WI</td>
<td>Traditional</td>
<td>15-20 (na / 16)</td>
<td>4 hrs/wk by faculty</td>
<td>na</td>
<td>3 hrs/wk by Lab Asst.</td>
<td>2 semester Sequence / Will begin using Workshop Physics in P97</td>
</tr>
<tr>
<td>Dickinson College Carlisle, PA</td>
<td>Workshop Physics</td>
<td>na (na / 25)</td>
<td>na</td>
<td>na</td>
<td>6 hrs/wk by faculty &amp; 2 TAs</td>
<td>2 semester sequence with minimal lecture / uses undergraduate TAs</td>
</tr>
<tr>
<td>Drury College Springfield, MO</td>
<td>Workshop Physics</td>
<td>na (na / 10)</td>
<td>na</td>
<td>na</td>
<td>6 hrs/week by faculty</td>
<td>2 semester sequence with minimal lecture</td>
</tr>
<tr>
<td>Moorhead State University Moorhead, MN</td>
<td>Workshop Physics</td>
<td>na (na / 20)</td>
<td>na</td>
<td>na</td>
<td>6 hrs/week by faculty</td>
<td>2 semester sequence with minimal lecture</td>
</tr>
<tr>
<td>Nebraska Wesleyan Univ. Lincoln, NE</td>
<td>Workshop Physics</td>
<td>na (na / 25)</td>
<td>na</td>
<td>na</td>
<td>6 hrs/week by faculty</td>
<td>2 semester sequence with minimal lecture</td>
</tr>
<tr>
<td>Skidmore College (SKD) Saratoga Springs, NY</td>
<td>Workshop Physics</td>
<td>na (na / 25)</td>
<td>na</td>
<td>na</td>
<td>6 hrs/week by faculty</td>
<td>2 semester sequence with minimal lecture</td>
</tr>
<tr>
<td>Prince Georges Community College, MD</td>
<td>Traditional</td>
<td>25 (25 / 25)</td>
<td>3 hrs/wk by faculty</td>
<td>1 hr/wk by faculty</td>
<td>2 hrs/wk by faculty</td>
<td>3 semester sequence with no lab in the first semester</td>
</tr>
</tbody>
</table>
Table 8-3.
Table 8-4: summary of the Data collected for each course at each schools
division pre-meds in Workshop Physics implementations at other schools including Dickinson.\textsuperscript{57} She suggests that these pre-med students often fail to see how physics might apply to medicine and want to be able to get an A with a minimum time investment. We will come back to this issue in chapter 10.\textsuperscript{58}

We collected FCI and MPEX Data from both calculus-based WP classes taught in the 1996 fall semester.

\begin{enumerate}
\item Personal observations from site visits and private communications with Edward F. Redish (UMD), Richard N. Steinberg (UMD), Priscilla Laws (DCK), Tom Foster (MIN), Chris Cooksey (OSU), Pat Cooney (Millersville College), Peter Shaffer (University of Washington), and Gregory Francis (Montana State University, Bozeman).
\item Test score averages are for students who entered that institution in 1995.
\item L.C. McDermott and P.S. Shaffer, Tutorials in Introductory Physics (Preliminary edition) (Prentice Hall, Upper Saddle River NY, 1997); For a description of tutorials as used at the University of Washington and an experimental result using them, see P.S. Shaffer and L.C. McDermott, “Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of an instructional strategy,” Am. J. Phys. 60, 1003-1013 (1992); L.C. McDermott, P.S. Shaffer, and M.D.
\end{enumerate}


11 See Ref. 10

12 See Refs. 7 & 10.


14 There is an indication of this found in two comments frequently heard by physics instructors, “I understand the material, but I just can’t solve the problems,” and , “I can do the physics but I don’t understand ‘the theory’.” Note: Students in introductory courses often refer to the physics concepts and/or formula derivations as ‘the theory’.


17 See Ref. 13.

18 The University of Minnesota Web pages offer a detailed description and course materials from the Group Problem Solving and Problem Solving Labs curriculum. (URL: www.physics.umn.edu/groups/physed/ )


21 See Ref. 15.


23 See Ref. 22.

24 The material in this section comes from Ref. (Heller), private communications with Tom Foster (the senior graduate student in the PER group at University of Minnesota, 1997), and the University of Minnesota Physics Education Research and Development web page at www.physics.umn.edu/groups/physed/.


26 See Ref. 15.

27 See Ref. 15.


29 Private communication with Priscilla Laws, Dickinson College, Fall 1995.


34 Private communication with Priscilla Laws, Dickinson College, Summer 1997.

35 This does not refer to the MPEX survey but another survey that Laws developed and used for her own purposes in the late 1980’s.

36 See Ref. 28.


41 We used Intel 386, 486, and Macintosh SE personal computers. The ULI, sonic ranger, and force probes are from Vernier software, Portland, OR. Only the Motion and Datalogger software that comes bundled with the ULI were needed. Two Pasco carts or their equivalents are also required for the Newton 3 tutorial. A fan assembly was used in the Newton 2 tutorial. The current cost of the required equipment is about $2000 per station (including the computer.)


44 See Ref. 43, Figure 2..


47 See Ref. 43.

48 Private communication with Tom Foster, who ran the laboratory and recitation parts of the curriculum from 1994-5 (Summer 1997).
There are no written materials available describing the implementation of the group problem solving curriculum at OSU. Consequently, the sources for the information on this sequence at OSU are private communications with Alan Van Heuvelen, Ed Adelson, and graduate students in the physics education research group at OSU as well as personal observations during a site visit.


Private communication with Alan Van Heuvelen at the Ohio State University, Summer 1997.

Private communication with Chris Cooksey, a physics education graduate student at the Ohio State University who worked closely with this sequence during Ohio State’s participation in this study, Fall 1997.

See Ref. 53.

Private communications with Priscilla Laws, Dickinson College, Fall 1997.

See Ref. 55.

See Ref. 37.

See Ref. 56.
CHAPTER 9: Conceptual Understanding

I. OVERVIEW

As we discussed in chapter 2, conceptual understanding plays an important role in the way experts solve complex problems. Unfortunately, the examples of interviews and qualitative exam problems in chapter 2 (Mazur\(^1\) and Hammer\(^2\)) as well as the surveys of concept tests discussed in chapter 4 (Hestenes \textit{et al.}\(^3\), Hake\(^4\), Thornton and Sololo\(f^5\)) indicate that many students have little improvement in their understanding of physics concepts after traditional lecture instruction than when they began. There are two main causes for this situation. One is the persistence of students’ common sense beliefs based on the students’ previous real world experiences. The second is the heavy emphasis in many traditional lecture courses on typical end-of-chapter problems that students can often solve without understanding or applying the relevant concepts (see the discussion in chapters 2 & 6). In classes like this, conceptual understanding becomes part of the hidden curriculum, a course goal that is neither explicitly stated nor adequately reinforced through grading.

However, as we saw in chapters 2 and 4, PER indicates that instruction that takes the students’ common sense beliefs into account and provides a mechanism for conceptual change\(^6\) can be effective for improving students’ conceptual understanding of physics. One of the main goals of this dissertation is to see if students taught with the three researched-based curricula discussed in the previous chapter, Tutorials (TUT), Group Problem Solving (GPS), and Workshop Physics (WP), show significant improvement in students’ conceptual understanding of physics. In this chapter, we will look at two aspects of students’ conceptual understanding: how well students learn the
basic concepts as measured by concept tests and how well they use concepts and representations in solving specially-designed qualitative problems.

Research-based multiple-choice concept tests, like the Force Concept Inventory (FCI)\(^7\) and the Force Motion Concept Evaluation (FMCE)\(^8\), have been developed with questions that can trigger and identify students’ common sense beliefs. These concept tests are a good indication of how well a student understands basic concepts in introductory physics. However, this type of understanding is only one part of the understanding needed for students to achieve a good functional understanding of physics.

Students should also be able to use their conceptual understanding in solving complex problems. In addition, they should be able to express their conceptual understanding in the multiple representations often used by expert problem solvers. The studies on physics problem solving discussed in chapter 2 indicate that expert problem solvers make heavy use of concepts and conceptual understanding in their solutions. In particular, they often use a detailed qualitative representation of the problem before applying the relevant mathematical model. This qualitative representation is then used both as a guide to solving the problem and as a means to evaluate the solution. This ability to make use of conceptual knowledge and representations in problem solving is one reason experts can respond more flexibly to new, complex problems than novices. Many students in traditional introductory physics courses lack this skill.

In section II, I present results on students’ understanding of Newton’s laws of motion and force for each of the three research-based curricula as well as for traditional instruction for comparison. To determine students’ understanding of these basic
concepts, I have gathered, processed, and analyzed concept test data for each of the ten schools. Unless otherwise specified, the data presented in section II is “matched.” Only responses from students who took the tests both at the beginning and at the end of the semester or quarter of the sequence are presented in the tables and graphs. Overall results and results from specific concept clusters from the two concept tests are presented for each of the three research-based teaching methods in turn. Particular attention will be paid to questions on students’ understanding of velocity-time graphs (University of Maryland only) and Newton’s third law; two concepts that PER has shown to be difficult for many students to learn in traditional instruction.  

Section III looks at how well students use concepts and representations in solving long exam problems (as opposed to short answer or multiple choice problems). Because of the logistical difficulties discussed in the chapter 8, exam data was only available from University of Maryland. The results from four exam problems are presented.

Results from problem interviews with students are presented in section IV. The two-rock problem protocol discussed in chapter 7 was used in interviews with student volunteers from Maryland, Dickinson, and Ohio State. The two-rock problem is very difficult for students. However as we discussed in chapter 7, it also provides an opportunity to see how well the student understands the concepts of velocity vectors, kinematics, and energy as well as probing their approach to physics.
II. STUDENTS’ UNDERSTANDING OF BASIC CONCEPTS

A. Overall concept of Force and Motion

1. University of Maryland Tutorials

The results of pre- and post-course FCI’s from tutorial and non-tutorial classes at the University of Maryland are given in Table 9-1. Sixteen first semester classes in the introductory physics sequence for engineering majors gave the FCI as pre- and post-tests. Nine were taught with tutorials, seven with recitations. Two of the instructors, C and D, taught classes with both formats. A comparison of the averages of the pre-test scores for all students taking the pre-test with the matched subset show that the matched data is consistent with the unmatched data. Therefore, the matched samples are a reasonable representation of the classes in question.

The results are displayed as a figure of merit ($h$) histogram in Fig. 9-1. Recall that $h$ is the fraction of the possible gain achieved from the pre to post FCI results ($h$ is described in detail in chapter 4 in the section on Hake’s 6000 student study). The tutorial classes systematically produced better overall FCI gains than the non-tutorial classes. The average fractional gains of the classes are ($< h > \pm \text{Std. Error}$)

\[
< h > = 0.19 \pm 0.03 \text{ (7 classes, with recitations)}
\]

\[
< h > = 0.35 \pm 0.01 \text{ (9 classes, with tutorials)}
\]

Note that this average is taken as equally weighted over lecture classes, not by students. Every tutorial class had a larger $h$ than all but one of the non-tutorial classes and that was a small class taught by an award-winning lecturer.\textsuperscript{11} However, even the tutorial results are somewhat disappointing, achieving only about 1/3 of the possible
Table 9-1: Overall FCI results for University of Maryland Traditional and Tutorial classes. Instructors are coded by letter. Numbers indicate that instructor has taught more than one class and are sequenced chronologically. $h$ is defined as the fraction of the possible gain.

<table>
<thead>
<tr>
<th>Traditional Lecture w/ Recitations</th>
<th>N</th>
<th>FCI Pre</th>
<th>FCI Post</th>
<th>$h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>42</td>
<td>55.4</td>
<td>55.9</td>
<td>0.01</td>
</tr>
<tr>
<td>B1</td>
<td>27</td>
<td>51.2</td>
<td>65.5</td>
<td>0.29</td>
</tr>
<tr>
<td>B2</td>
<td>18</td>
<td>60.2</td>
<td>70.1</td>
<td>0.25</td>
</tr>
<tr>
<td>C1</td>
<td>35</td>
<td>41.8</td>
<td>54.2</td>
<td>0.21</td>
</tr>
<tr>
<td>C2</td>
<td>39</td>
<td>44.5</td>
<td>52.9</td>
<td>0.15</td>
</tr>
<tr>
<td>D2</td>
<td>21</td>
<td>53.0</td>
<td>52.6</td>
<td>0.20</td>
</tr>
<tr>
<td>E</td>
<td>76</td>
<td>51.4</td>
<td>61.8</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>36.9</td>
<td>51.1</td>
<td>60.4</td>
<td>0.19</td>
</tr>
<tr>
<td><strong>Std. Dev.</strong></td>
<td>19.5</td>
<td>6.3</td>
<td>6.3</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>Std. Error</strong></td>
<td>7.4</td>
<td>2.4</td>
<td>2.4</td>
<td>0.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Traditional Lecture w/ Tutorials</th>
<th>N</th>
<th>FCI Pre</th>
<th>FCI Post</th>
<th>$h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>46</td>
<td>51.5</td>
<td>69.6</td>
<td>0.37</td>
</tr>
<tr>
<td>D1</td>
<td>55</td>
<td>53.9</td>
<td>67.8</td>
<td>0.30</td>
</tr>
<tr>
<td>F1</td>
<td>38</td>
<td>47.8</td>
<td>66.7</td>
<td>0.36</td>
</tr>
<tr>
<td>F2</td>
<td>102</td>
<td>54.5</td>
<td>72.8</td>
<td>0.40</td>
</tr>
<tr>
<td>G1</td>
<td>69</td>
<td>50.3</td>
<td>67.5</td>
<td>0.35</td>
</tr>
<tr>
<td>G2</td>
<td>65</td>
<td>45.3</td>
<td>61.1</td>
<td>0.29</td>
</tr>
<tr>
<td>H</td>
<td>59</td>
<td>47.6</td>
<td>67.2</td>
<td>0.37</td>
</tr>
<tr>
<td>I</td>
<td>24</td>
<td>52.7</td>
<td>69.4</td>
<td>0.35</td>
</tr>
<tr>
<td>J</td>
<td>88</td>
<td>50.3</td>
<td>68.9</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>60.7</td>
<td>50.4</td>
<td>67.9</td>
<td>0.35</td>
</tr>
<tr>
<td><strong>Std. Dev.</strong></td>
<td>24.1</td>
<td>3.1</td>
<td>3.1</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Std. Error</strong></td>
<td>8.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Figure 9-1. Overall FCI figure of merit histogram for classes at the University of Maryland. Figure of merit $h =$ fraction of possible gain on the full FCI, for tutorial classes (TUT) and traditional (TRD) lecture classes.
gain. Both results, however, are consistent with Hake’s findings from his study of 6000 introductory physics’ students discussed in chapter 4. He found that 14 classes taught in the traditional lecture style and 48 classes taught with active engagement (research-based) curricula had the following average fractional gains:\(^\text{12}\)

\[
\begin{align*}
\text{Traditional Classes} & \quad \langle h \rangle = 0.23 \pm 0.04 \text{ (std. dev.)} \\
\text{Active Engagement Classes} & \quad \langle h \rangle = 0.48 \pm 0.14 \text{ (std. dev.)}
\end{align*}
\]

where \( h \) is averaged over classes, not students. Both the non-tutorial and the tutorial results from the Maryland classes are within one standard deviation of the respective average \( h \)-values measured by Hake. Note that the active-engagement activity in the Maryland Tutorial classes was limited to only one hour of four per week while many of Hake’s active engagement classes had a much higher fraction of and total time spent on active engagement activities.

Assuming that all 16 University of Maryland classes are drawn from the same population, the probability that the difference of the means is random is less than 2\% using a 2-tailed t-test with pooled variance.\(^\text{13}\) If class A is excluded as an outlier, the probability that the difference in the means is random is less than 1\%.

The same amount of instruction was offered students in both environments (3 hours of lecture and 1 hour of small class section). The primary difference between the tutorial and traditional classes is that the tutorial classes spend one hour per week on explicit concept building in a small-class group-learning-style environment, while the traditional classes spend one problem-solving hour per week in a small-class lecture-style environment.
2. FCI results at other schools

The FCI was also used as a pre/post evaluation for the first term in the introductory sequence at 7 of the 9 other schools participating in this study. The overall FCI results are shown in Table 9-2. Overall FMCE results are given for the other two schools as well as more recent WP classes at Dickinson College are shown in Table 9-3. The values for each school in the table except Ohio State are averaged over the number of classes. The Ohio State data is averaged over students. The reader is reminded that the Workshop Physics classes at Drury, Nebraska Wesleyan, & Skidmore and the Group Problem Solving classes from Minnesota & Ohio State were in the first two years of the implementation of their respective curricula.\(^{14}\)

The pre-course FCI scores for the eight schools appear to represent two distributions. The pre-course averages at Drury and the three large state universities cluster around 50%. This is significantly larger (std. error \(\approx 0.03\)) than the 40% pre-course average at the other schools. A similar difference can be seen in the pre-course FMCE data between Moorhead State and the other two schools. The pre-course average for classes at Moorhead and at Maryland is about 40% while the classes at Carroll and Dickinson start with an average around 26%. These differences do not correlate with the selectivity of the college.

The average fractional gain \((h)\) for all of the research-based classes is significantly greater than the average for the traditional lecture classes at Maryland and the community college. The average \(h\) for all classes using research-based instruction each fall within one standard deviation of Hake’s average \(h\) value for active-engagement
Table 9-2: Overall FCI Scores for all curricula (scores ± std. error)

<table>
<thead>
<tr>
<th>University of Maryland (F93-S97)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (# of classes)</td>
<td>FCI Pre</td>
<td>FCI Post</td>
<td>$h$</td>
</tr>
<tr>
<td>Recitations</td>
<td>258 (7)</td>
<td>51.1 ± 2.4</td>
<td>60.4 ± 2.4</td>
<td>0.19 ± 0.03</td>
</tr>
<tr>
<td>Tutorials</td>
<td>546 (9)</td>
<td>50.4 ± 1.0</td>
<td>67.9 ± 1.0</td>
<td>0.35 ± 0.01</td>
</tr>
<tr>
<td></td>
<td>Other Traditional Lecture courses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N (# of classes)</td>
<td>FCI Pre</td>
<td>FCI Post</td>
<td>$h$</td>
</tr>
<tr>
<td>PGCC (F94&amp;F95)</td>
<td>40 (4)</td>
<td>37.2 ± 2.2</td>
<td>43.0 ± 4.4</td>
<td>0.09 ± 0.04</td>
</tr>
<tr>
<td></td>
<td>Group Problem Solving</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N (# of classes)</td>
<td>FCI Pre</td>
<td>FCI Post</td>
<td>$h$</td>
</tr>
<tr>
<td>MIN (F94)</td>
<td>524 (5)</td>
<td>49.0 ± 1.0</td>
<td>65.2 ± 0.7</td>
<td>0.32 ± 0.004</td>
</tr>
<tr>
<td>MIN (F95)</td>
<td>653 (5)</td>
<td>49.7 ± 1.1</td>
<td>72.9 ± 2.0</td>
<td>0.46 ± 0.03</td>
</tr>
<tr>
<td>OSU (F95)</td>
<td>258 (2)</td>
<td>50.4</td>
<td>69.4</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Workshop Physics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N (# of classes)</td>
<td>FCI Pre</td>
<td>FCI Post</td>
<td>$h$</td>
</tr>
<tr>
<td>DC (F92)</td>
<td>62 (3)</td>
<td>42.2</td>
<td>66.7</td>
<td>0.42</td>
</tr>
<tr>
<td>DRY (F96)</td>
<td>8 (1)</td>
<td>47.8</td>
<td>77.6</td>
<td>0.57</td>
</tr>
<tr>
<td>NWU (F95&amp;F96)</td>
<td>68 (6)</td>
<td>38.0 ± 1.7</td>
<td>62.3 ± 1.9</td>
<td>0.39 ± 0.03</td>
</tr>
<tr>
<td>SKD (F96)</td>
<td>33 (2)</td>
<td>41.2</td>
<td>63.9</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Average $h$ for classes using research-based curricula: $h = 0.39$ (N = 2150)

Average $h$ for classes using traditional lecture instruction: $h = 0.15$ (N = 298)

Table 9-3: Overall FMCE results for Workshop Physics schools

<table>
<thead>
<tr>
<th></th>
<th>N (# of classes)</th>
<th>FMCE Pre</th>
<th>FMCE Post</th>
<th>$h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAR (F95)</td>
<td>24 (1)</td>
<td>25.9</td>
<td>39.4</td>
<td>0.18</td>
</tr>
<tr>
<td>MSU (F95)</td>
<td>16 (1)</td>
<td>37.5</td>
<td>72.2</td>
<td>0.56</td>
</tr>
<tr>
<td>DC (F94-F96)</td>
<td>154 (7)</td>
<td>24.1 ± 1.2</td>
<td>73.4 ± 2.2</td>
<td>0.65 ± 0.03</td>
</tr>
</tbody>
</table>
Figure 9-2. FCI (above) and FMCE (below) figure of merit histogram for classes from all ten schools participating in the study. The figure of merit $h = \text{the fraction of the possible gain on the full FCI, for traditional lecture classes (TRD), Tutorial classes (TUT), Group Problem Solving classes (GPS) and Workshop Physics classes (WP).}$
classes. The average $h$ for all the traditional lecture classes at Maryland and the community college is smaller, but not significantly different than Hake’s value.\textsuperscript{15}

The average $h$-value for each class in this study using research-based curricula is better than all but one of the traditional lecture classes at Maryland and all but one of Hake’s traditional undergraduate classes.\textsuperscript{16} The worst $h$-values for the research-based curriculum classes are as good as these two best traditional classes. The best average $h$-values in this study were achieved by the Workshop Physics classes at Dickinson and Drury and by the fall 96 Group Problem Solving classes at University of Minnesota. Not surprisingly, these programs were considered the most successful implementations of their respective curricula by the instructors and outside evaluators. Also, Dickinson and Minnesota are the development sites for their respective curricula. The class at Drury College was in the second year of their implementation and had the second highest $h$-factor for any class participating in this study. This class is unusually small even for Drury and should not necessarily be considered typical.

The highest $h$-factors at Dickinson, NWU, Minnesota, and Maryland were achieved by classes that were taught by instructors directly involved in the development and/or implementation of the research-based curricula. I suggest that this may be due to better integration of the active-engagement elements and major themes in these courses. A senior graduate student in the physics education group at Ohio State made a similar observation in their implementation of the Group Problem Solving curriculum.\textsuperscript{17} The highest FCI $h$-factor achieved by any class participating in this study was 0.59 for the GPS class taught by Ken Heller at Minnesota, who was involved with the development of the GPS curriculum.
The increase in the average $h$-factors at Minnesota between the fall 94 and fall 95 quarters is very large even discarding Heller’s class as an outlier. When asked about this, the physics education researchers responsible for implementing the GPS curriculum at Minnesota commented on two differences between the two quarters. The implementation of the problem solving labs was better in 1995 and the post-course FCI was given as part of the final exam in 1995 instead of in the laboratory in the last week of classes. The PER people at Minnesota suggest that perhaps the increase in scores is due to the students taking “the FCI much more seriously as part of the final exam.”

There are three possible explanations for this remarkable improvement:

1. The students actually understood the concepts much better because of improvements in the curriculum or because they finally understood the concepts better after studying for the final exam.
2. The students don’t take the FCI as seriously when they are not graded on their score.
3. In the context of the exam, the students are responding with what they were taught rather than what they believe.

There is not enough data here to distinguish between these three possibilities. That is left for future research. However, it is interesting to note that in Hake’s study the highest $h$-factors were observed in classes where the post FCI was incorporated into the final exam.

Last, since Carroll College, Moorhead State University and the more recent Workshop Physics classes at Dickinson use the FMCE instead of the FCI for their evaluation of students conceptual understanding, FMCE results from both schools are included in Table 9-3 for comparison with the FCI results. Laws feels these FMCE results for Dickinson are more representative of their Workshop Physics classes due to improvements in the curriculum and some unusual difficulties in the 1992 fall semester.
By the end of the first term, the Workshop Physics classes at Moorhead State and Dickinson show significantly greater improvement in understanding of basic concepts compared to the traditional class at Carroll College both in terms of actual score and fractional gain $h$.

B. Velocity Graphs

Velocity graphs are essential to an understanding of mechanics and address the general issue of the relationship between a quantity and its rate of change. Velocity graphs are also known to be difficult for many introductory physics students (see the brief discussion on representations in chapter 2).\textsuperscript{19} Thornton and Sokoloff found that student understanding of velocity graphs could be significantly improved using an MBL curriculum they developed.\textsuperscript{20} They evaluated the effect of their curriculum using a set of multiple-choice velocity graph questions (VQ) in which students were required to match a description of a motion to a velocity graph. The VQ questions are shown in Figure 9-3. Thornton and Sokoloff demonstrated that students who were given four hours of their group-learning guided-discovery active-engagement MBL curriculum were significantly more successful in choosing the correct graphs than those who only received traditional lecture instruction.

The results are dramatic, with a large fraction of the students missing all but the simplest of the five velocity graph questions after traditional instruction.\textsuperscript{21} After the MBL activities, the error rate drops to below 10\% on all questions. This result is very robust and has been confirmed at dozens of colleges and universities. The results of
An object's motion is restricted to one dimension along the + distance axis. Answer each of the questions below by selecting the velocity graph below that is the best choice to describe the answer. You may use a graph more than once or not at all.

a. Which velocity graph shows an object going away from the origin at a steady velocity?

b. Which velocity graph shows an object that is standing still?

c. Which velocity graph shows an object moving toward the origin at a steady velocity?

d. Which velocity graph shows an object changing direction?

e. Which velocity graph shows an object that is steadily increasing its speed?
Thornton and Sokoloff are cited as an indication that interactive-engagement MBL activities are highly effective. However, some members of the physics education community question whether the improvement is due to the MBL activity or to the extra time on the topic? The following study addresses this question.

The VQ were given in two of the Maryland classes taught by Redish, my advisor. In the first class (G0), Professor Redish did his best to teach the material explicitly in lecture, devoting nearly three full lecture hours to the topic of instantaneous velocity. Lecture demonstrations using the same MBL apparatus as in the tutorial were conducted with much student interaction and discussion. The professor had the students watch and plot the professor’s motion as he walked a variety of paths, and a number of problems relating to students’ personal experience were presented, but no worksheets were distributed. While the students were prompted for predictions for many of the situations, they were not given time to explain their ideas and predictions with one another. In the recitation sections, graduate-teaching assistants spent one hour going over textbook problems on the same material.

For class G2, the tutorial system was in place, and the one-hour MBL velocity tutorial written by Redish and the author (this tutorial is described in more detail in chapter 8) was given. The professor reduced the lecture time on the topic to a single hour, which was more typical of a traditional lecture and had little student interaction. In both classes, the questions were given as part of an examination and were not previously given to the students as homework. The results for the error rates are given in Table 9-4 and shown in Fig. 9-4. Note that since only post instruction data was taken, this data is not matched.
Table 9-4: Percentage error on the VQ with and without MBL. TRD indicates traditional lecture instruction tutorial; TUT indicates class was taught with tutorials.

<table>
<thead>
<tr>
<th>Instruction without MBL</th>
<th>VQ1</th>
<th>VQ2</th>
<th>VQ3</th>
<th>VQ4</th>
<th>VQ5</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Maryland (TRD) N = 100</td>
<td>59</td>
<td>7</td>
<td>62</td>
<td>37</td>
<td>14</td>
</tr>
<tr>
<td>Tufts (Thornton &amp; Sokoloff)$^{23}$ N = 177</td>
<td>18</td>
<td>7</td>
<td>41</td>
<td>18.5</td>
<td>17</td>
</tr>
<tr>
<td>Six School Average (Thornton &amp; Sokoloff)$^{24}$ N = 505</td>
<td>41</td>
<td>17</td>
<td>63</td>
<td>37</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Instruction with MBL</th>
<th>VQ1</th>
<th>VQ2</th>
<th>VQ3</th>
<th>VQ4</th>
<th>VQ5</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Maryland (TRD) N = 100</td>
<td>16</td>
<td>2</td>
<td>30</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td>Tufts (Thornton &amp; Sokoloff)$^{25}$ N = 177</td>
<td>2</td>
<td>…</td>
<td>7</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Six School Average (Thornton &amp; Sokoloff)$^{26}$ N = 505</td>
<td>11.5</td>
<td>2</td>
<td>13</td>
<td>11</td>
<td>7</td>
</tr>
</tbody>
</table>
Figure 9-4. Error rate on velocity questions (VQ). TRD indicates classes taught without tutorials or MBL, TUT indicates Maryland class was taught with tutorials, MBL indicates classes taught with 4 hours of MBL activities.
The Maryland results with four hours of traditional instruction on velocity and velocity graphs but no tutorial (class F0) resembled the 6-school average of traditional lecture classes reported in Thornton’s lecture at the Raleigh conference.27 Not surprisingly, the Maryland result from class F2 with one hour of MBL tutorial and one hour of lecture was substantially improved from four hours of traditional instruction, but not as good as the improvement shown with four hours of Thornton and Sokoloff’s MBL activities.

These results are consistent with those given by Thornton and Sokoloff. The fact that these results have been obtained with both the lecturer and the time of instruction controlled strongly supports the finding by Thornton and Sokoloff that cooperative group activities with MBL can be more effective for helping students learn to understand graphical representations than traditional instruction. In this case, the improvement was achieved with two hours of instruction compared to four hours of traditional instruction so the improvement was not due to additional time on task. These results demonstrate that MBL group activities can play a significant role in improving student understanding of the concept of velocity. It is not simply the extra time that is responsible. It also suggests that simply enhancing lectures is not necessarily effective in producing an improvement in the learning of the velocity concept for a significant number of students.
C. Newton’s Third Law

1. FCI Newton 3 Cluster at University of Maryland

Student understanding of Newton’s third law was evaluated using the four FCI questions 2, 11, 13, and 14 (N3 FCI) shown in Figure 9-5. The results from the 16 traditional and tutorial classes at Maryland discussed previously are given in Table 9-5 and shown as a histogram in Figure 9-6. The table gives the fraction of students answering each of the N3 FCI questions correctly at the beginning (pre) and end (post) of the semester. A figure of merit, \( h = \frac{\text{class post-test average} - \text{class pre-test average}}{100 - \text{class pre-test average}} \), is calculated for each question in analogy with the Hake figure of merit for the full FCI. The four \( h \)-values are then averaged in the last column to give a figure of merit for the Newton 3 FCI cluster, \( h_{N3} \).

The fractional gain results are systematically better for the tutorial classes. Indeed, every tutorial class has a higher value of \( h_{N3} \) than every non-tutorial class (though a similar statement is not true for the \( h \)-values for every individual question). The average values of \( h_{N3} \) for each group of classes are (\( < h > \pm \text{Std. Error} \)):

\[
\begin{align*}
< h_{N3} > &= 0.28 \pm 0.04 \quad (7 \text{ classes, with recitations}) \\
< h_{N3} > &= 0.60 \pm 0.03 \quad (9 \text{ classes, with tutorials}) \\
< h_{N3} > &= 0.41 \quad (1 \text{ class with no MBL tutorials}) \\
< h_{N3} > &= 0.64 \pm 0.05 \quad (4 \text{ classes, with Newton 3 MBL tutorial}) \\
< h_{N3} > &= 0.60 \pm 0.02 \quad (4 \text{ classes, with MBL tutorials but no Newton 3 MBL tutorial})
\end{align*}
\]

where the last four classes had other MBL tutorials including the velocity tutorial and the Newton’s 2\textsuperscript{nd} law tutorial discussed in chapter 8. Using pooled variance, the
2. Imagine a head-on collision between a large truck and a small compact car. During the collision:

(A) the truck exerts a greater amount of force on the car than the car exerts on the truck.

(B) the car exerts a greater amount of force on the truck than the truck exerts on the car.

(C) neither exerts a force on the other, the car gets smashed simply because it gets in the way of the truck.

(D) the truck exerts a force on the car but the car does not exert a force on the truck.

(E) the truck exerts the same amount of force on the car as the car exerts on the truck.

11. In the figure at right, student "a" has a mass of 95 kg and student "b" has a mass of 77 kg. They sit in identical office chairs facing each other.

Student "a" places his bare feet on the knees of student "b", as shown. Student "a" then suddenly pushes outward with his feet, causing both chairs to move.

In this situation:

(A) neither student exerts a force on the other.

(B) student "a" exerts a force on student "b", but "b" does not exert any force on "a".

(C) each student exerts a force on the other, but "b" exerts the larger force.

(D) each student exerts a force on the other, but "a" exerts the larger force.

(E) each student exerts the same amount of force on the other.
Figure 9-5. Newton’s third law FCI questions (N3 FCI) continued

Refer to the following statement and diagram while answering the next two questions.

A large truck breaks down out on the road and receives a push back into town by a small compact car.

13. While the car, still pushing the truck, is speeding up to get up to cruising speed:
   (A) The amount of force with which the car pushes on the truck is equal to that of the truck pushing back on the car.
   (B) The amount of force of the car pushing on the truck is smaller than that of the truck pushing back on the car.
   (C) The amount of force of the car pushing against the truck is greater than that of the truck pushing back on the car.
   (D) The car's engine is running so it applies a force as it pushes against the truck, but the truck's engine is not running so the truck cannot push back against the car. The truck is pushed forward simply because it is in the way of the car.
   (E) Neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.

14. After the car reaches the constant cruising speed at which its driver wishes to push the truck:
   (A) The amount of force of the car pushing on the truck is equal to that of the truck pushing back on the car.
   (B) The amount of force of the car pushing on the truck is smaller than that of the truck pushing back on the car.
   (C) The amount of force of the car pushing on the truck is greater than that of the truck pushing back on the car.
   (D) The car's engine is running so it applies a force as it pushes against the truck, but the truck's engine is not running so it can't push back against the car. The truck is pushed forward simply because it is in the way of the car.
   (E) Neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.
Table 9-5. N3 FCI results for Maryland classes.
Figure 9-6. Histogram of average figures of merit \( (h_{N3}) \) for the Newton 3 FCI cluster for University of Maryland traditional (blue bars) and tutorial classes (yellow bars). All classes which used MBL tutorials with or without the Newton 3 MBL tutorial are shown in solid yellow. The crosshatched bar represents the one tutorial class taught without any MBL tutorials.

Newton 3 FCI Questions for University of Maryland classes

- **UMD (TRD)**
- **UMD (TUT)**
- **UMD (TUT-No MBL)**
standard error of the last two $<h_{N3}>$’s is $\sigma = 0.05$. Thus, if the other MBL tutorials are used, there is no significant difference in the student scores on the Newton 3 FCI cluster when the Newton 3 MBL tutorial is not used.

In the first semester in which tutorials were tested, there were no MBL tutorials and there was no tutorial specifically oriented towards Newton 3. The first MBL tutorials were implemented the following year. As a result, the first Maryland tutorial class, G1, used tutorials but no MBL tutorials. The same instructor taught the class with all three MBL mechanics tutorials in a later semester. This gives us a control for individual lecturer as well as for the presence of tutorials. (No special effort was devoted to Newton 3 in lecture in either case.) The result was:

\[
< h_{N3} > = 0.41 \text{ (F1: with no MBL tutorials)}
\]
\[
< h_{N3} > = 0.65 \text{ (F2: with velocity, Newton2, and Newton 3 MBL tutorials)}
\]

However, in later semesters a non-MBL Newton 3 tutorial was substituted for the MBL version while keeping the velocity and the Newton 2 MBL tutorials. As can be seen in the results reported on page 9-19, no significant difference in either the overall FCI or the FCI Newton 3 cluster results was observed.

2. FCI Newton 3 cluster for other research-based and traditional lecture curricula

A similar analysis was performed on the FCI data for other schools where the itemized FCI data was available (some schools only submitted the overall FCI score for each student). FCI Newton 3 cluster results from Ohio State, Minnesota, Dickinson, Skidmore, and Nebraska Wesleyan are summarized in Table 9-6. The distribution of fractional gains for classes taught with the four curricula is shown in a histogram in Table 9-6. N3 FCI results for all four curricula
Figure 9-7. Histogram of average figures of merit ($h_{N/3}$) for the Newton 3 FCI cluster for all four curricula: Traditional lecture classes at University of Maryland (blue bars), Tutorial classes at University of Maryland (yellow bars), Group Problem solving at University of Minnesota (F94) and Ohio State (red bars), and Workshop Physics classes at Dickinson College, Skidmore College, and Nebraska Wesleyan University.

**Newton 3 FCI Questions for all four curricula**

![Graph showing the histogram of average figures of merit ($h_{N/3}$) for the Newton 3 FCI cluster for all four curricula: Traditional lecture classes at University of Maryland (blue bars), Tutorial classes at University of Maryland (yellow bars), Group Problem solving at University of Minnesota (F94) and Ohio State (red bars), and Workshop Physics classes at Dickinson College, Skidmore College, and Nebraska Wesleyan University.](image)
Figure 9-7. The Ohio State and the Skidmore data only represent the results of two classes each. Accordingly, the results from Ohio State are averaged over students instead of classes.

The average pre-course Newton 3 scores vary from 26% at Skidmore to 41% at Maryland with most of the scores clustering at 31%. Surprisingly, the pre-course averages for Minnesota and Ohio State are significantly smaller than Maryland’s.

One of the Maryland Tutorial classes and four of the Workshop Physics classes achieved a fraction gain $h > 0.70$ (two of the classes at Dickinson, one of the classes at Skidmore, and one of the classes at Nebraska Wesleyan). The result for the Skidmore class is especially notable since the pre-course score for the Newton 3 cluster was the smallest of any class participating in this study. In fact, the students at both Skidmore and Nebraska Wesleyan had significantly lower average pre-course Newton 3 FCI scores than the other four schools whose results are shown in Table 9-6.

The school average fractional gains for classes using one of the three research-based curricula were significantly higher than those using traditional lecture instruction at Maryland. The scores for all the research-based classes are significantly larger than all the traditional lecture classes except for classes B1 and one Workshop Physics class at Dickinson. Class B1 had the best overall and Newton 3 FCI results of the traditional lecture classes. As mentioned earlier, the Dickinson class had many difficulties including poor attendance.

The best fractional gains for the research-based classes were achieved by classes that incorporated at least some MBL group learning activities to address student difficulties with Newton’s third law. The Group Problem Solving classes from Ohio
as well as all but two of the Tutorial and Workshop Physics classes. This result is not solving, not conceptual understanding explicitly like Tutorials and Workshop Physics. Also, no itemized FCI dat implementation taught with the GPS curriculum in both the primary and secondary implementations show greater improvement in Maryland students taught with traditional instruction.

S AND APPLICATION OF CONCEPTS IN COMPLEX PROBLEMS

-choice questions tell us whether students “have” the desired information, it gives no information on whether they can access it in an appropriate complex problem. Accordingly, the main two questions in this section are the following:

Q1. on multiple choice mechanics questions imply that the students can use these

Q2. Mechanics is a large part of the first term course, but it is only one of many topics covered in the typical calculus based introductory physics sequence. Are students -based curricula better able to apply their improved conceptual

To answer these questions, an analysis of student responses for four pro specially designed for this study, are presented in this section. These problems look at students’ use and/or understanding of velocity graphs & Newton’s third law, position interference.
A. Mechanics

In order to address question Q1, I developed an examination problem that required students to display both an understanding of a velocity graph and to use Newton’s third law in a complex physical situation. The problem is shown in Figure 9-8 (this problem is a variation of the problem in Figure 6-2a.). The problem in Figure 9-8 was given on the final exam in one tutorial class and one non-tutorial class, classes G2 and C2 respectively (FCI data from these two classes is shown in Tables 9-1 and 9-5). Overall, performance on the problem was better for the tutorial than for the non-tutorial students. However, here we will only discuss issues related to the velocity graph and Newton’s third law.

1. Velocity

Part of the examination question asked the student to generate a velocity vs. time graph for a complicated situation. The critical elements of a fully correct solution show the velocity starting at 0, increasing linearly until $t = 3$ seconds, and then decreasing linearly to some negative value.29

Students from both classes struggled with this question. Table 9-7 shows a breakdown of student responses. Only a small fraction of the students in either class were able to draw a graph that reflected the critical features, but the tutorial students did better than the students in the recitations. After traditional instruction, 12% of the
Two carts, A and B (Mass$_A$ > Mass$_B$), are placed on a table then stuck together with Velcro. Using pulleys, two small blocks, C and D (mass$_C$ < mass$_D$), are connected by light strings to the carts as shown below. Initially, the carts are held in place. Ignore all friction in this problem.

At t = 0, the carts are released. At t = 3 seconds, the Velcro pulls apart and the two carts separate. At some later time, cart A returns to its starting point.

a. Draw and label two separate free-body diagrams, one for each cart, for a time after the carts start moving but before the Velcro pulls apart.

b. Rank all the horizontal forces from both your diagrams by magnitude, from largest to smallest. Explain the reasoning that you used to rank the forces.

c. Briefly describe the motion of cart A from t = 0 until it returns to its starting point. On the graph below, qualitatively sketch the velocity vs. time for this time period.
Table 9-7: Results on student constructions of the velocity graph in the long qualitative exam problem from two classes at the University of Maryland. The Tutorial class is in bold.

<table>
<thead>
<tr>
<th></th>
<th>% correct</th>
<th>% apparently correct, but ending at v = 0</th>
<th>% other incorrect response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional (N = 50)</td>
<td>12</td>
<td>10</td>
<td>78</td>
</tr>
<tr>
<td>Tutorial (N = 82)</td>
<td><strong>22</strong></td>
<td><strong>21</strong></td>
<td><strong>57</strong></td>
</tr>
</tbody>
</table>

Table 9-8: Results on student use of Newton’s third law in the long exam problem from two classes at the University of Maryland. The Tutorial class is in bold.

<table>
<thead>
<tr>
<th></th>
<th>% correct</th>
<th>% stated third law force pair have different magnitudes</th>
<th>% used the same symbol but did not compare forces</th>
<th>% with no identification of contact forces</th>
<th>% other incorrect response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional</td>
<td>42</td>
<td>22</td>
<td>6</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>Tutorial</td>
<td><strong>55</strong></td>
<td><strong>40</strong></td>
<td><strong>0</strong></td>
<td><strong>1</strong></td>
<td><strong>4</strong></td>
</tr>
</tbody>
</table>

Table 9-9: Results on Newton 3 FCI questions for University of Maryland classes C2 and G2. See Tables 9-1 and 9-5 for comparison with other Maryland classes. The Tutorial class is in bold.

<table>
<thead>
<tr>
<th></th>
<th>% correct Pre</th>
<th>% correct Post</th>
<th>$h_{N3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional</td>
<td>35.9</td>
<td>50.0</td>
<td>0.23</td>
</tr>
<tr>
<td>Tutorial</td>
<td><strong>39.6</strong></td>
<td><strong>69.2</strong></td>
<td><strong>0.51</strong></td>
</tr>
</tbody>
</table>
students drew a correct graph. After MBL tutorials, 22% of the students drew a correct graph.

Analysis of the incorrect graphs along with the accompanying explanations revealed some of the students’ difficulties. Many students showed in a variety of ways that they had the well-documented confusion between position and velocity (see the discussion on representations in chapter 2). Some drew graphs that at first glance appear correct: the graph increased linearly for the first 3 seconds and then decreased linearly after. However the graph ended at v=0, and some of these students indicated that this coincided with the cart returning to its starting location (an example of this type of solution is shown in Figure 6-2a). Many students drew graphs that had incorrect combinations of linear segments, including discontinuities in velocity. Others drew dramatically curved features in their velocity-time graphs. Most of these graphs indicated severe conceptual difficulties even if interpreted as a position vs. time graph. It is worth noting that it is clear from their explanations that most of these students intended to draw a velocity vs. time graph.

Both the percentage of correctly drawn graphs and the nature of the incorrect graphs confirm that while student difficulties understanding kinematics is pervasive even after instruction, the modified instruction described earlier in this paper appears to be helping address these difficulties somewhat. Although the VQ were not given in these classes, approximately 70% of the students in the comparable tutorial class F2 answered all of the multiple choice questions correctly, while only about 40% of those in the recitation class A1 answered them all correctly. The relative results on the long-problem are qualitatively consistent with the results of the VQ, but the absolute number of
students getting correct answers on the long-problem was substantially lower (22% of the tutorial students correct vs. 12% of recitation students correct). Since no classes were evaluated with both the VQ and the long problem, we cannot completely answer Question Q1, but our indications are that the VQ does not suffice. Our results suggest that answering multiple-choice questions correctly is not sufficient to guarantee a robust and fully functional understanding of the relevant concepts for a significant number of students.

2. Newton 3

Another part of the same examination question tested student facility with dynamical concepts, specifically Newton’s 2nd and 3rd laws of motion. The students were asked to draw a free body diagram of each cart shown in Figure 9-8 and to rank the magnitudes of the horizontal forces. Note in particular that by Newton’s third law, the magnitude of the force of cart A on cart B is equal to that of cart B on cart A.

The breakdown of student responses to this part of the question is shown in Table 9-8. In the tutorial classes, 55% of the students correctly identified and compared the third law force pair. In the non-tutorial class 42% identified and correctly compared these forces. Many students identified that the two carts were exerting forces on one another, but stated explicitly that the two forces were not of equal magnitude. In addition, there were also many students who did not even recognize that the two carts exert forces on each other. This was particularly common in the non-tutorial class.

These results should be compared with the results on the Newton’s third law FCI questions for the same two classes shown in Table 9-9. The two classes’ pre-course Newton 3 FCI scores are not significantly different ($\Delta < \sigma$), but the post-test results of
69% and 50% respectively are very similar to the ratio of correct responses to the exam problem for the two classes. The discrepancy between the multiple-choice and long-answer problems (in this case both questions were done by both groups) also suggests that the answer to question Q1 might be: the short answer results provide an indication but overestimates the students’ knowledge.

B. Beyond Mechanics

1. Harmonic Oscillator

As we have seen in chapter 2 and in the section above, even with enhanced instruction many students seem to have difficulty using velocity graphs in complex mechanics’ situations; however, students seem to have much less trouble with position graphs of mechanics situations. This has been demonstrated by Beichner (see chapter 2) as well as Thornton and Sokoloff. Our own pretests at the beginning of the sequence indicate that most students have little trouble with simple position graphs in mechanics. But what about position graphs for non-linear motion?

In a study of students’ understanding of oscillations and waves, we noticed that many students had trouble connecting sinusoidal graphs to physical quantities such as velocity, acceleration, and force. Redish and I designed and implemented an MBL tutorial for harmonic oscillators to help the students overcome this difficulty using both force and motion sensors. This tutorial went through several iterations. After the second iteration, the problem shown in Figure 9-10 was given on the final exam. Keep in mind the harmonic oscillator tutorial is the first tutorial at the beginning of the

Figure 9-9. Vertical harmonic oscillator problem for first tutorial class
The figure at the right shows two identical, massless, frictionless, springs, each with spring constant from a bar. Attached to one spring is a mass \( m_1 \) to the other spring is a mass \( m_2 > m_1 \). At \( t = 0 \), the two masses are connected to the springs and released.

(Note: When the masses are at their starting height, springs are at their unstretched lengths.)

**single** \( y = t \) \( y = t \)

is time), sketch the motion of each of the masses. Label to which mass.

B. i) Determine the \( y = t \) \( m \).

\( y = t \) \( m \).

\[
Y = D + A \cos(\omega t)
\]

\[
\omega > \frac{1}{2}
\]
to the two harmonic oscillators problem shown in Figure 9-oscillator tutorial was modified (N = 72 students). If the solution is not clear, the student is giv

The sketch of the motion of each of the masses on a single axes.

Starting points of the drawn curves:

- 
  (starts at maximum displacement)

  - Sine Curves
  - Other 5%

Distance from initial positions to equilibrium positions of the two masses:

- $d_1$ 2 28% Correct

  - Same equilibrium distance
  - Can’t tell

- $A_1 < A_2$
- $A_1 = A$
- $A_1 2$ 6%

Comparison of the periods of oscillation of the two mass:

- $T_1$ 2 49% Correct
- $T_1$ 2 28%
- $T_1$ 2 11%

  - Other (either inconsistent or indeterminate) 12%

Only 19% of the sketches drawn were basically corre sinusoidal starting from the same position at maximum displacement and the equilibrium position $d$. 

Equations for the two oscillating masses:

- The equations are correct 6%
- The equations are incorrect & inconsistent with the sketch 46%
- Equations are incomplete or no equations given 10%
semester. Even so, the results from my analysis of the student responses to this problem were very surprising. The results are summarized in Table 9-10.

A graph was considered basically correct if the two curves were sinusoidal, they started from the same point at maximum displacement, and the equilibrium positions of the two curves were different. Although 85% of the students clearly showed that the amplitude for \( m_2 \) was greater than that of \( m_1 \), only 20% of them drew a graph that was basically correct. The main difficulties are listed below:

- Almost two-thirds of the student drew sinusoidal curves that started at the equilibrium positions of the two masses;
- Fewer than half the graphs showed that the period of mass \( m_2 \) was greater than the period of mass \( m_1 \); and
- Only 30% of the students showed in their graphs that the two masses would oscillate around different points.

In addition, only 45% of the students wrote an equation that was consistent with their graph. Another 45% of the students wrote an equation that was both incorrect and inconsistent with the graph including 10% of the class who explicitly used terms from equations that describe traveling or standing sinusoidal waves.

As one can tell from the above results, very few students were able to present a consistent and correct solution. However, many students had several parts of the correct answer. This is consistent with the view expressed in earlier chapters that students’ conceptual knowledge is fragmented and not well organized. The fact that almost two-thirds of the graphs started the masses in the equilibrium position and that fewer than half the equations were consistent with the graphs suggest that, at least on this exam, many of the students saw only weak connections between the graph, the equation, and the physical situation, even for a position-time graph.
After I discussed the results with the PER group at Maryland, two members of 33) revised the harmonic oscillator tutorial to help sical quantities. The revised tutorial is included in Appendix B. The exam problem in Figure -9 was used as a homework problem for the new tutorial. I designed a new harmonic context. In this problem, the two masses are the same but one spring constant is four times greater than Figure 9- em was placed on the final exam after the new tutorial was taught at the beginning of the second semester. My -11.

This time a graph was considered basically correct if the period of cart one wa longer, the two curves had equal amplitudes, and both curves started from the same maximum displacement. Almost two thirds of the students drew the curves correctly. Roughly 80% of the students recognized that the curves should be sinusoidal, start at maximum displacement, and have equal amplitudes. Although only 36% of the students indicated that the period of cart 1 was twice that of cart 2, 78% of the students indicated own an equation that was roughly consistent with the graphs they had drawn. One third of the used terms from equations that describe traveling waves and 6% who used one or more
Two identical frictionless mass \( m \) (see figure at right). Initially, both springs unstretched and both carts are at rest at \( x = x_0 \). The massless springs in systems 1 and 2 have \( k_1 \) and \( k_2 \) where \( k = 4k_1 \).

Both c and released at \( d \) and released at \( t = 0 \) sec on the axes below. Clearly identify which curve corresponds to which cart. Label axes clearly.

b) Determine the equation which gives \( x \) as a function of \( t \) for cart 2. Explain how you arrived at your answer.

Correct Response: \( x = x_0 + d \cos \omega_2 t \) where \( \omega_2 = \sqrt{\frac{k_2}{m}} \). Note \( \omega_2 = 2\omega_1 \).
Table 9-11: Summary of student responses from a University of Maryland Tutorial class to the two harmonic oscillators problem in Figure 9-10 after the harmonic oscillator tutorial was modified (N=67 students). If the solution is not clear, the student is given the benefit of the doubt.

The sketch of the motion of each of the carts on a single axes.

Starting points of the drawn curves:

<table>
<thead>
<tr>
<th>Type of Curve</th>
<th>Percentage</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosine Curves (starts at maximum displacement)</td>
<td>79%</td>
<td>Yes</td>
</tr>
<tr>
<td>Sine Curves (starts at equilibrium position)</td>
<td>16%</td>
<td>Yes</td>
</tr>
<tr>
<td>Other</td>
<td>4%</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Comparison of the amplitudes of oscillation for the two masses:

<table>
<thead>
<tr>
<th>Amplitude Comparison</th>
<th>Percentage</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1 = A_2$</td>
<td>84%</td>
<td>Yes</td>
</tr>
<tr>
<td>$A_1 &gt; A_2$</td>
<td>4%</td>
<td>Yes</td>
</tr>
<tr>
<td>$A_1 &lt; A_2$</td>
<td>7%</td>
<td>Yes</td>
</tr>
<tr>
<td>Drew only 1 curve</td>
<td>4%</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Comparison of the periods of oscillation of the two mass:

<table>
<thead>
<tr>
<th>Period Comparison</th>
<th>Percentage</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1 = 2T_2$</td>
<td>36%</td>
<td>Yes</td>
</tr>
<tr>
<td>$T_1 &gt; T_2$ but the ratio is not clear</td>
<td>30%</td>
<td>Yes</td>
</tr>
<tr>
<td>$T_1 = 4T_2$</td>
<td>12%</td>
<td>Yes</td>
</tr>
<tr>
<td>$T_1 = T_2$</td>
<td>10%</td>
<td>Yes</td>
</tr>
<tr>
<td>$T_1 &lt; T_2$</td>
<td>10%</td>
<td>Yes</td>
</tr>
<tr>
<td>Can’t tell</td>
<td>1%</td>
<td>Yes</td>
</tr>
</tbody>
</table>

After modifications to the tutorial, 66% of the sketches drawn were basically correct, i.e. the period of cart one was longer, the two curves had equal amplitudes, and both curves started from the same maximum displacement.

Equations for the two oscillating masses:

<table>
<thead>
<tr>
<th>Equation Type</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>The equations are correct</td>
<td>65%</td>
</tr>
<tr>
<td>Roughly consistent</td>
<td>57%</td>
</tr>
<tr>
<td>No equation was given</td>
<td>10%</td>
</tr>
<tr>
<td>The equations have some elements from wave motion</td>
<td>13%</td>
</tr>
<tr>
<td>Student used a kinematic equations in their solution</td>
<td>6%</td>
</tr>
</tbody>
</table>
The results of the two exam problems strongly suggest that the revised MBL harmonic oscillator tutorial helped to improve student understanding of the connection between the position graphs, the equations, and the physical situation. Even though the first harmonic oscillator problem is marginally harder because of the vertical oscillations around different equilibrium, less than a third (29%) of the students from the first tutorial class drew graphs that were basically correct even when the equilibrium positions are removed from consideration compared with two thirds of the students in the second tutorial class. An evaluation of the student understanding of the graphs of other physical quantities pertaining to harmonic oscillators such as velocity and/or force is left for future studies.

2. Two-Slit Interference

One area covered by the tutorials in the third semester of the introductory physics sequence at University of Maryland is physical optics. The problem shown in Figure 9-11 was part A of an exam problem written by Richard Steinberg of the Maryland PER group to see how students taught with physical optics tutorials would do on a semi-conventional interference problem. This problem was given on exams in one tutorial class and two traditional lecture classes. The student responses to part A were analyzed by Sabella and Steinberg.\textsuperscript{34} The results of their analysis are summarized in Figure 9-12. The tutorial class did significantly better on this problem, 60% vs. 16%. It is interesting to note that 40% of the students taught with traditional instruction applied the conditions for the first maximum next to the central maximum (instead of the first compared to only 9% of the tutorial students. This indicates that a much larger
Light with $\lambda = 500 \text{ nm}$ is incident on two narrow slits separated by $d = 30 \mu m$. An interference pattern is observed on a screen a distance $L$ away from the slits. The first dark fringe is found to be 1.5 cm from the central maximum. Calculate the distance, $L$, to the screen. Show all work.

Model Solution:

$\Delta D = d \sin \theta$

$\Delta D = (m + \frac{1}{2})\lambda = \frac{1}{2}\lambda$

Assume $L \gg y$, then

$\sin \theta$ is almost equal to $\tan \theta$

since $\theta$ is small so $\tan \theta = \frac{y}{L}$

therefore $L = \frac{2dy}{\lambda} = 1.8 \text{ m}$

Figure 9-12: Graph of Student Responses to Double Slit Interference Problem.

Note: DD = $\Delta D$
fraction of the traditional class was using an equation without considering the conditions of the physical situation. Other incorrect responses included algebraic mistakes and using higher order minima.

IV. PROBLEM SOLVING INTERVIEW

The two-rock problem protocol discussed in chapter 7 was given as a think-aloud interview exercise after expectation interviews with some of the students during site visits at Maryland, Dickinson, and Ohio State. This problem was used in interviews by Hammer in his dissertation study of student expectations (see chapter 2). Since all our students are volunteers, most of the students interviewed were in the top half of their introductory physics class. Even so, only one or two students at each of the three schools were able to work through the problem with only a few hints. The students who were able to work through the problem demonstrated a good conceptual understanding of kinematics but only one of these students thought about using conservation of mechanical energy to compare the speed of the rocks when they hit the ground.

The student volunteers were first asked what they thought the answers would be and why. Almost all of the students were able to answer just one of the two questions correctly. Many of them recognized that the rock thrown downward would hit the ground first but then assumed that meant the rock thrown downward would hit at the higher speed. The other students thought the two rocks would have the same speed when they hit the ground. This implied to them that the two rocks would hit the ground at the same time. Because acceleration was the same for both rocks, the students connected equal time to equal speed. Very few of the students seemed to recognize the
vector nature of velocity or look at this problem in terms of energy. The latter is an indication that word cues in the problem trigger the students to think of this problem as a kinematics problem.

Many of the students were reluctant to attempt a mathematical solution. This was particularly true of the Dickinson Workshop Physics students. After discussing what they qualitatively predicted the answer would be, they wanted to test their prediction with an experiment. This is understandable since prediction followed by experiment is a major component of the Workshop Physics curriculum. But it is disappointing since students are also expected to be able to apply analytical tools as well.

Most of the student volunteers were able to recall the kinematic equations; however, very few of them were able to derive them or state that the kinematic equations require constant acceleration when asked if the kinematic equations are always valid. Many of the student volunteers were surprised when we went over the derivation of the kinematic equations using integration and claimed they had never seen that derivation before. And as observed previously in Hammer’s dissertation study,35 few of the students had a good conceptual understanding of the equation \( v(t) = v_0 + at \).

One thing that makes this problem difficult is that while it is possible to derive a symbolic expression for the time it takes each rock to fall, the symbolic expressions are difficult to compare. Most of the students were able to derive the two equations below:

\[
    t_1 = \frac{-v_0 + \sqrt{v_0^2 + 2ah}}{a}
\]

\[
    t_2 = \frac{2h}{a}
\]
where \( t_2 \) is the time it takes the rock thrown horizontally with speed \( v_0 \) to fall a distance \( h \) while \( t_1 \) is time it would take the rock thrown straight down with the same speed to cover the same distance. The students were encouraged to find a symbolic solution to the problem. Most of the students plugged numbers into the two equations above. Only a few students were able to come up with additional ways of approaching the problem on their own including using average velocity or conservation of mechanical energy. This implies that if the students’ knowledge is flexible enough they are able to come up with another approach to solve the problem.

Now, recognizing that these students were put on the spot in what was essentially a videotaped oral exam, they did not do that badly. However, the inability of most of the students to derive the kinematic equations using integration or conceptually understand the velocity equation is more disturbing, particularly since all the students had at least one term of physics instruction that emphasized learning with understanding.

IV. SUMMARY

In this chapter, results have been presented from two types of measurements of students’ conceptual understanding of physics in the calculus-based introductory course:

1. How well students know basic physics concepts as measured by multiple choice tests like the FCI and FMCE when taught with Tutorials, Group Problem Solving, Workshop Physics, or traditional lecture instruction at various institutions.

2. How well students use physics concepts in solving complex problems when taught with tutorials or traditional lecture instruction.

Pre- and Post-course concept-test data were collected from classes at all ten schools participating in this study. The overall FCI and FMCE results from the 10 schools clearly show that classes taught with one of the research-based curricula improve
significantly more than traditional lecture classes. This is even true for classes where the research-based curriculum is in the first year of adoption at the institution. Instructors at University of Maryland who taught with both tutorials and traditional recitations had better fractional gains from the classes with Tutorials.

Similar results were observed for the Newton’s third law cluster of FCI questions. All the Tutorial classes at Maryland taught with MBL tutorials had significantly larger fractional gains than the traditional lecture classes at Maryland, even though both classes had similar pre-course scores on these questions. Although some the Workshop Physics classes started with much lower scores initially, the fractional gains of the classes taught with Workshop Physics and Group Problem Solving were also significantly better than the fractional gains for the traditional classes at Maryland. The Workshop Physics classes and the Tutorial classes, which used at least some MBL activities, appeared to improve more than the classes that did not. In addition, a Tutorial class which used the MBL velocity tutorial did better on velocity graph questions than a traditional lecture class taught by Redish who spent twice as much time (4 hours vs. 2 hours) going over velocity in the traditional class.

Four specially designed exam problems were used to see how well University of Maryland students were able to apply concepts when solving complex problems. One of these problems, which looked at students’ understanding of velocity graphs and Newton’s third law, was placed on the final exams of a Traditional class and a Tutorial class. The Tutorial students showed a significantly better understanding of the two concepts. The difference on Newton’s third law between the types of classes was similar to the differences in the class scores on the Newton three FCI cluster.
The other three problems were used to look at student understanding of concepts beyond mechanics. Two similar harmonic oscillator problems were given on final exams before and after the harmonic oscillator tutorial was revised. Each problem required students to sketch a graph the motion of two harmonic oscillators on the same axes. The first problem was marginally more difficult because the two oscillators were oscillating vertically around two different equilibrium positions. If this difference is removed from consideration, the fraction of students who were able to draw a qualitatively correct graph of the two oscillators increased from 29% to 66% after the tutorial was revised. The fraction of students who wrote an equation to describe the motion consistent with their graph increased from 45% to 57%. These results indicate that the classes taught with Tutorials not only had higher scores on concept test but also did better on exam problems using similar concepts. Moreover, the students in the tutorial classes showed a better understanding of concepts on exam problems beyond mechanics. In addition, Sabella et al. showed that Maryland Tutorial students performed better on a two-slit interference exam problem than Maryland students taught with traditional lecture.\textsuperscript{36}

However, only a few students from classes taught with research-based curricula were able to solve a difficult mechanics problem in interviews. Because of the nature of the problem, this was not surprising. What was surprising was that many of the students were unable to derive the kinematic equations and became stuck when the kinematic equations produced an expression that was hard to evaluate. This suggests that the research-based curricula may need further improvement to help students develop a more flexible and functional understanding of physics concepts and connect the concepts better to the equations.


7 See Ref. 3.

8 See Ref. 5.


11 In terms of absolute final scores, one tutorial class with a low pre-test score finished below some of the non-tutorial classes, and one non-tutorial class with a high pre-test score finished above some of the tutorial classes.

12 See Ref. 4.


14 The fall 1995 class at University of Minnesota was the third year in the implementation of the Group Problem Solving curriculum in the calculus-based introductory physics sequence.
See Ref. 4.

See Ref. 4.

Private communication with Chris Cooksey (August 1997).

Private communication from Laura McCullough and Tom Foster (November 1997).

See Ref. 9.


This class was taught prior to our use of the FCI at University of Maryland and is not shown in table 9-1.

See Ref. 21.

See Ref. 21.

See Ref. 21.

See Ref. 21.


Private communication with Tom Foster and Laura McCullough, Fall 1997.

A graph reversed with respect to the horizontal axis would also be considered correct.

A few students in the non-tutorial class used the same symbol for these two forces, but did not state whether the forces were equal, so it was impossible to determine if they were identifying these two forces as having equal magnitudes. Note that many students used the same symbol to represent forces that clearly had different magnitudes.


R. Thornton, “Using large-scale classroom research to study student conceptual learning in mechanics and to develop new approaches to learning,” in *Microcomputer-

33 Richard N. Steinberg is a Post Doc with the University of Maryland Physics Education Research Group; Michael C. Wittmann is a senior graduate student in the group.


35 See Ref. 10.

36 See Ref. 31.
I. OVERVIEW

A. What are expectations?

Students bring to introductory physics classes more than just common sense beliefs about how things work. They also bring assumptions, beliefs and attitudes about the nature of what they are learning, the skills they need to succeed, and what they need to do to achieve success. As previously discussed in chapter 2, these cognitive “expectations” can affect how students respond to an introductory physics course. Students’ expectations affect what students choose to listen to and what they ignore in the firehose of information provided during a typical course. Students’ expectations also affect how students build their own understanding of the course material and the type of understanding they build. All of this plays a major role in what students take away from a course.

B. Why study expectations?

Although we don't often articulate them, most physics instructors have expectation-related goals for their students. In the calculus-based introductory physics courses, we try to get students to make connections, understand the limitations and conditions on the applicability of equations, build their physical intuition, bring their personal experience to bear on their problem solving, and see the link between classroom physics and the real world. These kinds of learning goals — goals not listed in the course's syllabus or the textbook's table of contents — are part of the course's hidden curriculum. Even when articulated explicitly, course goals like these are often not
adequately reinforced through testing and grading. Yet, they are key components of our main goal of helping students learn a useful, functional understanding of physics.

As instructors, we are frustrated by the tendency many students have to seek “efficiency” — to achieve a satisfactory grade with the least possible effort — often with a severe unnoticed penalty on what and how much they learn. In the examples of Hammer and Tobias in chapter 2,\(^1\) the structure of the traditional lecture course seemed to encourage students to pursue this so-called efficient strategy. These traditional lecture courses emphasized algorithmic problem solving. This played into students’ beliefs that conceptual understanding and problem solving are separate and that the latter is more valuable for success in the course. The courses also discouraged debate and discussion of the course material. Understanding and cooperative learning were not valued by the course format. Some students with expectations more favorable to achieving our main goal of functional understanding, those that tried to make sense of the course material for themselves, did not find their efforts supported or encouraged by their courses.

In the same way that research into students’ understanding of physics concepts has resulted in new curricula which improve students’ conceptual understanding of physics, it is hoped that better awareness and understanding of the role of student expectations in introductory physics can lead to further improvements in instruction. Although most studies, including Hammer’s, are short term and show no evidence that students change their expectations over time, there are some indications that such changes do occur. In the long term studies of Perry\(^2\) and Belenky \textit{et al.}\(^3\) (described in more detail in chapter 2), they frequently found their young adult subjects starting in a
“binary” or “received knowledge” stage in which they expected everything to be true or false, good or evil, etc., and in which they expected to learn “the truth” from authorities. Both studies observed their subjects evolving to a “consciously constructivist” stage. In this stage, the subjects accepted that nothing can be perfectly known, and accepted their own personal role in deciding what views were most likely to be productive and useful for them. In introductory physics courses, the analogous situation to the evolution of students’ general epistemology is a change in the way students view physics and what they do to learn it. For example, we would like students to see physics as less of an exercise of taking disjointed facts and problem solutions handed down by authority (as represented by the instructor and the textbook) and more of an internal process to build a deep understanding of a few key concepts that can be used to understand many situations in the students’ real world. In this “constructivist” physics stage, students do not just accept the pieces they are given by authority but try to make sense of it for themselves and fit the pieces into a coherent framework.

C. Description of Study

This study is one attempt to better understand the evolution of student expectations towards the constructivist physics stage and on how student expectations are affected by traditional and enhanced instruction. To evaluate the distribution and evolution of student expectations in different types of courses, we developed the Maryland Physics Expectations (MPEX) survey, a Likert-style (agree-disagree) questionnaire designed to probe some aspects of these cognitive expectations (see chapter 5 for more information on the development and validation of the survey).4 For
this study, the MPEX survey was used to measure the distribution of student views at
the beginning of introductory calculus-based physics classes (pre), at the end of the first
semester or quarter (mid), and at the end of the first year (post) for classes using the
three research-based curricula and traditional lecture instruction (the curricula and
implementations are described in chapter 8). I distributed, collected, and processed
MPEX surveys from the ten colleges and universities participating in this study. Other
members of the Physics Education Research group at University of Maryland checked
the processed data for accuracy against hard copies of the students’ responses. The data
was analyzed by Redish, Steinberg, and the author using both traditional and innovative
techniques (see the description of the Redish plot in section E below and in chapter 5). I
conducted interviews with students at five of the ten schools including at least one
school using each or the three research-based curricula. The schools involved, the
teaching methods used, the number of students in the sampling, and the types of data
taken are summarized in table 8-4. (Summary descriptions of the schools, courses, and
course contents can be found in tables 8-1, 8-2, and 8-3 respectively.)

D. Research Questions

Because so little is known about the distribution, role, and evolution of student
expectations in the university physics class, many questions can be asked. To limit its
scope, this chapter is restricted to three questions.

Q1. How does the initial state of students in university physics compare with the
views of experts?

Q2. To what extent does the initial state of a class vary from institution to
institution?
Q3. How are the expectations of a class changed as the result of instruction with the three research-based curricula as well as traditional lecture instruction at our ten participating schools?

Other questions, such as how individual students’ expectations evolve, the relationship between student expectations and success in the course, and how the survey results compare with other evaluations of student learning are left for further analysis and future studies in this area.

E. Explanation of the MPEX survey results

Although the survey itself uses a five point Likert scale (strongly disagree = 1 to strongly agree = 5), we have chosen to group the student responses into three categories: agree, disagree, and neutral. The reasons for this transformation are discussed in chapter 5.

As described in chapter five, the responses of undergraduate physics instructors committed to implementing research-based active learning teaching methods in their own classrooms are considered here as expert responses. Almost 90% of our expert responses agreed with a particular position for each survey item. The preferred responses of our expert group are defined as the “expert response.” A response in agreement with the expert response is defined as “favorable” and a response in disagreement with the expert response is defined as “unfavorable.” When data is presented in tables it will be presented as (% of favorable responses) / (% of unfavorable responses).

The “agree-disagree” (A-D) or “Redish” plot introduced in chapter five is also used to display the MPEX. In these plots, the percentage of respondents in each group answering favorably is plotted against the percentage of respondents in each group
answering unfavorably. More detailed comments on how to read a Redish plot can be found on page 167.

In addition to the overall result, the MPEX survey has clusters of items that probe six dimensions of expectations (described in detail in chapter 5): concepts, independence, coherence, reality link, math link, and effort. The first three dimensions were used by Hammer (see chapter 2)\textsuperscript{7} to classify student beliefs about the nature of learning physics. The survey items are described in chapter 5. A full copy of both the scantron and pencil & paper versions of the survey are included in Appendix B.

In order to eliminate the confounding factor of differential dropout rates from the evolution of student expectations, unless otherwise specified, the MPEX data presented in this chapter is matched.\textsuperscript{8} Our results show some differences among different classes at the same institution, but the variation is statistically consistent with the sample size. To simplify the presentation of the MPEX results, the individual class results have been combined for similar classes at a given institution.

F. Site visits and Interviews

During the course of this investigation, I conducted over 120 hours of interviews with student volunteers at the University of Maryland, the Ohio State University, Dickinson College, Drury College, and Nebraska Wesleyan University. The interviews followed the MPEX survey or the Open MPEX open protocols described in chapter 7. The students were interviewed individually or in groups of two or three. If there was sufficient time the students were asked to work individually on a problem as a think aloud exercise (see chapter 7), usually the two-rock problem described in previous
chapters. These interviews were conducted mainly to validate the students’ interpretations of the MPEX survey items and to understand the reasons behind their answers. The interviews also revealed a great deal about the students’ expectations and their perceptions of the material and the teaching method used. The interviews with the students at the Workshop Physics schools were particularly interesting and are discussed in detail later in this chapter.

G. Chapter Layout

In section II, I present the distribution and evolution of student expectations obtained with the MPEX survey from introductory calculus-based physics sequences at each of the ten schools. Both the overall survey results and the results from the six dimensions or clusters are discussed. For each survey dimension, the pre student responses are compared with the following:

- with the expert responses to address question Q1,
- with each other to look for differences in student populations to address question Q2, and
- with the mid and post results to look for changes to address question Q3.

In section III, expectation results from interviews with students at three of the schools using Workshop Physics (Dickinson, Drury, and Nebraska Wesleyan) are discussed in detail. Note that this includes the school where the Workshop Physics curriculum was developed (Dickinson) and two secondary implementations.

Section IV is a chapter summary going over the results in the context of the three research questions and discussing the implications of the results.
II. STUDENT EXPECTATIONS: DISTRIBUTION AND EVOLUTION

To understand the effects of the instruction in the introductory physics sequence on students, one needs to measure the initial state of the student coming into the sequence and their state at later times during the sequence. In this section, we discuss the pre, mid, and post sequence results obtained from students taking the MPEX survey in the calculus-based introductory physics sequence at their institution. A summary of the schools, classes, and the number of students surveyed is shown in Table 10-1.

Pre MPEX survey data from our “expert group” and the US Physics Olympics Team from are included for comparison with the students’ initial state as measured by the pre results. To better compare the pre results with the later measurements, the pre-course data presented here only includes students who took the survey at the beginning and end of the first semester or quarter of introductory physics. Thus, these pre student responses are only from students who completed at least the first course in the sequence. The overall pre results and the pre results from the six clusters are presented in Table 10-2. To observe differences in the initial state at different schools, the University of Maryland data is used as a baseline. A school whose MPEX score is at least $2\sigma$ different from Maryland is considered significantly different. (The estimation of the uncertainty $\sigma$ is described in chapter 5)

The matched pre/mid and pre/post MPEX survey results are used to learn about the effects of the different methods of instruction on student expectations. The two types of matched data are included for the following reasons. Data matched over one semester allows for a more controlled study. Since over one semester or quarter
Table 10-2
Table 10-3
students have only one instructor, this allows us to look at the effect of a single teaching style on a particular class. Matched pre-mid survey results are shown in Table 10-3. (Because of logistical problems, no mid MPEX data was collected from the traditional class at Carroll College.)

Although the introductory courses participating in this study vary in content, they all begin with Newtonian mechanics. While there are some differences in content coverage, for the most part all the first term courses cover the same main ideas and concepts. This allows for a large degree of control for comparing classes and allows us to focus on the effects of the implementation of a curriculum by a particular instructor. However, my interviews and observations by the members of the University of Maryland PER group suggest that for at least some students, more than one semester or quarter may be needed to produce significant changes in student expectations.

The pre and post matching over one year allows us to look for longer term effects. A period of one year allows us to look at student responses from the beginning and end of the introductory sequence at all the participating schools except University of Maryland. Although the University of Maryland sequence has three semesters, post data is used from the end of the second semester so that all students reported on in this section will have had approximately equal class time. For eight of the ten schools matched pre/post data are presented in Table 10-4. Because of logistical problems, no post MPEX data was collected from the Workshop Physics classes at Skidmore College or the traditional classes at Prince Georges Community College and University of Maryland. However, we do have data from one group of students at Maryland who had
traditional lecture instruction for the first semester and Tutorials for the second semester. The curriculum for these students is listed as TRD-TUT.

A. Overall Results from all schools

The overall survey results for the ten schools are presented in Redish plots in Figures 10-1 and 10-2. In order to simplify the reading of the graphs, we have displayed the results from the three large research universities in one part of the figure and those from the smaller schools in another. The pre-course results are shown with green markers, the post-course results with red markers. A cross shows the result of the expert group.

We make three observations.

1. *The initial state of the students at all the schools tested differs substantially from the expert results.*

The expert group was consistent, with 87% agreeing on which survey responses were desirable for their students. Except for the WP class at DRY, beginning students only agreed with the favorable (expert) responses about 50-65% of the time, a substantial discrepancy. Furthermore, students explicitly supported unfavorable positions about 10-25% of the time.

2. *There are some significant differences in the student populations.*

Three of the student populations have significantly more favorable expectations than the students at University of Maryland (DRY, NWU, & MIN). The students at Drury College (DRY) responded even more favorably than the Physics Olympics Team (79% vs. 68%). However, this is an extremely small class (even for Drury) with a corresponding large uncertainty that is shown to demonstrate what is possible but
Figure 10-1
Figure 10-2. Pre/Post Redish Plots for all schools, average of all items
should not be considered typical. In general, coming into the introductory physics sequence the students at the small liberal arts (SLA) schools and at University of Minnesota seem to have more favorable expectations than students at Maryland, Ohio State, and the community college.

3. In all cases, the result of instruction on the overall survey was an increase in unfavorable responses and a decrease in favorable responses (though some changes were not significant). Thus, instruction produced an average deterioration rather than an improvement of student expectations.

Three sequences showed a significant decrease in favorable expectations overall. These include TRD instruction at Minnesota and both GPS sequences (MIN & OSU). Note that both GPS sequences had significantly better fractional gains on the Force Concept Inventory (FCI) than the TRD classes. In fact, the GPS sequence at Minnesota was considered by the physics education group there to be a very successful implementation.10

The overall survey includes items that represent a variety of characteristics, as discussed in chapter 5. To better understand what is happening in the classes observed, let us consider the initial state and the change of student expectations in our various clusters. The clusters and the associated survey items are discussed in detail in chapter 5.

**B. The Independence Cluster**

One characteristic of the binary thinker, as reported by Perry and BGCT (discussed in chapter 2), is the view that knowledge comes from an authoritative source, such as an instructor or a text, and it is the authority’s responsibility to convey this knowledge to the student. More mature students understand that developing knowledge
is a participatory process. Survey items 1, 8, 13, 14, 17, and 27 probe students’ views along this dimension. The pre/post results for this cluster are displayed in a Redish plot in Figure 10-3.

Our expert group responded favorably to the survey items in this cluster 93% of the time. On this cluster, students' initial views were favorable in a range from 36% (TYC) to 59% (SKD). For comparison, the POT showed favorable views on these items 81% of the time. Only the students at Drury and Minnesota showed any significant change as a result of instruction. The WP students at Drury and both the TRD and GPS students at Minnesota showed significant decreases in the percentage of favorable responses after one year on instruction. Note that Drury students had exceptionally favorable pre expectations in every dimension and ended with expectations that were still among the most favorable.

Survey items 1 and 14 are particularly illuminating and show the largest gaps between experts and novices.

1. All I need to do to understand most of the basic ideas in this course is just read the text, work most of the problems, and/or pay close attention in class.

14. Learning physics is a matter of acquiring new knowledge that is specifically located in the laws, principles, and equations given in the textbook and in class.

The expert group was in 100% agreement that students should disagree with item 1 and in 84% agreement that they should disagree with item 14. Disagreeing with these items represents a rather sophisticated view of learning, but favorable shifts on these items are exactly the sort of changes that indicate the start of a transition between a binary and a more constructivist thinker. The interviews strongly support this view. Students who
Figure 10-3. Pre/Post Redish Plots for all schools, independence cluster
disagreed with these items were consistently the most vigorous and active learners.

This cluster of items, and items 1 and 14 in particular, appear to confirm that most students in university physics enter with at least some characteristics of binary learners, agreeing that learning physics is simply a matter of receiving knowledge in contrast to constructing one’s own understanding. We would hope that if a university education is to help students develop more sophisticated views of their own learning, that the introductory semester of university physics would begin to move students in the direction of more independence. Unfortunately, this does not appear to have been the case. In the touchstone items of 1 and 14, the only significant improvement was in the WP sequences at DC on item 14 (26% to 43%) and the TUT sequences at UMD (4% to 22%), the GPS sequence at OSU (7% to 13%), and the TRD sequence at MIN (12% to 16%) on item 1. The TRD class at MIN significantly improved on item 1 and significantly deteriorated on item 14.

C. The Coherence Cluster

Most physics faculty feel strongly that students should see physics as a coherent, consistent structure. They feel that a major strength of physics is its ability to describe many complex phenomena with a few simple laws and principles. Students who emphasize science as a collection of facts fail to see or appreciate the integrity of the structure. This lack of a coherent view can cause students many problems, including a failure to notice inconsistencies in their reasoning and an inability to recall information through crosschecks. Survey items 12, 15, 16, 21, and 29 were included to probe student views along this dimension. The pre/post results are shown in Figure 10-4.
Figure 10-4. Pre/Post Redish Plots for all schools, coherence cluster
Our expert group was in agreement as to what responses were desirable on the elements of this cluster 85% of the time. The initial views of students at our ten schools were only favorable between 48% and 58% of the time. Most sequences showed a small deterioration on this cluster, except for DC which improved slightly (57% to 63% favorable responses) and the two GPS sequences at MIN and OSU which deteriorated significantly (56% to 48% and 49% to 39% favorable responses respectively).

Two specific items in this cluster are worthy of an explicit discussion.

21. If I came up with two different approaches to a problem and they gave different answers, I would not worry about it; I would just choose the answer that seemed most reasonable. (Assume the answer is not in the back of the book.)

29. A significant problem in this course is being able to memorize all the information I need to know.

Item 21 is a touchstone. Coming up with two different answers using two different methods indicates something is seriously wrong with at least one of your solutions and perhaps with your understanding of the physics and how to apply it to problems. Our expert group and POT students feel strongly that students should disagree with item #21 at the 85% level. Initially, only 42-55% of the students responses were favorable for this item, and the only significant change is the decrease in favorable responses in the TUT sequence at UMD (55% to 46%) and both sequences at MIN (53% to 40% for the GPS sequence and 46% to 40% for the TRD sequence). No sequence showed any significant improvement on this item.

D. The Concepts Cluster

The group of items selected for the concepts cluster (items 4, 19, 26, 27, and 32), are intended to probe whether students are viewing physics problems as simply a
mathematical manipulation of an equation, or whether they think about the underlying physics concepts. For students who had a high-school physics class dominated by simple “problem solving” (find the right equation, perhaps manipulate it, then calculate a number), we might expect largely unfavorable responses to the items in this cluster.

However, we would hope to see substantial improvement in this cluster as the result of even a single college physics course. The pre/post results are shown as a Redish plot in Figures 10-5.

Our experts agree on their responses to the items of this cluster 89% of the time. The initial views of the students at the six schools were favorable between 42% (PGCC) and 57% (SKD) of the time. Only two sequences showed significant improvement on this cluster, the WP sequence at DC (54% to 62% favorable responses) and the TUT sequence at UMD (52% to 60% favorable responses) although two of the other WP schools showed some improvement as well. Both GPS sequences showed deterioration in this cluster although only MIN’s was significant (54% to 49%)

Within this cluster, the results on items 4 and 19 are particularly interesting.

4. “Problem solving” in physics basically means matching problems with facts or equations and then substituting values to get a number.

19. The most crucial thing in solving a physics problem is finding the right equation to use.

While these items are similar, they are not identical. Agreeing with item 4 indicates a naive view of physics problems or a lack of experience with complex

Figure 10-5. Pre/Post Redish Plots for all schools, concepts cluster
problems. A more experienced student could reject 4 but still agree with 19 because of the phrase “most crucial”. However, one would hope that increased experience with complex physics problems would lead a student to disagree with this item as well. For example, 54% of the POT students gave a favorable response on this item as compared to only 22% of beginning students at UMD. Our personal observations of these students indicate that as expected, the POT students have considerably more experience with complex problem solving than the typical beginning engineering student.

Most of the schools begin with favorable responses on item #4 of 35-58%. Our community college (PGCC) is an anomaly, with only 16% of the students responding favorably on this item. This suggests that the group of students at PGCC may be considerably less sophisticated, at least along this dimension, than the average beginning university student. The shifts on this item for the TUT sequence at UMD and the WP sequences at DC and NWU were favorable and significant (e.g., UMD 46% → 68% favorable, DC 47% → 75% favorable, and NWU 46% → 68%). The other sequences deteriorated with the exception of DRY, which did not change significantly but started with a very high value.

All groups showed a low initial favorable response on item 19 (13% (TYC) to 33% (MSU)) but all showed a significant deterioration after one year of instruction except for MIN-GPS, and CAR-TRD which shifted favorably but not significantly and OSU-GPS which did not change.
E. The Reality Link Cluster

This cluster looks at student beliefs about connections between physics and their personal experience. The four items (items 10, 18, 22, and 25) that compose the reality link cluster do not just probe whether the students believe the laws of physics govern the real world. These items also probe whether students feel that what they learn in their physics course is relevant to their personal real world experiences and vice versa. The pre/post results are shown in a Redish plot in Figure 10-6.

Our expert group of instructors was in almost unanimous agreement (93%) with the favorable response on our reality cluster. Interestingly, the POT students only gave favorable responses at the 64% level. Examining their written comments as well as their responses gives one possible explanation. The POT students saw physics as being associated primarily with interesting and exotic phenomena, such as cosmology, relativity, and particle physics, and they did not see a link between this physics and their personal experiences.

The student groups at our six schools started out with fairly strong favorable responses, ranging from 61% (UMD) to 82% (CAR). Unfortunately, every group (except DRY) showed a deterioration on this measure as a result of instruction, and many of the shifts were substantial (CAR-TRD from 82% to 50%; MIN-TRD from 71% to 59%; MSU-WP from 71% to 44%; SKD-WP (over one semester) from 75% to 56%; and MIN-GPS from 72% to 53%). The WP students at NWU started at 80% favorable responses and ended at 77% favorable, the highest of any student group in this study. I will come back to this point when we discuss the interview results.
Figure 10-6. Pre/Post Redish plots for all schools, reality link cluster
F. The Math Link Cluster

One of the hidden goals of the introductory physics course is to help students develop the ability to use abstract and mathematical reasoning in their study of physics. Experts view mathematical equations as concise summaries of relationships between physical quantities. Many students fail to grasp these relationships and instead focus on just using the right equation to calculate the answer to a problem.

The survey items probing students' apparent expectations of the role of mathematics are 2, 6, 8, 16, and 20. The pre/post results are shown as a Redish plot in Figures 10-7.

Our expert group is in strong agreement on the favorable answers for this cluster, agreeing at the 92% level. Since high school physics courses tend to be decidedly less mathematical than university physics courses, we were not surprised with the initial response of the students in our test classes, which range from 57% to 71%.

Although these lower expectations may be appropriate for high school students and therefore for beginning university students, one might hope that these attitudes would change towards more favorable ones as a result of a university physics class. Unfortunately, none of the sequences probed show improvement in the favorable/unfavorable ratio and six (CAR-TRD, MIN-TRD, MSU-WP, NWU-WP, MIN-GPS, and OSU-GPS) show a significant and substantial deterioration.

Among the items of the cluster, the results on item 2 are particularly interesting.

2. All I learn from a derivation of a formula is that the formula obtained is valid and that it is OK to use it in problems.
Figure 10-7. Pre/Post Redish Plots for all schools, math link cluster
From our interviews and informal discussions, we note that many students today have had little or no experience with formal mathematical proof. A few did not understand the meaning of the word “derivation,” mistaking it for “derivative.” This lack of experience can produce a severe gap between the expectations of instructors and students and cause serious confusions for both groups. On item 2, the students at no institution showed favorable responses (disagree) at higher than the 53% level (CAR). At PGCC, only 20% of the students gave a favorable response with item 2 initially. They improved somewhat after the class (to 33%). Two other sequences also improved significantly on this item (DC-WP from 35% to 47% favorable, MIN-TRD from 36-44% favorable), but three sequences deteriorated significantly (CAR-TRD from 53% to 33% favorable, MSU-WP from 42% to 17% favorable, MIN-GPS from 42% to 36%). The improvement at DC implies that the deterioration at MSU may not be associated with the Workshop Physics structure, which tends to emphasize hands-on and laboratory activities over purely abstract and mathematical reasoning.

G. The Effort Cluster

Many physics lecturers will expect students to use whatever resources they have available to make sense of the material. Unfortunately, many students do realize that if they do not see something right away, there are steps they can take that will eventually help them make sense of the topic.

An important aspect of the hidden curriculum is to realize that one's current understanding might be wrong, and that the mistakes one makes can be useful in helping to correct one's errors. This dimension is probed by items 3, 6, 7, 24, and 31 on
Figure 10-8. Pre/Post Redish Plots for all schools, effort cluster
the survey. For this cluster, the pre/post results are displayed in a Redish plot in Figure 10-8.

Our experts are in strong agreement on the answers to the items of this cluster, at an 85% level. The initial views of the students at the various institutions begins quite high, ranging from 66% favorable (at OSU) to 88% favorable (at CAR). By the end of the year, the shift is dramatically downward, with three institutions dropping in the favorable percentages by roughly 20% (UMD-both sequences, DC-WP, NWU-WP, MIN-TRD, and OSU-GPS, and PLA) or more (CAR-TRD 39%, MSU-WP 30%, DRY-WP 29%, MIN-GPS 47%). In one sense, this may be interpreted that the students expected to make more of an effort in the course then they actually did, as the shifts were largest on items 3 and 6, but the downward shifts on items 24 and 31 were also substantial.

III. WORKSHOP PHYSICS: SITE VISITS AND INTERVIEWS

During the course of this study I was able to make site visits and interview students at three institutions using Workshop Physics: Dickinson College (DCK) and two of the adopting schools, Nebraska Wesleyan University (NWU) and Drury College (DRY). Note that no survey data is shown for Drury College earlier in this chapter because the unusually small class size of their calculus class (eight students) made for unusual results and large uncertainties. However, the interviews I had with students from both the algebra/trig-based and the calculus-based Workshop physics classes provides another view of an adopting institution. To study their expectations, students
were interviewed with either the MPEX Survey protocol or the Open MPEX protocol, both described in Chapter 7.

A. Dickinson College: Pre-course and mid-year

I conducted pre-course, mid-year, and post-sequence interviews with students in the Workshop Physics classes at Dickinson College during the 1994-5 and 1995-96 school years. In the 1994-5 school year, eight students were interviewed at mid-year at the end of the fall semester about a week before finals. The students were asked questions from the MPEX Survey protocol in pairs and some were individually given problem interviews (see chapter 7 for a description of problem interviews) including the two-rock problem. In the 1995-96 school year, pre-course interviews were conducted in the second week of the fall semester and post sequence interviews were conducted a week before finals in the spring semester. In the pre-course interviews, students were again asked questions from the MPEX Survey protocol to validate the survey items at the beginning of instruction. Some additional open-ended questions concerning their views on the nature of physics and mathematics were also asked. In the post-sequence interviews at the end of spring semester, the Open MPEX protocol was used with six students including four who were interviewed previously. The students were all volunteers who were paid a small stipend by Dickinson for participating in the interviews. Note that the student sample was not randomly selected. While I will not go into the interviews in detail there are a few points that should be mentioned.

In the first pre-course MPEX survey results from Dickinson College from the 1994-95 school year, we noticed that their students’ pre-course results were substantially
more favorable than the other schools participating in the study at that time. In the subsequent years, the survey was given before the course instructors gave their course overviews on the first day of class. The results were unchanged. However, in the pre-course interviews described, most of the Dickinson students were found to be familiar with at least some aspects of the Workshop Physics course either from contact with the physics faculty or discussions with friends who had already taken the course before the first day of class. In addition, other departments now offer classes using the workshop approach including calculus, which is a co-requisite for the Workshop Physics course. I hypothesize that either Dickinson is attracting students with more constructivist expectations or this foreknowledge of how the class is run is affecting the survey results.

Two surprising expectation findings came out from the mid-year interviews. First, in addition to the conceptual difficulties discussed in connection with the two-rock problem in chapter 9, the students were very reluctant to demonstrate their predictions to the two-rock problem mathematically. Instead they wanted to go into the Workshop Physics Classroom and set up an experiment. In light of the dominant role of the laboratory in Workshop Physics, this attitude is understandable and perhaps not even an issue for liberal arts students. But for students in science and engineering majors, this prejudice against using mathematical models for predictions before an experiment is a serious concern.

The second finding was that the majority of students interviewed did not see the need to reflect on or reconsider the course material. These students felt that once they had learned the material through the appropriate activities, they knew it and could move on. In physics, the knowledge and understanding of experts is always under
reconsideration as new ways are discovered of looking at things and as experts extend what they know to new areas. This finding suggests further probing into this issue and, perhaps, future work into assessing the knowledge structures students build in introductory courses more directly.

B. Nebraska Wesleyan University: Post-Sequence

Nebraska Wesleyan University completed their second year of implementing Workshop Physics in the spring 1997 semester. As described in Chapter 8, their implementation is unique in that they only teach one flavor of introductory physics, algebra/trig-based and calculus-based in the same class. I made a site visit two weeks before finals in April 1997 where I attended one session each of both Workshop Physics classes and interviewed 10 student volunteers.

During my class observations I noted some similarities and differences with the Dickinson implementation. The room layout was conducive to group activities and class discussion as well as lecture and demonstrations. The students worked in groups of three or four on opposites sides of lab benches aligned parallel to the front of the room. Each group had a computer and a lab set up. The day’s lesson was on flux and was very difficult for most of the students. The morning had an additional facilitator, a graduate student who was working part time. Thus, there were two instructors available to help the student groups. In the afternoon when the professor was alone, he used what he had learned from the students in the morning to give them a better, more detailed presentation before the students began the group lab activity. Despite this, several groups had to wait for the instructor to assist them when they got stuck. This is a major
difference between the implementation of Workshop Physics at adopting schools and Dickinson College. At Dickinson, they use undergraduate TAs to assist the student groups in the classroom so that there are three facilitators in the classroom at all times. There are ways to address this problem without adding staff, but they require adjustments and fine-tuning over a period of time.

Another difference was that the instructors at NWU had just begun customizing the material this past year to meet the need of their students while Dickinson has been running this curriculum since the late eighties. The instructors at NWU are aware of their current problems and are working to fix them. They will need at least one more iteration of the research and development cycle to make their activity guide and the class better suited for their unique bi-level approach. This is not to say that the course is not doing well, just that there are difficulties that need to be addressed. Part of this evaluation is to help identify less obvious difficulties.

I met with all ten student volunteers separately. Nine of the ten were interviewed using the MPEX Survey protocol. The tenth student was interviewed with the Open MPEX protocol. The results were very intriguing. Note that the analysis presented below is more detailed than for other interviews in this chapter. There are several reasons for this. First, this was the only site visit to a school adopting Workshop Physics that met the study criteria, twenty or more students in a calculus-based class closely following the implementation at Dickinson College in a similar environment. Since Workshop Physics used guided-discovery laboratory as the primary method of instruction, we would expect Workshop Physics to have the largest favorable effect on student expectations. As we saw from the survey, only Dickinson showed any
significant improvement in student expectations (in the concepts cluster). However, the WP sequences at DC, NWU, and DRY ended the year with more favorable expectations than the other sequences. Also, DRY and NWY had less deterioration (in units of $\sigma$).

Second, these were the only MPEX interviews conducted with a good sample from a sizable calculus-based class that used the most current version (version 4.0) of the survey. Third and last, two additional controls were added to address concerns raised in the earlier interviews.

- To prevent the interview itself from unknowingly influencing the student responses, the students completed the survey before starting the interview (No more than 24 hours before)
- This student sample was very representative of the entire class. At least three students were interviewed from the top third, middle third, and bottom third of the class. The students were rated by the instructor based on their overall grade at the time of the interviews.

As the interview transcripts are too long to include in this dissertation, transcript summaries of the interviews (organized by topic) with all nine students interviewed with the MPEX Survey protocol are included in Appendix D. In addition, their interview responses to the survey items listed by item are included in Appendix C. Nine of the ten students interviewed and eight of the nine interviewed with the MPEX Survey protocol were majoring in biology, biochemistry, or biopsychology.

While all the interviews were useful for learning about the students’ expectations, the five interviews described below were selected to see what the students’ expectations are like at the top and bottom of the class.
1. The view from the top: Charlie, John, & Amy

The results from the three students in the top third of the class were particularly interesting. The three pre-med students, code-named Charlie, John, and Amy, had very different expectations and views of Workshop Physics.\textsuperscript{13} All three were very bright, motivated, and very articulate students.

Charlie:

Charlie is a pre-med molecular biology and biochemistry major. He had no prior knowledge of workshop physics or the workshop approach until he showed up in class. In high school, he had math through calculus and five years of science including one year of physics.

During the interview, I found Charlie to have generally very favorable expectations. He is strongly constructivist. This is reflected in his responses to items 1, 14, and 34 from the MPEX survey shown in table 10-5. Note that in all three of his responses he refers to building his own understanding. He recognizes that at least sometimes he needs to step back, reflect, and piece things together himself. He also uses physics to understand the real world and tries to tie the physics he is learning to familiar experiences. He both values derivations and goes over them on his own. He goes over both his homework and exams so that he can understand his mistakes and correct them.

He prefers the Workshop Physics methods to the traditional methods used in his other science courses. That is not say he thought the course was perfect. While his first semester group included his friends who were also well motivated to learn and
Table 10-5a: Interview responses for item 1 from students in the top third of their class at NWU.

Item 1. All I need to do to understand most of the basic ideas in this course is just read the text, work most of the problems, and pay close attention in class.

**Favorable Responses**

Charlie: *And I disagreed with this one. Uh ... basically because the under- — the big word there that got me was "understand." To understand it, just reading and doing the problems and going to — paying attention in class just feels like going through the motions, to me. To understand it, I think it takes a lot more — critical thinking — going — going back and going over it in your mind and stuff, because you'll have new concepts, I think, and stuff that sometimes just doesn't go over well the first time you see it or consider it. A lot of it just takes rehashing.* [Okay. Now, what do you mean by "critical thinking"?] ... *I mean not just memorizing formulas or taking things because they're presented, you know, just to be true; to go a step beyond that and maybe why or how this is — how they can — how come the book can come and say this, or where they get this from. If they give you a reason, try and understand it. I mean you — not just take it as fact. Try and understand why it is.*

**Unfavorable Responses**

John: *And I put "strongly agree," number five. And ... that's just ... all I really need, usually, for most problems. There're a few that I ... have to ask questions on — although it has happened. But most of the time, I can just read the text. Even if I have to read it over and over, I'll get it eventually.*

Amy: *And I said "strongly agree," because if you were to do all of these things, you're pretty much guaranteed to learn it.*
Table 10-5b: Interview responses for item 14 from students in the top third of their class at NWU.

14. Learning physics is a matter of acquiring knowledge that is specifically located in the laws, principles, and equations given in class and/or in the textbook.

Favorable
Charlie: *This one I disagreed with. And I think I disagreed with this one because it just sounds like this means that you're just memorizing the — the basics. And — and, again, I like to go beyond that and — and go to the understanding of where it all comes from — where the laws, principles and equations arise from. And so I disagreed with that one.*

Neutral
John: *Three, "neutral." Again, I think it depends on the type of person you are. Learning physics as a whole — really learning physics — is more than just ... principles, equations, laws. It's ... in understanding and incorporating all this. But to pass a course, learning physics, I think that's all you really need to get by. And I think some people maybe aren't interested in it and are — or maybe even aren't capable. It just doesn't — doesn't do it for 'em, and so they don't think that way — which isn't bad. It doesn't mean they're not smart. It just means that that's a different way of thinking ... they don't think about that. So, therefore, I ... just chose "neutral."

Unfavorable
Amy: *I agree with that. Learning the fundamentals of it — you've got to do that.*
Table 10-5c: Interview responses for item 34 from students in the top third of their class at NWU.

34. Learning physics requires that I substantially rethink, restructure, and reorganize the information that I am given in class and/or in the text.

**Favorable**

Amy: I agree. ... If you were given a problem to solve ... you definitely need to figure out, you know, "In what area do I need to apply what I know? And what part of that area do I need to pick at most? And what do I need to solve for to get this problem solved?" So, you're constantly picking apart and thinking about what you've learned and what you ... can apply from what you know — along with, you know, of course, submolecular level you can't, hands-on, say what's going on; but from what — Your creativity, I guess would come into that point as far as being able to restructure what you've learned from ... Just the same as the people who would come into a test with photocopied answer sheets from the homework problems. They're not restructuring, rethinking, reevaluating what they've learned. They're trying to get the answers to all the problems and hope there's problems like that on the test that they can, quick, go through and pinpoint something, so they can plug and chug something. That's not rethinking or reanalyzing. Those people won't do well, and they won't get an 'A' out of the class. So you have to do that.

**Neutral**

Charlie: And I was neutral also, because I — there's times when I have to restructure and rethink; there's times when I don't. And it just depends on how clear the material is presented or how well I understand it at the time. Sometimes it just hits me all — real fast and hard, and I just have to — "Woah!" — step back from it for a while and go back later and slowly pick at it ... take it a step at a time.

**Unfavorable**

John: "Disagree." Again, I took that question to mean that you ... have to sit down — I know people who will go through a class — right? Then they'll go back and they'll rewrite out notes on a page, from the textbook, or from the class, or lecture. I don't need to do that. Usually, just when I get in class and what I read out of the book, homework problems I have to do — that's good enough.
understand the material, his second semester group partners were not as serious. In his words,

_Because in the first semester, I was with three friends, and the — You know, we were a group of three. So, I was with two other guys. And one of ’em was going on to physical therapy, and the other was going to take MCAT with me. So, it was really important — We knew we had to really learn and understand some material. And we were friends, so we could criticize each other and — and it was just a lot easier to work with than to — We ... just were there for business, it seemed like. Even though we were friends. It sounds like you would be more apt to goof off, but we're all good students and we were serious about why we were there. And we got a lot done. This second semester, I'm with three guys that have — I just met this semester, really. And it's a lot harder, because they're not as serious about it, it doesn't seem like, it seems like we spend a lot more time goofing off or just chatting and get sidetracked. And — And then when it comes to understanding stuff, sometimes they don't seem to be hung up on understanding as much as I am, so that they'd rather just — You know, they're there to get it done, get through the stuff so they can leave. Whereas, it takes me a while longer to get through stuff and to keep up with them, because I want to understand things as I go along._

Although he was fairly constructivist, he felt that it was not necessary to do experiments to introduce a topic. For example, he felt that in the case of learning that there were two kinds of charge that it was pretty self-explanatory and _that [it] just seems like stuff that could be, you know, told you. ... I’d just rather go spend more time on more complicated stuff than the real simple._

**John & Amy:**

John and Amy were interviewed separately, but they have similar mixed expectations. They are both pre-med biology majors (Amy is actually a biochemistry major) who were both successful with at least one semester of traditional undergraduate physics instruction prior to transferring to NWU from other schools (Iowa State University and University of Nebraska, Lincoln, respectively). Both had heard about the
Workshop Physics course before taking the class, John from last year’s students and
Amy from the faculty. In high school, John had math through AP Calculus and at least
four years of science including two years of physics. Amy did not discuss her high
school background in detail but she did not have physics or calculus in high school.
However, after her first semester of traditional physics in which she received an A+, 
Amy considered becoming a physics major. She avoided taking calculus until this year
(she did not want to take it).

They both are more transmissionist than constructivist in their view of physics
learning and knowledge. For example, their responses to survey items 1, 14, and 34 in
Table 10-5 indicate more of an emphasis on problems and equations than in building
their own understanding of the material. Although in item 14 John indicates that while
it’s not necessary for learning physics for the class, understanding is important for really
learning physics. And in item 34 Amy states that she is rethinking and restructuring what
she is learning and what she knows while she solves problems.

Another indication of their mixed expectations is that while they both believe it’s
important to see derivations, they don’t actually go through them themselves. While
John believes it’s important to see where equations come from and to see how the
concepts link to the equations, he does not believe it’s necessary to work out the algebra
(because then it becomes more of a math exercise) or that derivations are strongly
useful. Amy thinks it’s necessary to know [derivations] and be able to implement them.
However she doesn’t go over derivations in the book. This is not surprising since she
doesn’t read the book much unless it’s to clarify ideas. (Note that this is not that
unusual. Many students taking Workshop Physics at Dickinson use the activity guide as
their main text and pretty much ignore the assigned text. In fact for at least one year Dickinson did not use a textbook with the Workshop Physics class.) In class, she thinks it’s important to watch the instructor do derivations though she doesn’t ponder them. She thinks it’s important to watch because what I get out of ... watching someone derive a derivation or a proof is it gets me closer to being that type of person who’ll be able to do it, myself. I'm not, by nature, somebody who does that. It is interesting to note that the interviews hint that both Amy and John do not like math very much.

They are both in the binary stage\textsuperscript{14} at least with regards to physics knowledge. They both tend to see physics knowledge in terms of facts and situations. While both value concepts and understanding, both value the ability to apply equations to new, similar situations more. However, neither Amy nor John values building understanding from experience, both prefer to learn from authority, i.e. the instructor and the text. Also, both believe that the ideas in physics are relevant to other sciences and that science is coherent in that respect, but not that physics itself is a coherent framework of ideas. Amy’s link between physics and reality is as strong as Charlie’s is and she believes that thinking about the connection is an important part of the course. John’s reality link is also strong and he gives several examples of real world situations that use ideas from class. He believes that while the connection between physics and the real world is useful it was not essential for this course. Both of them prefer to learn the material and work out problem solutions on their own.

Not surprisingly, both John and Amy prefer traditional Lecture to workshop physics. John feels that WP might be good for learning the concepts in a course where the concepts are hard but for him this course was mostly review. Although John worked
with a “good group” where I’m telling them things sometimes and they’re telling me things sometimes, he would have preferred a faster pace to cover more material. John felt that physics had two parts, a basic fundamental part and an applied, complex, interrelated part. Like Charlie, he felt that while discovery learning wasted time on the basics, he thought discovery of complex information and applications was good. This is somewhat reinforced by his need to work in a group to learn physical chemistry which he describes as being like trying swimming and trying to hold your head up. Here he found working with a group useful because together we were able to solve problems we probably wouldn’t have been able to do individually.

Amy felt that she learns best on her own or one on one with a “knowledged” person such as a professor. She learns when things are explained to her by the book or by a professor. She would rather go off and read the book on her own than do Workshop Physics. However, Amy did not have a pleasant experience in the Workshop Physics course. Amy’s group partners were friends and they tended to exclude her from the group when both partners were present; although, Amy found she was able to learn effectively in Workshop Physics when one of her partners was absent. Based on this and other experiences with group work, she thought working in groups was helpful because no one knows everything and they could share information. However, when it comes to interactive activities, I think interactive work is appropriate in a research setting, when you already are knowledged, and maybe the two people are learning together, and you know, you have one professor working with you when you’re doing research. That, I think, should be group work.
It is interesting to note that in most traditional lecture classes both Amy and John would be considered ideal students. They have succeeded grade-wise in both the traditional class and the Workshop Physics class. They both learn well on their own. With the exception of their attitude towards derivations, many physics majors including myself have had similar expectations at the end of the introductory physics sequence. So what is the problem? What is wrong is wrong with the picture painted here? Nothing, if they were physics majors. At that point in their academic careers, many physics majors have learned to apply the equations of physics without a good understanding of physics concepts or a good understanding of what physics is (as discussed in chapter 2). It is only later iterations of core courses, preparing for the qualifier, or sometimes after teaching introductory physics, that physics majors develop a good understanding of physics and become more constructivist physicists.

But Amy and John are not physics majors and this is the last physics course they will take. While they are able to recognize physical situations similar to things they have seen before, it is doubtful they could apply the concepts of physics to unfamiliar situations. Their knowledge of physics is not as flexible as Charlie’s. In terms of the Hammer dimensions of constructivism, they believe that knowledge comes from authorities, that physics is weakly coherent in that it applies to other sciences, and that physics knowledge is based on equations and facts rather than concepts. I believe that their expectations of what physics is and how it is done may have prevented them from seeing the benefit of the Workshop Physics approach.
2. The view from the bottom: Hanna & Kim

Now that we have looked at the expectations of students in the top third of the class, it is instructive to look at the expectations of students in the bottom third to try to understand why they are not doing better. The three biology majors, code-named Hanna, Kim, and Roger, had expectations and views of Workshop Physics that provide an interesting contrast to the previous section. All three were having difficulties with the course. Their responses to items 1, 14, and 34 are shown in Table 10-6. Hannah’s and Kim’s interviews are discussed below. However, because Roger was an older returning student with a very unusual perspective, his interview is not discussed here.

Kim:

Kim is a senior majoring in biology preparing for a career in physical therapy. She is an Afro-American. She was the only minority student I interviewed at NWU and one of the only minority students in the class. She had no prior knowledge of the Workshop Physics course coming into the class. In high school, she had three years of science including one year of physics and mathematics up to pre-calculus including statistics. Like five of the seven students I interviewed at NWU who had high school physics (including Charlie), Kim did not get much out of her high school physics class. In her words, *The only thing I think I might have mastered was friction and the section in dealing with waves.* Unlike most of the other students I have interviewed, Kim has a history of learning problems with mathematics beginning in second grade. However, her problems with mathematics in this course are at least partly due to a lack of use and a lack of conceptual understanding. She took calculus two years ago at NWU but *it*
Table 10-6a: Interview responses for item 1 from students in the lower third of their class at NWU.

Item 1. All I need to do to understand most of the basic ideas in this course is just read the text, work most of the problems, and pay close attention in class.

Favorable

Roger: "Disagree." And why I said "disagree" is because I know it's just a straightforward science and it hasn't changed in so many years — just different philosophies of the actual premise of physics. But there're so many different angles that one question can take on. If I read the book — and since we have — we can write down anything we want, long as we generate it ourself [sic] and bring it into the test, I could copy that book word for word and take it into the test; but it's not going to help me. Because his question, or your question, or anybody else's question, most likely is not going to match up to anything that's in the text. So, it's more of a — You need more cognitive development than what a test is actually going to give you.

Neutral

Hannah: I marked agree. … I mean that covers, like, basically everything I do. I mean when I read this, I thought that covered, like, everything. I don't know what else you could do, besides read the text, work the problems and pay close attention in class. I mean, I guess I shouldn't say — Maybe I should disagree with it, because it's not all. You always have to think about it. You can't just, like, be a robot and do physics. But, mostly, that's...that about covers it. So, I don't know if I should change my answer or not. [That's something you have to decide.] … Okay. [She changes her answer to disagree]

Unfavorable

Kim: And I said "agree." … But that doesn't totally answer everything. That's part of what I need to do. So, that's why I picked "agree." I need to spend a lot of time with the teacher. So — but I need to read, work most of the problems and pay close attention. [What do you need to do with the teacher?] I have him — after I read, if I have problems with the — the problems. Or if I was trying to pay attention in class and didn't understand, I'd go to my professor for just verification — to see if I'm on the right track. [Okay.] Clear up any misunderstandings I have about the homework or the concept that we are doing...
Table 10-6b: Interview responses for item 14 from students in the lower third of their class at NWU.

14. Learning physics is a matter of acquiring knowledge that is specifically located in the laws, principles, and equations given in class and/or in the textbook.

**Favorable**

Hannah: *I think I should disagree with this, because it's more than just acquiring knowledge. You also have to put it all together. It just can't be a bunch of facts.*

[Okay. Why'd you initially agree with it?]

*Because, well, I don't know. Because I was in an agreeable mood. I mean I think that was the basic part of learning physics, but I guess to understand physics — which is, like, a deeper meaning of learning ... you have to be able to, like, put all that stuff together. So, then I disagree.*

Roger: *And I strongly disagreed. [Why do you strongly disagree with that?] ... I guess ... it's kind of coming to a sense of — Talking to other people that have taken physics, or — or hearing people in class, well, some people are getting it and some aren't. And some have stronger backgrounds than others. So, it's like, "Okay. Why is he getting it?" and I might not be. And why is this cluster not getting it? And, okay, they might've taken this, this and this — which then is allowing them to manipulate the formulas better than this group. That's why this — Even in the book — Even in the books, they don't actually say, "Okay, do this, and then next do this, and then do this. It's like saying, "Do this, this, and here you go." There's a — lot of — And even in teaching it, in — And that — My teacher's way — it's just, you know, a lot, in general, of ... taking for granted what — I'm losing my ... thought. [What students are able to do.] Yeah. Yeah, you know, it's like, "Okay. Here's the formula. Here's the end product. And you do this, this and this and this; but there's no "why" — why's and how's — to get to this. Or a lot of people, sometimes they'll go, "Okay, I get it to here, and you've lost me here." And that's probably a ... mistake for many of us in not going, "Okay. Why? Show me this part."*[So, it's ... not just the laws, principles and equations; but a lot of it has to do with the how and the why.] Um-hmm. [And based on your observations, you feel this depends a lot on what you bring into this class from the — from what classes you've had previously in your background.] Right.*

**Neutral** — No neutral responses

**Unfavorable**

Kim: *(Kim sounds like she is on the verge of a shift here.) I put "agree," and I think that ... Well, all things that we learn are the concepts, the laws and the principles and the equation — I guess it's a matter of how you acquire that knowledge. ... Okay. The way I'm taking the question would be is that physics is more than just memorizing the laws, principles and equations. To me, it's about how you learn. I'm not making sense. I agree with that statement — that it's a matter of understanding the laws, principles and equations; but I think it's all in how you learn it.*
Table 10-6c: Interview responses for item 34 from students in the lower third of their class at NWU.

34. Learning physics requires that I substantially rethink, restructure, and reorganize the information that I am given in class and/or in the text.

---

**Favorable**

Kim: And I put "agree." Ahem. Um ... Like, you can — with this, I took it to mean with all the different concepts we learn — they combine concepts together. And so you need to be able to take one concept and add it to another and make it all tie in together. So, that's why I chose "agree."

Roger: I said "strongly agree." And the reason why I say that is just the way it's ... taught — is that to make it sink in to any of the individuals here, we're restructuring and rethinking and formulating it ... into the way our mindset learns. And I think that's what I was thinking about, too.

---

**Neutral**

Hannah: I said "neutral," because ... it does require that I substantially rethink — it's — I don't know what "restructure" really means, but "reorganize" ... But the "substantial" part — I don't know what — to what degree. I'm not sure. That's why I put "neutral" — because I find out that, like, it does require that. I wouldn't say "disagree." ... But I'm just not — I put "neutral," 'cause I wasn't really sure to, like, what extent they actually mean.

---

**Unfavorable** — No unfavorable responses
hasn't helped me in this course because I don't remember half of the stuff I learned sophomore year. I haven't used it. I know what it is, but I can't explain it.

Kim has some favorable expectations for learning physics. She values building her own understanding and has some sense of physics as a coherent structure. She understands that different people can come up with different ways of looking at concepts or problem solving and that this is a creative process. She understands that looking at things from different points of view is helpful. As shown in Table 10-6, she is mixed in independence with indications of reliance on the professor to explain things in office hours.

She believes that physics is strongly connected to the real world and that this connection is strongly emphasized. However, she does not connect her real world experiences to the physics she is studying.

She strongly favors concepts over equations, but her math skills are very weak and this seems to be affecting her problem solving on exams and homework. In her words,

I said earlier sometimes I have a problem with tying in the concept with the math part of it. I think the concepts will tell me how it works or why it works this way, but sometimes I just have a problem tying in the math to give the mathematical answer. But I can look at the problem and see why and how the concept will work.

She is an example of a student who sees problem solving and conceptual understanding as separate. For example, when asked about which activities were most useful for learning physics, she replied,

Just to learn physics or pass a test? [To learn physics?] ... the things we do in class — the hands-on type of activities we do. To me, if you want me to learn physics, I think it'd be more understanding the
concepts than to plug in the math. Because physics is more than math. Physics is more than can you fill in the equation. Well, what good is it if I don't understand what it means? So, I think learning the concepts — which is what we do in class — more so than the math part — helps me to learn the physics. But on the test, it's more math problems than it is telling concepts. [So, what do you find most helpful preparing for the exams?] Working the homework problems. [Okay. Now, of all the things that you do, which would you say were least helpful for helping you learn physics?] None of it is ... I want to rephrase that. All of it is good to have. I can't say, "Throw out the homework," or, "Throw out the" ... But it just — It just depends on where you ... 'Cause the homework is good for the tests, but the concepts are good for just understanding — the understanding of physics. ... Both of them are good.

She reminds me of several students I have worked with in algebra/trigonometry-based introductory courses who have an intuitive feel for the concepts of physics but have difficulty connecting the math to the concepts to do the homework problems. However, even in problem solving she is moving towards the expert view. When addressing one of the survey items dealing with problem solving, she replied,

I was thinking more in terms of understanding the problem. Anybody can work a problem just plugging in numbers. But for me, when it comes to problem solving, I — sometimes I can't even draw the picture; so, it's more than just — for me, it's more than just plugging in the numbers. It's being able to understand where to start and how to even get to the right principle.

And later when asked whether finding the right equation is the crucial step in problem solving,

I put "agree," but I would — For me, I think I have ... to change it. Sometimes, the most crucial thing for me is setting it up, setting up the picture. ... For me, knowing what factors are in this picture to come up with the right equation. That's what I worry about.
She tries to make a qualitative representation to understand the problem, a characteristic of expert problem solvers discussed in chapter 2. Even here, she sometimes had problems,

Well, sometimes when you get ready to set up a problem, you can always draw pictures first, so you can see what you're working with. And sometimes I don't know where to start. I don't know how to draw a picture — the diagram — that goes with the problem. So, that's one of my biggest problems — being able to visualize it.

She values derivations and spends a lot of time on them. However, she seemed to be confusing a derivation with a derivative and her idea of spending time on them was to ask the professor about them. Also, she looks at example problems in the textbook similar to the homework problems but she doesn’t try to figure them out herself. However, she does try hard to work on homework problems alone even when she gets stuck. She meets with a group before homework is due to compare answers and also meets with them to prepare for exams.

When asked if learning physics in the Workshop Physics style would be an advantage to her in physical therapy school or in her later career, she responded,

Oh definitely. From this class, I've learned more than just physics. I've learned how to ask questions, how to try to come up with a method of solving problems — those types of things — critical thinking skills. Those types of things I get from this course. And then physics as a — the topic — you can use it in the real world as well. I learned to move things with physics.

She seems to appreciate Workshop Physics although she doesn’t like physics because she finds it very difficult. When asked what changes she would like to see in the course, she answered,

Um ... I think maybe in the what part of lecture we do have in class, maybe working more on math — tying in the math part with the
concepts. Not just coming up with an equation; but actually applying the equation to the problem in class, to have examples. We go over our homeworks sometimes, but I think I would put more math, actually tying the concept and the math together — besides just coming up — we do a lot of coming up with the equations, but we don’t really use the equation to the homework outside of the class.

Basically, this is a request for the instructor to model problem solving for her and perhaps for additional instruction to help her improve her problem solving skills.

Students like Kim with a weak understanding of math are unfortunately very common in algebra/trig-based and calculus-based introductory classes. They construct a qualitative understanding of physics that is not reflected in the homework and exams. They are often frustrated in traditional lecture classes and just do whatever is necessary to get a decent grade. I can’t determine from our interview if Kim is learning more in the Workshop Physics format, but she has developed more positive attitudes towards physics and the class.

Hannah:

Hannah is a junior pre-med Biology major. She was on the forensics team in high school and now competes at NWU. She had four years of high school science including one year of physics and high school mathematics through calculus. When asked about her high school physics class, she replied,

*I didn’t get very much out of it. It was kind of...There were only, like, six students; and it was, like, all boys. And they were just playing with, like, remote-control cars and stuff. And they wouldn’t let me (laughing) play with them, so I didn’t — I don’t know — I don’t feel like I got very much out of the physics class.*

Hannah filled out the survey in 10 minutes the day before the interview. As she was going through her answers in the interview she changed several of her responses.
On these items, she did not seem to have really thought about her answers. When asked about this, she responded that she was bitter about physics when she filled out the MPEX survey for reasons described below.

In her interview responses, her expectations seem very mixed but much closer to novice views than John or Amy. She shows indications of being in the binary stage; she was very concerned about giving the right responses during the interview. However, she is at least somewhat constructivist. Although she initially answered items 1 and 14 with unfavorable responses, she changed her mind while explaining her answers for both items, as shown in Table 10-6. She weakly indicates on some items that you need to take responsibility for constructing her own understanding, yet her emphasis seems to be on understanding what she is given by authority. Indeed, she seems to have trouble putting things together for herself. Surprisingly, she does not expect to intuitively understand anything. Some things she intuitively understands and some she doesn’t. She doesn’t seem to recognize that she can build an intuitive understanding of physics. This sounds very much like Belenky et al.’s description of one aspect of the binary state.\textsuperscript{15}

These women either get an idea right away or they do not get it at all. They don’t really try to \textit{understand} the idea. They have no notion, really, of understanding as a process taking place over time and demanding the exercise of reason. They do not evaluate the idea. They collect facts, but do not develop opinions.

Hannah is more focused on equations than concepts, yet she knows she needs more than just the equations. She recognizes that she is seeing the same math and situations again and again but doesn’t recognize the conceptual framework. However, she does know that two approaches to the same problem should give the same answer. She believes
derivations are important but only goes over them herself before exams. Also, while she feels that physics is strongly connected to the real world and that this is an important aspect of the class, she was not able to give specific examples.

She seems to have three main problems that are keeping her from doing better in the course. One, she recognizes that she doesn’t spend enough time studying for this class and going over the class lab activities (only two hours per week). As her study habits show, she crams for the exams (five hours of preparation for an exam). Second, her group this semester is ineffective in part because of clashing goals and bad teamwork. Like Charlie, Hannah had a much better relationship with her group last semester when the students were allowed to form their own groups. She implies that this semester one of her group partners is rushing the pace so that the group is going too fast for her to understand what is going on. Third, the group has had trouble getting timely help from a facilitator in class. They have had many difficulties with the activity guide materials and the lab equipment this semester. They spent a lot of time floundering, waiting for the help they needed. Often, the group would get *hung up on something, like, really small that doesn’t even matter but it takes a lot of time.* This issue had become a problem only in the weeks preceding the interview. This was one source of her frustration because she felt that at a private school like NWU she should have better access to the instructor during class. Because of the latter two difficulties, she goes through the activities but she still doesn’t understand the concepts. This is another source of frustration for her because she cannot figure things out for herself as she claims she was able to do earlier in the class. In describing her biggest difficulty with the course, she said,
I think just staying focused on the material, probably. 'Cause I mean it's hard — I often — like, rarely do I go home and study it after class, you know? Because I just — I just get so frustrated with it. It just is so... I don't know. It's just, like, really hard to stay focused in the class. There's just — So much is going on...there needs to be more organization or something. And I should probably — I should. I know I should go home and, like, study it on my own, and stuff like that. But I don't. And that's — that's probably my greatest — like, the reason why I'm not doing as well as I could be. ... I don't study on my own enough.

Another issue was how she felt when the class challenged her common sense beliefs on how things work. In her words,

*Just seems like everything is backwards from what I thought. Most stuff. [Is that discouraging?] Yeah, in some ways. I don't know. It depends, like, how it's presented. You know, if it's presented in a way that I just, like, feel real stupid, then it's discouraging. But if it's, like — If I figure it out myself, it's not so discouraging at all. [What happens when it's in class?] Um...it depends on the day. ... Like I said before, like, at the beginning of the semester, I was, like, beginning to figure stuff out for myself, you know. But now, it just seems like mostly [the instructor] gets disgusted and ends up writing it on the chapter [sic] and so, like, I know he doesn't like to do that, and so, I mean in that way, it's — It's just like I don't even have a chance to open my eyes... . I don't know. It's just hard.*

It was this combination of factors that led to the bitterness mentioned previously.

In addition to learning how to solve physics problems she also learned,

... A lot of interpersonal interaction skills and, like, how to ask questions, and how to work in a group ... [What would you say was the main skill you learned, if any?] ... (laughter) ... The main skill ... I would say to understand physics more than just to solve physics problems. Just to understand it in a deeper way.

But a little later when she was asked if the main skill she got out of the course was to learn how to reason logically, she replied,

*I said, "agree," ... I get other skills as well. But this would probably be the main skill that I get out of it ... I hope it's what I get out of the course. I don't know. I mean that's the main skill I want to get out of...*
the course, but I don't know if it's the main skill that I have gotten out of the course.

And when asked if learning physics with this method would give her an advantage over other pre-med students, she said,

*I don't know. ... I mean I can see some advantages to it, but also I think in other physics classes they have, like, just one lab and stuff like that. So, they still get hands-on experience; but maybe they're learning the concepts, like, in a more solid way. ... You know, like more concretely presented, instead of, like ... instead of just kind of, like, figuring it out. You know, they're just given, like, “Learn this”... and so they may really, really learn it — instead of just kind of like, "Play around with these cars and see what happens when they, like, crash." And then in three days, we'll, like, figure out what it means.

Hannah had the background to do well in this type of class. But her expectations, her study habits, and problems with the class implementation all contributed to her getting less out of this class than she might have.

**C. Drury College: Post Sequence**

At Drury College, I observed both the algebra/trig-based and the calculus-based Workshop Physics classes and interviewed twelve student volunteers, four of them from the calculus-based sequence. The calculus-based class was very small, only 7 students, but it is worth discussing this class briefly because this sequence had the second best improvement in FCI fractional gain and the most favorable expectations of all the classes in the study.

The interviews confirmed that the students have favorable expectations, but it is interesting to note some of the differences between the WP classes at NWU and Drury. The contrast was quite noticeable because I went to NWU immediately after my site visit to Drury. The Drury students were more constructivist and less mixed in their
expectations than the NWU students. Like the NWU students, the Drury students also
had very favorable responses to the reality link questions, but they were better at seeing
how understanding physics would be useful to them in their life and in their careers.
Also, the Drury students seemed more satisfied with both their groups and the Workshop
Physics curriculum. I will come back to this point in the next section.

IV. DISCUSSION

A. MPEX Survey

In this chapter we have discussed the use of the MPEX survey to measure student
cognitive attitudes in physics. We gave the survey in classes at ten schools that had
varying entrance selectivity and that used either traditional instruction or one of three
research-based teaching methods. We find explicit answers to the research questions we
posed in the introduction.

Q1. How does the initial state of students in university physics differ from the
views of experts?

At the ten schools tested, the initial state of students deviated significantly from that
of the expert calibration group with overall responses ranging from 50-60% favorable.
The results on the reality link cluster (60-80%) and the effort cluster (65-90%) were
much closer to the expert response than the overall results and the results from the other
clusters.

Q2. To what extent does the initial state of a class vary from institution to
institution?

The student attitudes at UMD, PGCC, MSU, and OSU as measured by the survey
were very similar. The attitudes of beginning students at MIN and the liberal arts
institutions (DC, CAR, DRY, NWU, & SKD) were more favorable in two to five of the measured dimensions. The beginning students at MIN, NWU, and DRY also had more favorable attitudes on the overall survey.

**Q3. How are the expectations of a class changed as the result of one year of instruction in various learning environments?**

At every school but one in this study (DC was the exception), the overall results deteriorated as the result of after instruction. Many of the schools showed significant deterioration on two important clusters: half on the math link (with the others showing some deterioration) and half on the reality link (with all but one of the others showing some deterioration. Also, the expectations of students taught with the GPS curricula were not better and in some cases were worse than the traditional classes at CAR and MIN.

**B. MPEX Interviews**

Interpreting the analysis of the MPEX interviews produced some interesting results both in terms of the students’ expectations and in terms of our understanding of the implementation of the Workshop Physics curriculum. In this section, interview responses from all nine NWU students are discussed. Leb, Krystal, and Ramsey were in the middle third of the class.

**1. Expectation issues**

In general, unlike the students in Hammer’s dissertation study who were either constructivist or transmissionist (see chapter 2), the NWU students I interviewed had mixed expectations. Some of the students like Amy and John were not prevented from succeeding in the class by their transmissionist expectations. But their expectations did
seem to prevent them from getting the full benefit of the Workshop method. In addition to the individual student expectations described above, there were also some general findings that I would like to point out.

As we saw earlier in the chapter, most classes (including the Workshop Physics classes at Dickinson College) show a decrease in the number of favorable responses in the reality link cluster over the year of introductory physics. At NWU, the students began and ended the sequence with more favorable responses in the reality cluster than most of the schools in this study, even Dickinson College (see the Redish Plot in Figure 10-6). The NWU students went from 81% favorable to 77% favorable on the reality link cluster over one year. For comparison, the DC students went from 72% to 68% favorable (combined std. error = 5%). Although they both change by the same amount, the NWU students start and end more favorable than the Dickinson students.

This extremely favorable response from the NWU students is reflected in the interviews. Almost all the students I interviewed at NWU believed that the physics they learned in the classroom was strongly connected to the real world. In fact, all the students interviewed in the upper two-thirds of the class gave explicit examples of how they were connecting physics to their own out-of-class experiences. Part of the reason for the favorable response on the reality link cluster is due to the way many of the students look at learning and part is undoubtedly due to Workshop Physic’s heavy emphasis on laboratory activities. But part of the favorable response is due to the instructor connecting what the students are learning to everyday experience. In Kim’s words,
I think every problem we have in class, every example he’s [the instructor’s] used he’s related to a real world concept. It’s not anything that we’ve never heard of before. … We use real life examples everyday.

There are two items on the survey, items 17 and 33, that deal directly with students’ perception of physics and the class. They read as follows:

17. Only very few specially qualified people are capable of really understanding physics.

33. It is possible to pass this course (get a “C” or better) without understanding physics very well.

Both of these items can reveal a lot about students views on the class itself. The NWU students’ responses are shown in Table 10-7. Item 17 often reveals whether or not students believe they could have learned and understood physics in this course. This question was asked in eight of the ten interviews including John, Amy, Hanna, Kim, and Hanna. Everyone except John and Krystal felt that anyone is capable of learning physics; although, some of them did add conditions such as if you apply yourself (2 students), or if you really want to (2 students). John, on the other hand, responded neutral to the question because he felt that there are some people who just can’t learn to really understand physics no matter how hard they try. Krystal responded that many people can really understand physics, but you have to have some intelligence. This result indicates that even though some of these students are struggling, they believe that
Table 10-7a: Responses for MPEX survey item 17 from all NWU students interviewed.

17. Only very few specially qualified people are capable of really understanding physics.

Favorable

Amy: I strongly disagree with that. I think — I think if you take the time and you concentrate on it, you can learn it. So, I — I never — When I first started taking it that first semester, I never thought I was going to learn it. First of all, "What is this stuff?" — you know. But I — That's why I spent a lot of time on it.

Kim: I put I disagree. And I think it depends on the level of physics. Some — I'm taking a course. I understand some things, so I can say I'm capable of understanding. But some of the more technical or mathematical things that I'm not — that I don't grasp, some people do have an understanding of it. But I think, on a college level or a high school level, the materials that they — they teach are — people are capable of grasping some of those concepts.

Hannah: I said disagree. I think a lot of people can understand physics if they want to try and put the time into it.

Leb: I said "strongly ... disagree." I think if you apply yourself, you can do just about anything you want. [That works for you, or anyone?] It works for anyone.

Ramsey: I said I disagreed. "Contrary to popular folk lore." [Can you elaborate a little more on that?] Yeah. Now, I'm almost rethinking that because of the number of physics majors here. I think that there's very few. I think that most people in our course — I would say that most people understand it. But I mean to what degree. "Really understand" it — I don't know what that means. I guess it... I think most of the people in our course — and I know I do — I get a lot of course understanding about the physical world around us. I don't know if I really, really understand it like I could teach it. That's what I would assume that means. "Really understand." Then I would say maybe "only a few specially qualified people would." But I don't know. Is that what you mean by 'really understanding' it? Like being able to tell somebody else? Or, is this my interpretation? [It's your interpretation.] I'm going to — But it's not like an elite club of people who understand physics. I'll stick with my answer. But I don't...it sounds so limiting ... Let me think. I don't know how I can elaborate, sticking to my answer. But I want to stick with my answer. "Only a few, specially qualified people are capable of really understanding physics ... " I think, for the most part, most of the students in this course understand it to a degree above average. I'll say that. [All right. And for that reason, you disagree.] For that reason, I disagree.
Neutral

John: Now, in retrospect to my previous answer, I'd still put "neutral," because I feel that the term "very few," to me, makes it sound kind of elitist in that question. Maybe I'm reading too much into it. But I think there are many people who can understand it. There's a matter of will involved. And then, indeed, there are people who can't; and that's why I put three. No matter how hard they try, they're just not going to get it. And there're some things that no matter how hard I try, I don't get, either. Like art. (laughter) I'm just not into art. And ... But I think that there's certainly a desire factor.

Krystal: (UNINTELLIGIBLE) neutral. I wouldn't understand it, but I think — You know, I don't completely understand it. But I think that a lot of people can. You don't have to be special to understand physics. I think you have to have some intelligence, though. That's for sure.

Roger: I was neutral on that. And the reason why I put "neutral" is pretty much ... If there's a will, there's a way. If you put your mind to do it, you can do anything you want it to do. And it's ... Anybody can learn it ... I think. It's just that it's what your research is about. It's how it's taught to the individual. If it's laid out in a very simple, straightforward manner, where they're actually looking for knowledge and the cognitive development of the student, versus getting through a textbook, the student can learn it.

Unfavorable — No unfavorable responses
Table 10-7b: Responses from all NWU interviews for MPEX Survey item 33

33. It is possible to pass this course (get a “C” or better) without understanding physics very well.

Favorable

Charlie: *And I disagreed basically because I think, since we get use our activity guide and you’ve got all the information there that you, you got to understand it and not just because — you don’t need to memorize it because you’ve got it all there. So, to do well, you’ve got to understand it.*

Leb: *And I said, ”disagree.” You can get a ’C’ or better without really understanding all the physics we’ve been through, but I think you definitely have to learn a substantial amount if you don’t know any physics coming into it, to do reasonably well in this class.*

Krystal: *I disagreed. I mean you definitely have to have a little bit of an idea to get better than a ”C”.*

Neutral — No Neutral Responses

Unfavorable

John: *I put ”agree.” I guess. You can ... be a plug-and-chug. You know what you need to do in a given situation and just deal with it and pass.*

Hannah: *I said disagree, but I don’t know — I might want to change my answer. I think you could pass it without understanding physics very well, if you could just — if you were a person who could just, like, plug and chug and figure out what goes in what formula. I think you could get a ’C’ in the class. [So, what would you change your answer to?] I would change it to agree. [So, you’re saying — Not just any plug and chugger — you have to be a really good plug and chugger.] Yeah, but you wouldn’t have to understand physics, necessarily. [Okay. So a good plug and chugger could go far.] Yeah.

Ramsey: *And I said four for ”agree.” I said, ”It could be done if someone really wanted to memorize the appropriate material and not be able to apply much of what they know.” I would say that someone could get by with a ’C’ or better without understanding, because I would say — I would define ’understanding’ it by being able to tell someone else what you know — relate to someone else what you learned. And those people who can — I’m sure someone could do a ’C’ or better if they memorized what they needed for a test and forgot it in a few hours or something like that. Because I know people do that.*

Roger: *I’d agree. … [Well, why ... do you say that?] … I mean there’s many people I’ve talked to that ... have gotten a ’C’ in the class, and they don’t understand it.*
most people can learn physics with understanding. This result from the interviews is close to the results from the survey. The NWU students started even with the Dickinson students at 88%. While the DC students only dropped to 78% favorable (DC had the largest favorable response to item 17), the NWU students dropped to 65%.

Agreeing with item 33 indicates that understanding physics is not required to succeed in the class and implies that understanding physics is part of the hidden curriculum. This question was asked of seven of the ten students including John, Charlie, and Hanna. Three of them (including Charlie) disagreed with this item. But John, Hannah, and Ramsey believed that despite the heavy course emphasis on conceptual understanding in Workshop Physics, it was still possible to pass this class by memorizing, plugging equations into problems, and chugging away for an answer. In other words, by the same means that students use to pass a traditional lecture course when they don’t understand the course or when they are trying to get the best grade possible with the least amount of effort. This may be an indication that the homework and the exams may not be adequately reinforcing the goal of understanding physics, i.e. that the problems may not be designed to encourage students to explicitly use their understanding of physics to do well.

Only two of the four students in the top third of the class (including Charlie) and three of the ten students overall preferred Workshop physics over traditional lecture instruction. However, six of the other students including Amy and John believed that learning by lecture was or would be more effective for them than figuring things out for themselves. In the case of Amy and John this was due to their misconception on what they were supposed to be learning, although John did believe that Workshop Physics was
a better way to learn concepts. The other four students preferred lectures, either because they wanted things explained to them or because they wanted more recipe-like activities.

Most of these students had at least some constructivist expectations, but these are what we call “apparent” constructivist expectations. Most of these students believe they should be constructing their own understanding, yet deep down they feel that being told is more effective. One reason for this is that working on something when you don’t know what is right or what you are supposed to be learning is uncomfortable for many students. Three of the students including Hannah commented that sometimes they felt they were not learning the right things or they did not know the right answers to their classroom lab activities.

Another reason why students might feel uncomfortable is that Workshop Physics requires a different type of learning from what they are used to in traditional instruction. Two of the students from the middle third of the class, Ramsey and Krystal both commented on this. When asked if the class was what he had expected it to be, Ramsey replied,

_There’s a lot more emphasis on using your hands in lab. I was ready to take notes and study for exams like every other class. So, this has required me to study in a different way. But I don’t really have anything to compare it to. But I don’t — I like learning like this. It's not too bad. It's just a different way to study, different way of learning._

When asked how studying for this class was different, he continued,

_In several of the courses I've taken here, you sat in your lectures for 50 minutes, take notes, review your notes, and basically memorize things for a test. And it might just be — This is the only physics class I've taken, too. It might be that, but that's a different, too ... just having to work all the problems. But the lab manual that we work out of — you_
just keep up by doing exercises every day in lab. You don't really take
notes. And sometimes I don't really learn the right thing, because we
don't review. A lot of time, we don't go over the material — like in
other classes, you have reviews or recitations, I guess, where you can
ask questions and those are kind of skipped over here.

Krystal made a similar comment when asked about what changes she would make to
improve the class,

Oh, I would add a lecture, somehow, and cut back on the amount of lab
time. And I understand what they're trying to get at with the computer
workshop, but I — I just — Again, I think it's important. Maybe if they
could just kind of just wean us into it, because just — boom — all of that
at once, when you're used to that learning style, of lecture, questions,
test; lecture, questions, test; and then three hours of lab, you know —
like most of us has — the way the programs — the classes are set up
here. But I — I mean I wouldn't want to get away from what they're
trying to do, so maybe if they could just take one hour out of our week,
lecture, and get rid of an hour of lab. Does that make sense?

And when asked which format she would prefer is she had the opportunity to do the
class over again, she answered,

Well, I definitely would prefer the traditional, because I'm not one to
accept a lot of change — especially thrown at you that fast. And
because I learn that way. That's just ... I mean I think I would have done
a million times better than how I'm doing right now, had it been not that
format. And I am one of those people — I'm really concerned about
what I get. My grades are really) important, if I want to reach my
career goal

An important result from these interviews is that success in the class,
expectations, and mathematical ability seem to describe three different aspects of student
achievement. Of the four students I interviewed in the top third of the class, two seemed
to be in the binary stage and two had weak math backgrounds.

One of the key aspects of science-education reform and research-based physics
curricula is the establishment of a community of learners where the students discuss the
central ideas of the course with one another both in and out of class. Although some of the students at NWU discuss their solutions with one another and help one another prepare for exams, they do not discuss or reflect on what they do in class when they are out of class. This is in contrast to Dickinson College and Drury College where the workshop physics students regularly discuss the class activities outside of class in the same way the NWU students discuss their homework.

Almost all the students I interviewed at Drury worked with other students outside of class at least once a week. They discussed all aspects of the class, not just the homework assignments and the exams. In addition, when asked about the advantages of taking workshop physics, many of the students answered as Charlie did, that this way of learning physics was helping them learn useful knowledge and skills that they could apply beyond the class. While some of the NWU students were able to see connections between the physics they were learning and their careers, the Drury students saw the connection more clearly and were able to cite specific examples.

In general, the implementation at Drury seemed more successful than the implementation at NWU. The groups functioned better and the instructors seemed to have learned how to manage the class more effectively. However, part of the success at Drury may be due to the nature of the institutions. It should be noted that many of the students in the WP classes at Drury knew each other from other classes and they were encouraged to form study groups for many of their classes.
2. Implementation issues:

Two factors essential to the success of any curriculum that uses cooperative
groups are how well the group members work together and whether the group’s main
goal is completing the activity or trying to really understand what is going on in the
activity. Six of the ten interviewed students including Charlie, Amy, Krystal and Hannah
mentioned problems with their groups.

In Amy’s case, the problem was the way the group worked together. The two
other members of her group knew each other outside of class and excluded her when
they were both present. However, in the weeks before the interview, one of the other
women had been absent and Amy was able to collaborate with her remaining lab partner.
Amy said that lab had been a much better experience with just the two of them.

In the case of Charlie, Krystal, and Hannah, the problem was one of conflicting
goals. Their goal of trying to make sense of their lab activities was thwarted by some of
their group members who stressed getting through the assignment as quickly
(“efficiently”) as possible. These students were like the poor students in the studies by
Chi et al.\textsuperscript{17} and Ferguson-Hessler and de Jong\textsuperscript{18} (see the discussion on the use of
example problems and readings in chapter 2) who looked through the material without
thinking about it and checking to see that they understood it.

In cooperative group activities, the instructor can counter this tendency by
asking semi-Socratic questions of the groups and/or the class as a whole. However,
problems with the group dynamics should not be taken lightly. It is an indication that the
students’ goals for learning do not match those of the instructor or the curriculum. In
addition, there were also indications in the interviews that some students were not
connecting the classroom activities to the homework and the exams. One student noted explicitly in their interview that they used their textbook to do the homework and the class activities to understand the concepts.

Several of the NWU students interviewed commented about wasted time due to the confusion at the beginning of the period. In Charlie’s words,

… so there was a lot of confusion time, just setting things up. It wasn’t explained well. And that’s the big thing, I think. He [The instructor] goes through each group individually and explains how to set something up. But a — rather than doing it as a whole. … I just thought with [the second semester instructor], we seem to waste a lot more time, because you couldn’t really progress with the lab until you had it set up right. There just was a lag time.

The students need to have a good idea of what they are doing when they begin the lab activity. Leaving student groups floundering for long periods of time can result in the students being frustrated. This is where having more than one facilitator for 25 students can make a big difference.

From the interviews, we learn that the group dynamics affect how the students viewed the curriculum. When students feel their group works well, they tend to feel Workshop Physics is effective. If the group has difficulties, students tend to believe they would learn more in traditional instruction.

The primary role of the facilitator in all three of our research-based curricula is to keep the group members exchanging ideas and on task. In all three of these curricula, it is essential for the facilitators not to explain the material to the students, but to try to help the students find their own answers. At the same time, the instructors need to monitor the student groups. Several of the interviewed students, including Charlie, Amy, and Hannah, mentioned problems with the groups. The instructor/facilitator also needs
to be aware of the dynamics in each group so that problems such as Amy described get resolved early in the semester. However, an instructor needs to be careful not to act prematurely; some groups need a little time to pull together. One option with assigned groups\(^{19}\) is to change the groups every 3-4 weeks so that a person in a dysfunctional group is not stuck for very long.

\begin{enumerate}
\item W. F. Perry, *Forms of Intellectual and Ethical Development in the College Years* (Holt, Rinehart, & Wilson, NY, 1970).
\item R. Likert, “A technique of the measurement of attitudes,” *Archives of Psychology*, No. 140.
\item On items 7, 9, and 34, a majority of our expert group’s agreed with a particular position, but one-third to one-fourth of them chose neutral.
\item Matched data includes only the results from students who completed the survey both at the beginning and at the end of the period in question.
\item With one exception, the data presented for Carroll College includes students who took the MPEX survey and the beginning and end of the two semester sequence.
\item Private communication with Tom Foster and Laura McCullough from the Physics Education Group at University of Minnesota (Fall 1997).
\item See Ref. 5-20.
\item This led us to include the phrase “or proof” in item 2.
\end{enumerate}
The interview code-names correctly reflect the gender of the students.

See Refs. 2 & 3.

See the description of cognitive beliefs of adult learners in Chapter 2 and Refs. 2 & 3.

See Ref. 7.


The NWU students were assigned into groups at the beginning of the second semester and kept the same groups for the entire semester.
PART IV. CONCLUSION

Chapter 11. Conclusion

In this dissertation, I have examined the methods used to evaluate instruction in terms of the hidden curriculum and used these methods to study classes at ten schools taught with one of three research-based curricula or traditional lecture instruction. In this chapter, I will summarize what I have learned in this study, discuss the implications for introductory physics instruction and physics education research, and suggest directions for future studies in this area.

I. SUMMARY

A. Why are Research-Based Curricula Necessary?

Physics Education Research (PER) has demonstrated that traditional lecture, recitation, and lab instruction is not helping many students in introductory physics classes develop a functional understanding of physics concepts. In chapter 2 (Mazur and Hammer), chapter 4 (Halloun and Hestenes, Hestenes et al., and Hake), and chapter 6 (example of a students’ solution to a quantitative end-of-chapter problem and a qualitative problem), we have seen that many students have difficulty with basic physics concepts on simple, qualitative questions even when they can successfully solve mathematically complex end-of-chapter problems. The studies by Mazur, Hammer, and Tobias suggest that the emphasis on typical end-of-chapter problems and the structure of the traditional lecture method encourage students to see learning physics as memorizing and applying the facts and equations without understanding the underlying concepts.
To the students in these traditional lecture courses, the main learning goal appears to be to demonstrate mastery of the material by solving typical end-of-chapter problems on exams and homework. However, most physics instructors want their students to achieve much more than that. Among other things, they want a majority of their students to achieve the following:

- to understand the main physics concepts including knowing when and where specific concepts apply;
- to be able to express what they have learned consistently in multiple representations including graphs, equations, and words;
- to see physics knowledge as a connected, coherent framework of ideas where a few key principles can be used to understand many physical situations; and
- to be able to apply what they know to new physical situations in and out of the classroom.

Learning goals like these, which are often neither stated explicitly to the students in class nor encouraged through grading and testing, are part of what we call the “hidden curriculum.”

In the last twenty years, PER has made significant progress in identifying and understanding student difficulties with introductory physics. One of the main findings is that students come to the introductory physics sequence with beliefs and attitudes based on years of experience with school and the world around them. In particular, they have their own ideas about how to solve physics problems, common sense beliefs about how things work, and cognitive beliefs on learning, physics, and mathematics. Many of these views are incompatible with what instructors want the students to learn, hinder the students’ learning, and outlast traditional lecture instruction. PER has also demonstrated that if students’ initial views are taken into account, it is possible to design active-learning activities that induce most of the students to develop a good better
understanding of many of the basic concepts or develop a more-expert problem-solving approach.

This last result has led to the development of new curricula to improve student learning by focusing on what is happening in the student rather than on what the teacher is doing. Most of these PER-based curricula use a strategy similar to Posner et al.’s four conditions for conceptual change (discussed briefly in chapter 2)\textsuperscript{10}, including a component where the students are actively involved in debating and discussing the course material in peer groups of 2-4 students. The three PER-based curricula examined in this study, Tutorials,\textsuperscript{11} Group Problem Solving,\textsuperscript{12} and Workshop Physics,\textsuperscript{13} all make use of cooperative student group activities. Tutorials and Workshop Physics are designed to improve students’ conceptual understanding of physics while Group Problem Solving emphasizes developing expert problem solving skills.

The developers of these three curricula have each presented evidence that their methods significantly improve student performance on problems and/or multiple choice tests in the areas the curricula were designed to address. In addition, students taught and engaged with one of these three curricula have been shown to do as well, if not better on conventional end-of-chapter problems as students taught with traditional lecture. However, to be effective, each of these PER-based curricula require additional resources and major changes in teaching style. Because of the effort and cost involved in implementing one of the research-based curricula as well as the difficulties of adopting a curriculum developed elsewhere, it is important to learn how to evaluate these PER-based curricula to see what students are learning, particularly with regard to the hidden curriculum.
B. How Do We Evaluate Research-Based Curricula?

In this study, I evaluated two aspects of the hidden curriculum, conceptual understanding and expectations, by collecting four types of assessment data to evaluate four curricula used in calculus-based introductory physics courses. The types were multiple-choice concept tests, the Maryland Physics Expectation (MPEX) survey, student interviews, and specially designed qualitative exam problems. Concept test results and surveys were collected from at least one class at each of the ten schools. Student interviews were conducted at five of the ten schools and qualitative exam problem results were available from classes at the University of Maryland.

From a research standpoint, the four methods are most useful for learning about what happens when students go astray and don’t learn what was intended. By studying what these students have learned from the class and why, we can begin to understand the nature of the students’ difficulties in learning physics and how to modify instruction to address these difficulties. Each of the four methods tells us different things about what students in introductory classes learn during the sequence.

In chapter 4, we saw that concept tests like the Force Concept Inventory (FCI) give an indication of how well students know basic concepts. In the case of the FCI, Halloun and Hestenes demonstrated this by comparing FCI results with results from interviews and open-ended questions from individual students.\(^\text{14}\) Hake demonstrated that the fractional gain between the pre- and post-course FCI results is a more consistent measure of how much students’ conceptual understanding has improved than the absolute gain or raw post score.\(^\text{15}\) However, Steinberg and Sabella found that a large minority of students did not respond consistently to similar FCI and open-ended
This suggests that while the FCI is a good indicator of whether students know Newtonian force concepts, it may not be a good indicator of how students use their conceptual understanding in problem solving.

An analysis of exam problem solutions can be a good indicator of students’ ability to apply what they know. A variety of carefully constructed problems like those in chapter 6 can reveal much about what students have learned to apply. However, as we saw in the case of the waves-math pretest in chapter 7, interviews may also be needed to understand what the students are thinking and why they answer a particular way. Interviews are the most effective way to determine how students are thinking about physics.

As we saw in chapter 2 (Hammer and Tobias) and chapter 10, interviews are also a good way to learn about student expectations. While interviews are the most effective research tool for determining how students think about physics and what they are learning, the time required to transcribe and analyze interviews makes them impractical for evaluating classwide effects in all but the smallest classes. Since one of the goals of this investigation is to study the distribution and evolution of student expectations in introductory classes, we developed and used the MPEX survey, as discussed in chapter 5, refined through the use of interviews.

The MPEX survey was constructed to probe student expectations with a focus on six structures: independence, coherence, concepts, the link between physics and the real world, understanding of the role of math in physics, and the kind of effort students expect to make. The survey was calibrated using five groups. The calibration group expected to be most sophisticated was in strong agreement (better than ~80% on almost
all the items) as to the desired responses on the items of the survey. Their preferred response was defined as favorable. The other calibration groups showed decreasing agreement with the expert group as predicted. Over 100 hours of interviews were conducted with student volunteers to validate the students’ interpretations of and responses to the 34 survey items. In addition, reliability of the survey was established by a Cronbach alpha of 0.81 and by demonstrating that the overall survey results and the results for the six dimensions were reproducible for several similar main sequence classes. The calculated standard deviations of each of the survey results for similar classes were comparable or less that the estimated distribution widths.

While the survey can provide a measure of student expectations are and how they change, interviews are also needed to see why they change and to better understand the process of change.

C. Evaluation of PER-Based Curricula

In this dissertation, the evaluation procedures described in the previous section were used to compare three PER-based curricula (Tutorials, Group Problem Solving, and Workshop Physics) to traditional instruction. We find explicit answers to the three research questions posed in chapter 1.
Q1. What are the characteristics of different student populations coming into the calculus-based introductory physics class?

a. Conceptual understanding of physics

To measure conceptual understanding, FCI or Force and Motion Conceptual Evaluation (FMCE)\(^{17}\) data were obtained from first-term classes at all ten schools. FCI data were collected from eight of the ten schools. The average overall pre-course FCI results at the three large state universities (UMD, MIN, and OSU) and one of the small liberal-arts colleges (DRY) were all very similar at approximately 50%. This was significantly larger than the average overall pre-course FCI score at the other three liberal arts schools (DC, NWU, and SKD) and the community college, where the average score was about 40%. The pre-course FMCE results show a similar gap. The students at MSU start with an average FMCE of 37.5% while the students at DC and CAR both start with an average of about 25%. For the FCI results it is worth noting that the average pre-course FCI scores at all ten schools are well below the Halloun and Hestenes’ 60% entry threshold for thinking about motion in terms of Newton’s laws.\(^{18}\) Halloun and Hestenes suggest that students who have not reached this threshold are not yet ready to solve physics problems with an understanding of the underlying concepts.

Of the 8 schools that supplied FCI data, 6 of them (UMD, MIN, OSU, DC, NWU, and SKD) provided itemized data, allowing the Newton’s third law FCI cluster to be evaluated. The students at UMD had the highest initial score (40%) on the Newton 3 cluster and the students at SKD had the lowest (26%). The classes at OSU, MIN, and DC averaged about 33% on the Newton 3 cluster while classes at NWU initially averaged 29%.
The overall concept test results indicate that students in introductory physics classes at the large public universities (UMD, MIN, OSU, and MSU) start off with a small but significant advantage in their understanding of Newtonian force over their counterparts at the smaller liberal arts colleges (with the exception of DRY). The differences on both overall concept test averages and on the Newton 3 cluster do not seem to correlate with the overall selectivity of the school.\textsuperscript{19}

\textbf{b. Student expectations}

MPEX survey data were also collected from classes at the ten schools. The initial state of students at the ten schools deviated significantly from that of the expert calibration group with overall responses of the students ranging from 50-65\% favorable in compared to 87\% for the expert group. The expert responses on the clusters varied from 85-93\% favorable. The student responses deviated most strongly on the independence, coherence, and concepts clusters. These varied from 40-60\% favorable with the community college near the bottom for all three clusters. The most favorable student response was on the reality cluster where responses ranged from 60-80\% favorable.

Many of the student populations showed some differences when compared with Maryland students. Starting with our three large public universities (UMD, OSU, & MIN), the Minnesota students had significantly more favorable expectations overall and in three of the clusters than the Maryland students. The Ohio State students had similar expectations to the Maryland students overall and in all the clusters except the reality link cluster where the OSU students were slightly but significantly better. Except for MSU, which is really a medium sized public university, the students at the small liberal
arts colleges all responded more favorably initially than the Maryland students on two to five expectation dimensions of the survey. Note that none of the student populations at any of the nine other schools consistently responded more favorably than the Maryland students for all six clusters. The NWU students responded more favorably than the Maryland students overall and for the reality and effort clusters. The DC students responded more favorably than the Maryland students to the independence, reality, and effort clusters. The students at CAR and SKD both responded more favorably than the UMD students to the reality and effort clusters. It is worth noting that the students at all five of the liberal arts schools and MIN responded more favorably to those two clusters than the University of Maryland students.

Q2. How do we determine if students are improving their knowledge of physics concepts and expectations?

Q3. Are the research-based curricula more effective for helping students to improve their conceptual understanding and their expectations of physics?

a. Conceptual understanding of physics

Multiple-choice concept tests, specially designed open-ended exam problems, and interviews were used to evaluate students’ conceptual understanding of physics. The change in the mechanics concept test scores from the beginning to the end of the first term of introductory physics are used to determine if students are improving in their knowledge of physics concepts. In this study, we used Hake’s fraction of the possible gain $h$ to measure the change. The larger the fractional gain, the greater the improvement in understanding of the basic concepts.
Classes that used one of the three research-based curricula (RBC) had significantly better overall fractional gains on the multiple-choice concept tests on Newtonian Force than the traditional lecture classes at UMD, CAR, and PGCC. This was even true for classes at schools that were in their first or second year of implementation of research-based curricula. In fact, all the classes using RBC had overall fractional gains on the FCI as good or better than the best Traditional lecture class in this study. The best results came from classes using Workshop Physics (DC & DRY) or Group Problem Solving (MIN). At Maryland, two instructors taught classes both with Tutorials and without in different semesters. In both cases, the classes taught with tutorials had significantly better overall fractional gains on the FCI.

Similar results were obtained with the Newton’s third law cluster on the FCI. The classes taught using one of the three RBC had significantly better average fractional gains on the Newton 3 FCI cluster than the classes taught with traditional lecture instruction. With two exceptions, the average Newton 3 FCI fractional gain for every RBC class was larger than the gains for the traditional classes. One exception was a small (≈ 40 students) traditional class taught by an award winning instructor at Maryland that achieved a larger fractional gain than any of the GPS classes and one of the Tutorial classes. This class also had the best overall fractional gain on the FCI of any traditional class. The other exception was a Workshop Physics class at Dickinson that had a fractional gain typical of traditional lecture classes. This class had unusually severe attendance problems. The Tutorial and Workshop Physics’ classes tended to have higher fractional gains than the Group Problem Solving classes. This may be due to the
fact that the Group Problem Solving curriculum emphasizes problem solving, while Tutorials and Workshop Physics emphasize conceptual understanding.

As we discussed earlier, concept tests are an indication of how well the students know the concepts, but not necessarily how well they can apply the concepts in problem solving. To address this issue, four specially designed problems were given on exams in Traditional and Tutorial classes at Maryland. The students are said to have improved in their ability to use their conceptual understanding if a significantly greater fraction of the students in the Tutorial class use the concept correctly in the context of a problem.

Since the total and fraction of time spent on active-engagement activities in Tutorials (one hour per week) is the least of the three RBC, we would expect these results to be suggestive of the results that would be obtained from classes at other schools.

The results from the exam problems at Maryland were consistent with the concept test results. Tutorial students displayed a better understanding of velocity graphs, Newton’s third law, graphs and equations describing the motion of harmonic oscillators, and 2 slit interference. The ratio of correct responses to the Newton’s third law problem from the Tutorial class and the Traditional class were very similar to the ratio of correct responses to the Newton 3 FCI cluster, although the number of correct responses on the exam problem was significantly less.

Hammer’s two-rock problem\textsuperscript{22} was used in interviews with student volunteers at Maryland, Dickinson, and Ohio State. Here the issue was not so much to see how students had improved, but how well they were able to use their physics knowledge. Very few students were able to solve the problem in the interview, but this was not surprising because of the nature of the problem. What was surprising was the inability of
many of the students to derive the kinematic equations or to recognize the condition that acceleration is constant. In addition, many students became stuck when the kinematic equations produced an expression that was hard to evaluate. In contrast, the few students who did solve the problem were able to come up with alternative approaches that indicated a better, more flexible understanding of physics concepts.

The concept tests and the exam problems clearly indicate that RBC are more effective for helping students learn and apply physics concepts than traditional instruction. However, all three types of evaluation indicate there is still need for further improvement. Only a few of the classes participating in this study achieved a fractional gain of at least 50% and none of the classes achieved an overall average score of 85% on the FCI, the threshold suggested by Hestenes and Halloun for confirmed Newtonian thinkers. Moreover, the results from the exam problems and the interviews suggest that while more of the RBC students than the traditional students are able to apply concepts correctly, many of the RBC students were still not able to use the concepts correctly.

b. Student expectations

The students participating in this study took the MPEX survey at the beginning (pre) of the introductory physics sequence, at the end of the first semester or quarter (mid), and again at the end of the first year (post). The mid and post responses are compared with the pre responses. If the students in a particular group gave significantly more favorable responses on the overall survey or in one of the six clusters, we say that the students have improved in those expectations. A change $\Delta$ is calculated by subtracting the percentage of pre favorable responses from the percentage of post...
favorable responses. The change is considered to be statistically significant if $\Delta > 2\sigma$
where our estimation of the uncertainty $\sigma$ is discussed in chapter 5. However, if
students’ expectations deteriorate significantly as a result of traditional instruction while
the expectations of students taught with one of the RBC do not change significantly,
then we can say the RBC has improved the students’ expectations relatively.

At every school we studied, the overall MPEX survey results deteriorated as a
result of instruction, although only five of the classes decreased significantly (CAR-TRD,
MSU-WP, MIN-GPS, MIN-TRD, & OSU-GPS). Note that this group includes all the
traditional classes for which data was available. A major part of this deterioration was
the significant decrease in favorable responses (deterioration) to the effort cluster at
every school tested. In their judgments at the end of a semester or the end of the year,
students felt that they did not put in as much effort as they had expected to put in at the
beginning of the sequence. This part of the result is well known and neither surprising
nor particularly disturbing. What is more troublesome is the result that half of the
schools showed significant deterioration on the math link and the reality link dimensions.
There was no significant increase in any of cognitive dimensions after one year of
instruction except for the concepts cluster where UMD Tutorials and DC Workshop
Physics both improved significantly.

It is interesting to note that the survey responses of both GPS classes deteriorate
significantly overall and in three of the cognitive clusters including coherence. The
students who had traditional instruction at Minnesota deteriorated significantly overall
and in all clusters except the coherence or concepts clusters but did better than the
Minnesota GPS students who deteriorated both overall and in every cluster.
These results suggest that students who have had a year of Workshop Physics and Tutorials have maintained better expectations than students taught with Group Problem Solving or Traditional lecture instruction. Only two schools showed any improvement in expectations, both in the concepts cluster.

Interviews with students at some of the schools were used to try to understand the interaction between the curriculum, student learning, and student expectations as well as validating the student interpretations of the MPEX survey items. Two of the four students interviewed from the top third of the Workshop Physics class at NWU had expectations that prevented them from developing a flexible, functional understanding of physics but did not prevent them from succeeding in the class. In addition, some of the interviewed students had favorable expectations but ran into difficulties with other members of their group who did not. The students with unfavorable expectations had different learning objectives in class, namely to get through the material, not to understand it and think about it. Since the groups work together as a team, this caused some of the interviewed students with favorable expectations to rush through the activity as well. The interviews also suggest that expectations, mathematical ability, and success in the course were independent in the NWU class.

II. IMPLICATIONS

In this dissertation, we compare three RBC with traditional instruction to determine how well these curricula improve student learning in terms of conceptual understanding and expectations. Compared with the three RBC, traditional instruction is clearly not working for many of the students. The students taught with traditional instruction had lower fractional gains on the FCI and FMCE and did not do as well as
the RBC students on the qualitative exam problems. In addition, the MPEX results show that after a year of instruction the fraction of favorable responses on the reality link, math link, and effort clusters as well as the overall survey result decreased significantly. The results for the remaining clusters also showed some deterioration as well. This confirms the need for change discussed earlier in the chapter.

First, let us consider the effect of the three RBC on conceptual understanding. The average fractional gain on the overall FCI and FMCE was significantly better for the three RBC than the traditional classes at CAR, MIN, and UMD. In addition, each class’ average fractional gain was as good or better than the best fractional gain for a traditional class. These results are consistent with the earlier findings of Hake. These results strongly suggest that the three RBC were more effective than traditional instruction for teaching the students the concepts of Newtonian force and motion. On the Newton’s third law cluster on the FCI, the average fractional gain was higher for the students taught with any of the three RBC than for those taught with traditional instruction. Although the GPS sequences had overall fractional gains on the FCI as good as the best classes using one of the other RBC, the Tutorial and Workshop Physics classes had higher fractional gains on the Newton 3 cluster. This implies that the Tutorials and Workshop physics may be more effective for teaching difficult concepts.

The evaluation of the Maryland students with both concept tests and specially designed exam problems showed that students taught with Tutorials had not only improved their knowledge of physics concepts, but also their ability to apply their conceptual understanding in problems. The Tutorial students demonstrated a better functional understanding of several concepts than the Traditional students. The scores
on the Newton’s third law problem were reflective of the differences on the Newton 3 cluster on the FCI. The students performed significantly better on the concept test than on the comparable problem. This suggests that while the concept tests indicate how well students know the concepts, they are not necessarily an indication of how well students can use what they know, an important component of functional understanding and the hidden curriculum.

The expectation results were less encouraging. Only 40 to 60% of the student responses at the ten schools were favorable on each of the three cognitive dimensions described in Hammer’s study. This suggests that many of the students enter the introductory physics course as “binary” learners who believe that knowledge is composed of facts (in this case, equations) that are transferred from the authorities (the instructor and the textbook) to the students. This in turn implies that the type B attitudes and beliefs that Hammer observed in his small sample are prevalent in a large fraction of students in calculus-based introductory courses at community colleges, liberal arts colleges, and large state universities.

With regard to the effects of the RBC on expectations as measured by the MPEX results, only the Tutorials and one of the Workshop Physics sequences showed any significant improvement, both in the concepts cluster. However, the Tutorial classes and most of the Workshop Physics classes did better than the other classes by not deteriorating as much. The MPEX results for the GPS classes were not as good as the TUT and WP classes. The GPS classes had MPEX results that were no better and in some cases worse than traditional instruction.
While this analysis of the MPEX results is complete, there is still more to learn from the data we collected in this study; for example, how the MPEX survey results correlate with other evaluations such as grades, FCI performance, or majors? Laws has commented that junior and senior biology majors seem the most resistant to RBC in physics classes. There are preliminary results from a Flemish study using the MPEX survey that biology majors have substantially different expectations than other majors. Also, the results presented in this dissertation are only part of what has been collected to date. We have also started to collect data from traditional and innovative algebra-based courses.

This study has several implications for developers and adapters of RBC. The best results (fractional gains) on the FCI and FMCE were achieved by classes taught by people involved in developing the curriculum (Redish-UMD, Heller-MIN, Laws and Pfister-DC) or people who had effectively incorporated the active-learning activities with cooperative student groups (Riley-DRY). Part of this may be due to the skill of the instructor, but I believe this result suggests students learn more effectively when the active learning activities are well-integrated into the course and the student groups are functioning well discussing their understanding of the material. However, even the FCI gains of the best classes in this study and Hake’s show there is ample room for improvement of PER-based teaching methods in helping students learn the concepts of physics.

Our understanding of expectations is considerably less developed than our understanding of students’ difficulties with conceptual understanding. Therefore, it is not surprising that efforts to address expectation issues are also less developed than the
PER-based curricula to address conceptual understanding and problem solving. The MPEX results of the RBC classes and particularly the GPS classes indicate that even while students’ conceptual understanding is improving, their expectations about what physics is and how to learn it are not. The MPEX interview at NWU showed that even when students’ expectations did not prevent them from succeeding in the class, it did affect their perception of what they were learning and what they took away from the course. More research is needed on the role of student expectations in learning physics, particularly on the interaction of student expectations and the curriculum, to help curriculum developers and adapters address this issue of the hidden curriculum more effectively.

This study represents a first step is exploring the issue of assessment of the hidden curriculum and in expanding our understanding of what is really going on in our classrooms.

6 See Ref. 1.
7 See Ref. 2.


15 See Ref. 5.


18 See Ref. 14.

Although, the seven DRY students came close. The DRY students responded more favorably than UMD students both overall and for every cluster except coherence.

See Ref. 5


See Ref. 14.

The third was not measurable since no data was collected from consecutive traditional classes at Maryland.

See Ref. 5.

See Ref. 21.


Private communication with E. Van Zele at Ghent University, Belgium (Sept. 1997).
Appendix C. Maryland Physics Expectation (MPEX) Survey

This section contains a list of the 34 MPEX Survey items and a copy of the scantron pre-course version of the survey form (v. 4.0). Note that the survey items on the scantron version are numbered 60-94 for use with scantron form NCS 4887.
The MPEX Survey items:

Note that individual items from this survey should not be used to evaluate individual students. On any single item, students may have atypical interpretations or special circumstances which make the “non-expert” answer the best answer for that student. Furthermore, students often think that they function in one fashion and actually behave differently. A more detailed observation is required to diagnose the difficulties of individual students. This survey is primarily intended to evaluate the impact of one or more semesters of instruction on an overall class. It can be used to illuminate some of the student reactions to instruction of a class that are not observable using traditional evaluations. In this context, it, together with evaluations of student learning of content, can be used as a guide for improving instruction. The numbers 1 to 5 represent a five point Likert scale where 1 = “strongly disagree” and 5 = “strongly agree.”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All I need to do to understand most of the basic ideas in this course is just read the text, work most of the problems, and/or pay close attention in class.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>2</td>
<td>All I learn from a derivation or proof of a formula is that the formula obtained is valid and that it is OK to use it in problems.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>3</td>
<td>I go over my class notes carefully to prepare for tests in this course.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>4</td>
<td>Problem solving in physics basically means matching problems with facts or equations and then substituting values to get a number.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>5</td>
<td>Learning physics made me change some of my ideas about how the physical world works.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>6</td>
<td>I spend a lot of time figuring out and understanding at least some of the derivations or proofs given either in class or in the text.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>7</td>
<td>I read the text in detail and work through many of the examples given there.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>8</td>
<td>In this course, I do not expect to understand equations in an intuitive sense; they just have to be taken as givens.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>9</td>
<td>The best way for me to learn physics is by solving many problems rather than by carefully analyzing a few in detail.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>10</td>
<td>Physical laws have little relation to what I experience in the real world.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>11</td>
<td>A good understanding of physics is necessary for me to achieve my career goals. A good grade in this course is not enough.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>12</td>
<td>Knowledge in physics consists of many pieces of information each of which applies primarily to a specific situation.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>13</td>
<td>My grade in this course is primarily determined by how familiar I am with the material. Insight or creativity has little to do with it.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>14</td>
<td>Learning physics is a matter of acquiring knowledge that is specifically located in the laws, principles, and equations given in class and/or in the textbook.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>15</td>
<td>In doing a physics problem, if my calculation gives a result that differs significantly from what I expect, I'd have to trust the calculation.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>16</td>
<td>The derivations or proofs of equations in class or in the text have little to do with solving problems or with the skills I need to succeed in this course.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>17</td>
<td>Only very few specially qualified people are capable of really understanding physics.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>18</td>
<td>To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>19</td>
<td>The most crucial thing in solving a physics problem is finding the right equation to use.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>20</td>
<td>If I don't remember a particular equation needed for a problem in an exam there's nothing much I can do (legally!) to come up with it.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>21</td>
<td>If I came up with two different approaches to a problem and they gave different answers, I would not worry about it; I would just choose the answer that seemed most reasonable. (Assume the answer is not in the back of the book.)</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>22</td>
<td>Physics is related to the real world and it sometimes helps to think about the connection, but it is rarely essential for what I have to do in this course.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>23</td>
<td>The main skill I get out of this course is learning how to solve physics problems.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>24</td>
<td>The results of an exam don't give me any useful guidance to improve my understanding of the course material. All the learning associated with an exam is in the studying I do before it takes place.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>25</td>
<td>Learning physics helps me understand situations in my everyday life.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>26</td>
<td>When I solve most exam or homework problems, I explicitly think about the concepts that underlie the problem.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>27</td>
<td>Understanding physics basically means being able to recall something you've read or been shown.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>28</td>
<td>Spending a lot of time (half an hour or more) working on a problem is a waste of time. If I don't make progress quickly, I'd be better off asking someone who knows more than I do.</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>A significant problem in this course is being able to memorize all the information I need to know.</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>The main skill I get out of this course is to learn how to reason logically about the physical world.</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>I use the mistakes I make on homework and on exam problems as clues to what I need to do to understand the material better.</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>To be able to use an equation in a problem (particularly in a problem that I haven’t seen before), I need to know more than what each term in the equation represents.</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>It is possible to pass this course (get a &quot;C&quot; or better) without understanding physics very well.</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Learning physics requires that I substantially rethink, restructure, and reorganize the information that I am given in class and/or in the text.</td>
<td></td>
</tr>
</tbody>
</table>
Introductory Physics Diagnostic Testing

Force Concept Inventory & MPEX Survey: Pre-Course

This diagnostic testing is to learn about what happens in introductory physics classes. The results will help us design materials to help us make this a better course. This testing is voluntary. The results will have no impact on your grade in this course. We do, however, appreciate your cooperation.

- Use a number 2 or H/HB pencil to fill out the scantron form.
- Fill in the bubbles completely.
- Erase completely any marks you made by mistake
- See marking instructions on side 2 of the scantron form if you have questions.

**SIDE 1:**

A. Fill out NAME and SEX in the spaces provided

B. For GRADE OR EDUC, please indicate which of the following Colleges or Universities you attend:

=> 0.) University of Maryland at College Park
   1.) University of Minnesota
   2.) Ohio State University
   4.) University of Maine
   5.) Prince Georges Community College
   6.) Joliet Junior College
   7.) Dickinson College
   8.) Moorhead State University
   9.) Carroll College
   10.) Skidmore College
   11.) Forsyth Technical Community College
   12.) Nebraska Wesleyan University
   13.) Whittier College
   14.) University of Northern Iowa
   15.) Drury College

C. Write TODAY’S DATE. in the space marked “birth date”.

D. Write your STUDENT IDENTIFICATION NUMBER in the space marked “identification number”. The first digit of your student ID number should go in box A.

E. For SPECIAL CODES:
   K and L.) What is your age? Example: If you are 20 K => 2, L => 0.

   M.) Where is this course in the sequence? Most of the classes we survey are part of a multiterm introductory physics sequence. Is this class
       1.) first term.  3.) third term.
       2.) second term.  4.) fourth term.

   N, 0, and P.) This is a code to determine which professor teaches your class and what discussion section you are in (where applicable). The code should be on a sheet given to you or provided by the instructor administering this diagnostic. For UMCP students for example, use the last 3 digits of your section number.
The Force Concept Inventory

If you are doing the Force Concept Inventory (FCI) with the survey, you will be given the FCI on a separate handout. It uses items 1-30. If you are not taking the FCI now, skip down to the next section and begin with item 31.

If you are doing the FCI, fill in items 1-30 on the right half of side 1 of your scantron sheet as follows:

Fill in the answers to questions 1-30 on the FCI in the bubbles numbered 1-30.

Example:

1.) Answer FCI question 1.

When you are done with the FCI return here, go down to the next section, and begin with item 31.

Your Background

We use items 31-60 to help us understand your background.
Place the answer for each item in the correspondingly numbered set of bubbles.

What is your major? Choose from among the items in the lists 31-35.
If you are undecided, choose 35g. If you are a double major, choose your primary major.

31.) Engineering
   a) Aerospace Engineering
   b) Agricultural Engineering
   c) Civil Engineering
   d) Chemical Engineering
   e) Electrical Engineering
   f) Fire Protection Engineering
   g) Engineering Materials
   h) Mechanical Engineering
   i) Nuclear Engineering
   j) Other Engineering

32.) Physical Science
   a) Physics
   b) Chemistry
   c) Mathematics/Statistics
   d) Astronomy
   e) Meteorology
   f) Geology
   g) Chemical Physics
   h) Other Physical Science

33.) Biological Sciences
   a) Biology
   b) Botany
   c) Marine Biology
   d) Biochemistry
   e) Marine/Environmental Science
   f) Agronomy
   g) Animal Science
   h) Zoology/Entomology
   i) Pre-med/Pre-vet
   j) Other Biological Science

34.) Applied
   a) Computer Science
   b) Telecommunications
   c) Business, Management, and Accounting
   d) Architecture
   e) Journalism
   f) Kinesiology
   g) Health, Nutrition, and Food Science
   h) Systems Analysis/Engineering
   i) Computer/Software engineering
   j) Pre-Law

35.) Other
   a) Philosophy
   b) Other Humanities
   c) Social Science
   d) Music/Performing Art
   e) Art
   f) Education
   g) Undecided
   h.) None of the majors listed
36.) Have you taken a physics course before? (Select only one of the following.)
   a) Yes, in high school
   b) Yes, in another college/university
   c) Yes, both in high school and another college/university
   d) Yes, in this college/university
   e) Yes, in high school and this college/university
   f) No, this is my first physics course

37.) Are you repeating this course?
   a) Yes, at this college/university.
   b) Yes, at another college/university
   c) No.

**Answer 38, 39, & 40 only if you have taken a physics course before.**

38.) Did you feel it was a good course?
   a) yes
   b) no
   c) so-so

39.) Did you feel you did well in that course?
   a) yes
   b) no
   c) so-so

40.) When did you take it?
   a) this year
   b) last year
   c) two years ago
   d) three years ago
   e) four years ago
   f) five years ago
   g) more than five years ago

41.) What was the last math class you completed prior to taking this course?
   a) first year Algebra
   b) Geometry
   c) second year Algebra
   d) Trigonometry/Math Analysis
   e) first semester Calculus
   f) second semester Calculus
   g) third semester Calculus
   h) Other

42.) When did you take it?
   a) this year
   b) last year
   c) two years ago
   d) three years ago
   c) four years ago
   f) five years ago
   g) more than five years ago

43.) Did you take calculus in high school?
   a) yes
   b) no
For 44 & 45, use the following responses
a.) 0 hours  f.) 9-10 hours
b.) 1-2 hours  g.) 11-12 hours
c.) 3-4 hours  h.) 13-14 hours
d.) 5-6 hours  i.) 15-16 hours
e.) 7-8 hours  j.) more than 16 hours

44.) The number of hours per week I plan to spend reading, studying, and doing homework for this course(not counting time in labs, writing lab reports, and reviewing for exams) is about: ______

45.) The number of hours I plan to spend preparing for an exam in this course is about: ______

46.) I plan to work with others _______ of the total time I plan to spend on physics outside of class (excluding labs).
   a.) 0 %  f.) 50 %
b.) 10 %  g.) 60 %
c.) 20 %  h.) 75 %
d.) 30 %  i.) 90 %
e.) 40 %  j.) 100 %

For the list of skills below in 47 - 58, rate some of your relevant abilities using the following scale:

A = Excellent  B = Good  C = Average  D = Weak  E = Poor  F = Can’t say

47.) Understanding physics textbooks
48.) Understanding physics lectures (if you had them)
49.) Understanding experiments and demonstrations
50.) Taking tests
51.) Laboratory skills
52.) Solving homework problems on your own
53.) Expressing yourself clearly in writing
54.) Convincing others of your point of view
55.) Algebra
56.) Trigonometry
57.) Calculus
58.) Vectors

PLEASE SKIP ITEMS 59 & 60 AND CONTINUE WITH THE MPEX SURVEY ON THE NEXT TWO PAGES.
SIDE 2: MPEX Survey

TURN THE SCANTRON ANSWER SHEET OVER AND START WITH NUMBER 61.

Here are 35 statements (Items 60 - 95) which may or may not describe your beliefs about this course. You are asked to rate each statement by circling a number between A and E where the letters mean the following:

<table>
<thead>
<tr>
<th>A: Strongly Disagree</th>
<th>B: Disagree</th>
<th>C: Neutral</th>
<th>D: Agree</th>
<th>E: Strongly Agree</th>
</tr>
</thead>
</table>

Answer the questions by circling the number that best expresses your feeling. Work quickly. Don’t over-elaborate the meaning of each statement. They are meant to be taken as straightforward and simple. If you do not understand a statement, leave it blank. If you understand, but have no strong opinion one way or the other, circle C. If an item combines two statements and you disagree with either one, choose A or B.

61.) All I need to do to understood most of the basic ideas in this course is just read the text, work most of the problems, and/or pay close attention in class.

62.) All I learn from a derivation or proof of a formula is that the formula obtained is valid and that it is OK to use it in problems.

63.) I plan to go over my class notes carefully to prepare for tests in this course.

64.) "Problem solving" in physics basically means matching problems with facts or equations and then substituting values to get a number.

65.) Learning physics will make me change some of my ideas about how the physical world works.

66.) I expect to spend a lot of time figuring out and understanding at least some of the derivations or proofs given either in class or in the text.

67.) I plan to read the text in detail and work through many of the examples given there.

68.) In this course, I do not expect to understand equations in an intuitive sense; they just have to be taken as givens.

69.) The best way for me to learn physics is by solving many problems rather than by carefully analyzing a few in detail.

70.) Physical laws have little relation to what I experience in the real world.

71.) A good understanding of physics is necessary for me to achieve my career goals. A good grade in this course is not enough.

72.) Knowledge in physics consists of many pieces of information each of which applies primarily to a specific situation.

73.) My grade in this course will be primarily determined by how familiar I am with the material. Insight or creativity will have little to do with it.

74.) Learning physics is a matter of acquiring new knowledge that is specifically located in the laws, principles, and equations given in class and/or in the textbook.
75.) In doing a physics problem, if my calculation gives a result that differs significantly from what I expect, I'd have to trust the calculation.
76.) The derivations or proofs of equations in class or in the text has little to do with solving problems or with the skills I will need to succeed in this course.

77.) Only very few specially qualified people are capable of really understanding physics.

78.) To understand physics, I expect to think about my personal experiences and relate them to the topic being analyzed.

79.) The most crucial thing in solving a physics problem is finding the right equation to use.

80.) If I didn't remember a particular equation needed for a problem in an exam there is nothing much I could do (legally!) to come up with it.

81.) If I came up with two different approaches to a problem and they gave different answers, I would not worry about it; I would just choose the answer that seemed most reasonable. (Assume the answer is not in the back of the book.)

82.) Physics is related to the real world and it sometimes helps to think about the connection, but it is rarely essential for what I will have to do in this course.

83.) The main skill I expect to get out of this course is learning how to solve physics problems.

84.) The results of an exam won't give me any useful guidance to improve my understanding of the course material. All the learning associated with an exam will be in the studying I do before it takes place.

85.) Learning physics helps or will help me understand situations in my everyday life.

86.) When I solve most exam or homework problems, I explicitly think about the concepts that underlie the problems.

87.) "Understanding" physics basically means being able to recall something you've read or been shown.

88.) Spending a lot of time (half an hour or more) working on a problem is a waste of time. If I don't make progress quickly, I am better off asking someone who knows more than I do.

89.) A significant problem in this course will be being able to memorize all the information I need to know.

90.) The main skill I expect to get out of this course is to learn how to reason logically about the physical world.

91.) I will use the mistakes I make on homework and on exam problems as clues to what I need to do to understand the material better.

92.) To be able to use an equation in a problem (particularly a problem that I haven't seen before), I need to know more than what each term in the equation represents.

93.) It is possible to pass this course (get a "C" or better) without understanding physics very well.

94.) Learning physics requires that I substantially rethink, restructure, and reorganize the information that I am given in class and/or the text.
REFERENCES


A. diSessa, Cognit. Instruct. 10 (2 and 3), 105-225 (1993).


R. Likert, “A technique for the measurement of attitudes,” *Archives of Psychology* No. 140 (1932).


W. F. Perry, *Forms of Intellectual and Ethical Development in the College Years* (Holt, Rinehart, and Wilson, NY, 1970).


E. F. Redish, “Is the computer appropriate for teaching physics?” *Computers in Physics* 7, 613 (December 1993).


