

## ABSTRACT

LIN, DING. Designing an Energy Assessment to Evaluate Student Understanding of Energy Topics. (Under the direction of Ruth W. Chabay and Bruce A. Sherwood.)

The well-established approaches to energy in traditional introductory mechanics courses are often oversimplified or even erroneous. Unlike the traditional courses, *Matter & Interactions (M&I) Modern Mechanics* presents students a scientific view of energy by emphasizing the energy principle and the atomic nature of matter. Motivated by a great need for an appropriate assessment tool that matches the accurate approaches to energy as employed in the *M&I* mechanics course, this study carries out a valid and reliable energy assessment to evaluate student understanding of energy topics. This energy assessment is a 33-item multiple-choice test and is suitable for the *M&I* mechanics course or courses of similar content and approaches. In general, questions in the energy assessment test higher-level thinking yet involve only short reasoning processes.

Students from different academic levels participated in completing the energy assessment. The majority participants are students from the *M&I* mechanics course who took both the pretest and posttest in the 2006 fall semester at North Carolina State University. Results from a series of quantitative analyses show that the *M&I* students performed significantly better in the posttest than in the pretest not only on the entire assessment, but also on most of the individual items and all the test objectives. Moreover, a small number of student interviews were conducted to probe student reasoning. Qualitative analyses of the student interviews indicate that students are able to use the energy principle correctly to tackle physics questions if they choose to start from the fundamental principles. Another aspect highlighted in the interviews is that students are capable of performing qualitative analysis without using exact formulas.

**Designing an Energy Assessment to Evaluate Student  
Understanding of Energy Topics**

by

**Lin Ding**

A dissertation submitted to the Graduate Faculty of

North Carolina State University

In partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

**Physics**

Raleigh, North Carolina

2007

**Approved by:**

---

Dr. Robert Beichner

---

Dr. John Risley

---

Dr. Ruth Chabay  
Chair of Advisory Committee

---

Dr. Bruce Sherwood  
Co-chair of Advisory Committee

## **DEDICATION**

To my dearest maternal grandparents

**YongFa Wu and ABao Hong**

Who

Once tugged me under their cozy wings of aegis

Who

Nourished me with their pure love and great wisdom

And who

Held my hands and

Weaved, along with me, the first thread of my dreams!

## BIOGRAPHY

With my initial interest in experimental sciences, I obtained my B.S. in electric light source engineering from Fudan University in Shanghai, China, and M.S. in atomic physics from Texas Christian University. In January 2003, I came to the physics department at North Carolina State University, where I switched my focus to physics education research. This major shift in research was largely due to my long existing interest in science education, and partially due to my rebellion against such a notion that “those, who can’t do, teach”.

My current interest in physics education research mainly concerns the effective and efficient means to evaluate student conceptual understanding of physics concepts. Aside from this particular focus, I am also interested in neurosciences, cognitive sciences, statistics and their applications in physics education.

## ACKNOWLEDGMENTS

Going through a doctorate program is rather similar to running a marathon; it requires patience, perseverance as well as intelligence. Along the way, I received much support and assistance, without which this dissertation wouldn't have been possible.

I thank my dissertation advisors Ruth Chabay and Bruce Sherwood for their professional guidance. As much as I enjoy their scholarly acumen, I am often amazed to see how their insightful comments can touch a dreadful impasse into a golden Broadway, or tear a seemingly promising idea into broken pieces. Their role of magician during my work has left me a life-long memory.

I also owe a great debt to other advisory committee members: Robert Beichner and John Risley. I especially appreciate Robert Beichner's humor, efficiency and timely manner in handling every case I witnessed. His generous support during my research and quick responses to any of my request has been indeed invaluable.

I thank Mark Haugan, Matthew Kohlmyer, Prabha Ramakrishnan, and the many faculty members at North Carolina State University (NCSU) who provided valuable suggestions and feedback to my studies. I am also grateful for the useful conversations with the former and current graduate students and post-docs in the Physics Education Research group at NCSU; special thanks to Shawn Weatherford who volunteered to proofread part of my dissertation manuscripts. Thanks also go to the National Science Foundation, who provided partial funding for this project.

Finally, I would like to thank my parents, ZhaoJing Ding and ZhengMei Wu, and my aunt MeiLi Wu, for their unwavering love and encouragement.

## TABLE OF CONTENTS

LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi
CHAPTER 1: Introduction.....	1
1.1 Highlights of the energy topics in the <i>M&amp;I</i> mechanics course.....	2
1.2 A brief introduction to assessment instruments.....	4
1.3 Research questions and organization of the dissertation.....	7
CHAPTER 2: Literature Review.....	9
2.1 Students' difficulties with energy and energy related topics.....	11
2.1.1 Students' difficulties with the concept of energy.....	11
2.1.2 Students' difficulties with the concept of work.....	14
2.1.3 Students' confusions among work, heat and internal energy.....	19
2.1.4 Students' difficulties in applying the energy principle.....	24
2.2 Energy in the <i>Matter &amp; Interactions Modern Mechanics</i> course.....	28
2.3 Assessment instruments in physics education research.....	33
2.3.1 Validity and reliability.....	33
2.3.2 Using a multiple-choice test.....	37
2.3.3 Assessment instruments on energy.....	40
CHAPTER 3: Methodology.....	48
3.1 Development of the energy assessment.....	48

3.1.1 Test Objectives.....	48
3.1.2 Item creation and review.....	52
3.1.3 Item revision.....	59
3.1.4 Two versions of the energy assessment.....	71
3.2 Validity, reliability and discriminatory power of the energy assessment.....	72
3.2.1 Validity of the energy assessment.....	72
3.2.2 Reliability and discriminatory power of the energy assessment.....	74
3.3 Student participants.....	76
3.3.1 Sample 1.....	77
3.3.2 Sample 2.....	77
3.3.3 Sample 3.....	78
3.3.4 Sample 4.....	79
3.3.5 Sample 5.....	80
3.4 Test Settings and procedures.....	81
3.4.1 Test settings and procedures for sample 1.....	81
3.4.2 Test settings and procedures for sample 2.....	82
3.4.3 Test settings and procedures for sample 3.....	84
3.4.4 Test settings and procedures for sample 4.....	85
3.4.5 Test settings and procedures for sample 5.....	86
3.5 Quantitative and qualitative analyses.....	87
3.5.1 Quantitative analyses.....	87
3.5.2 Qualitative analyses.....	88

CHAPTER 4: Result I—Results on the Energy Assessment.....	91
4.1 Reliability and discriminatory power of the energy assessment.....	91
4.1.1 Pilot version.....	91
4.1.2 Current version.....	98
4.2 Content and cognition level of the test items.....	103
4.3 Reasoning steps.....	112
4.4 Issues on symbolic/numerical representations and diagrams.....	118
4.4.1 Symbolic representation vs. numerical representation.....	118
4.4.2 Diagram vs. no diagram.....	126
4.5 Summary.....	129
CHAPTER 5: Result II—Student Performance on the Energy Assessment.....	131
5.1 Results for student sample 2— <i>M&amp;I</i> students.....	132
5.1.1 Results on pretest.....	133
5.1.2 Results on posttest.....	140
5.1.3 Comparisons between the pretest and the posttest.....	146
5.2 Results for student sample 3— <i>M&amp;I (U)TAs</i> .....	157
5.3 Results for student sample 4—senior physics major undergraduates.....	159
5.4 Results for student sample 5—first year physics major graduate students at Purdue.....	161
5.5 Factor analysis of student responses to the energy assessment.....	163
5.6 Qualitative discussion of student interviews.....	167
5.6.1 Application of the energy principle.....	168

5.6.2 Determination of energy and interpretation of energy graphs.....	179
5.6.3 Determination of work and differentiation between work and heat....	188
5.6.4 Specification of systems.....	193
5.6.5 Calculation of atomic spectrum.....	194
5.7 Summary.....	195
 CHAPTER 6: Conclusion.....	200
6.1 Summary of major results.....	200
6.1.1 Results regarding the energy assessment.....	200
6.1.2 Results regarding student performance on the energy assessment.....	202
6.2 Implications for instructions.....	205
6.3 Future work.....	207
 REFERENCES.....	209
APPENDICES.....	214
Appendix A: Faculty Review.....	215
Appendix B: Current Version of the Energy Assessment.....	217
Appendix C: Pilot Version of the Energy Assessment.....	234

## LIST OF TABLES

Table 2.1	Seven topics covered in the TUG-K.....	34
Table 2.2	Statistical evaluations of some standardized physics tests.....	37
Table 3.1	Four panel experts.....	51
Table 3.2	Ten faculty reviewers.....	56
Table 3.3	Comparison between the pilot version and current version of the energy assessment.....	72
Table 3.4	Student participants from the four sections of the <i>M&amp;I</i> mechanics course in lecture hall settings in the 2006 fall semester at NCSU.....	78
Table 3.5	Five samples of participants who took one of the two versions of the energy assessment.....	80
Table 3.6	Test settings and administration formats for the five samples of participants.....	86
Table 4.1	Item analysis of the pilot version of the energy assessment.....	93
Table 4.2	Evaluation of the pilot version of the energy assessment.....	98
Table 4.3	Student participants whose data were used for the item analysis and test analysis of the current version of the energy assessment.....	99
Table 4.4	Item analysis of the current version of the energy assessment.....	100
Table 4.5	Evaluation of the current version of the energy assessment.....	103
Table 4.6	The content dimension of the revised Bloom's taxonomy.....	106
Table 4.7	The cognition dimension of the revised Bloom's taxonomy.....	107
Table 4.8	The content and cognition level for the individual items in the current energy assessment.....	109

Table 4.9	The number of reasoning steps for each item in the current energy assessment.....	115
Table 4.10	Students who participated in taking either the symbolic or numerical version of the quiz.....	122
Table 4.11	The percentage of correct responses among non-physics major students....	123
Table 4.12	The percentage of correct responses among physics major students.....	124
Table 5.1	Five samples of students who completed either the pilot version or the current version of the energy assessment.....	132
Table 5.2	Pretest alternative choice distributions (The correct answers are underlined.).....	137
Table 5.3	Five major test objectives and pretest objective scores.....	139
Table 5.4	Posttest alternative choice distributions (The correct answers are underlined.).....	144
Table 5.5	A contingency table for Q5.....	154
Table 5.6	Results of the paired t test in objective scores.....	156
Table 5.7	The first factor from the factor analysis.....	165
Table 5.8	The second factor from the factor analysis.....	167

## LIST OF FIGURES

Figure 2.1	Two series of systems having identical positions but different motions (Vienott, 1979).....	12
Figure 2.2	A question from the multiple-choice test of energy and momentum concepts (Singh and Rosengrant, 2003).....	15
Figure 2.3	A mechanical problem (Loverude, Kautz, and Heron, 2001).....	17
Figure 2.4	Versions A and B of the bicycle pump problem and the insulated cylinder problem (Loverude, Kautz, and Heron, 2001).....	21
Figure 2.5	Isobaric expansion of a system (Meltzer, 2004).....	23
Figure 2.6	A two-puck demonstration (Lawson and McDermott, 1987).....	25
Figure 2.7	A question from the multiple-choice test of energy and momentum concepts (Singh and Rosengrant, 2003).....	42
Figure 2.8	A question from the multiple-choice test of energy and momentum concepts (Singh and Rosengrant, 2003).....	43
Figure 2.9	Choices for a question in the ECI (Hestenes and Swackhamer, unpublished).....	45
Figure 2.10	A question from MBT (Hestenes and Wells, 1992).....	46
Figure 2.11	A two-puck question from MBT (Hestenes and Wells, 1992).....	46
Figure 3.1	Hierarchical structure of the test objectives.....	50
Figure 3.2	Development of the energy assessment.....	58
Figure 3.3	Item 1 in the current energy assessment.....	59
Figure 3.4	The original form of item 1 .....	60
Figure 3.5	Item 6a and 6b in the current energy assessment.....	60

Figure 3.6	The original form of item 6a and 6b.....	61
Figure 3.7	Item 4 in the current energy assessment.....	62
Figure 3.8	Item 5 in the current energy assessment.....	62
Figure 3.9	Item 7 in the current energy assessment.....	63
Figure 3.10	Item 8 in the current energy assessment.....	64
Figure 3.11	Item 17 in the current energy assessment.....	65
Figure 3.12	The original form of item 17.....	66
Figure 3.13	Item 14 in the current energy assessment.....	66
Figure 3.14	Item 18 in the current energy assessment.....	67
Figure 3.15	Item 22 in the current energy assessment.....	68
Figure 3.16	Item 24 in the current energy assessment.....	69
Figure 3.17	Item 25 in the current energy assessment.....	70
Figure 3.18	An item that was removed during the expert panel review due to its low relevancy to physics.....	73
Figure 4.1	Item difficulty index, discrimination index and point bi-serial coefficient plots for the pilot version of the energy assessment.....	94
Figure 4.2	Three items with the lowest difficulty index values in the pilot version of the energy assessment.....	96
Figure 4.3	An item with a negative discrimination index and point bi-serial coefficient in the pilot version of the energy assessment.....	97
Figure 4.4	Item difficulty index, discrimination index and point bi-serial coefficient plots for the current energy assessment.....	101
Figure 4.5	The revised Bloom's taxonomy.....	104

Figure 4.6	A 3-D bar chart of the content and cognition levels for the items in the current energy assessment.....	110
Figure 4.7	An item in the current energy assessment that requires the application of concept.....	110
Figure 4.8	An item in the current energy assessment that requires the application of both the momentum principle and the energy principle.....	111
Figure 4.9	Items that require the comprehension of concepts.....	112
Figure 4.10	A histogram of the number of reasoning steps for the items in the current energy assessment.....	116
Figure 4.11	The symbolic version of the quiz.....	120
Figure 4.12	The numerical version of the quiz.....	121
Figure 4.13	A bar chart of the percentage of correct responses for the two versions among the non-physics major students.....	123
Figure 4.14	A bar chart of percentage of correct responses for the two versions among the physics major students.....	125
Figure 4.15	An item was initially created with a diagram.....	127
Figure 4.16	The percentage of correct responses for the versions of the item.....	128
Figure 5.1	Pretest histogram of the current energy assessment for four <i>M&amp;I</i> sections collectively.....	133
Figure 5.2	Histogram of the test scores produced by random guessing (computer generated data).....	134
Figure 5.3	Comparison between two sets of data: computer generated data and collected pretest data.....	135

Figure 5.4	Pretest item scores.....	136
Figure 5.5	A bar chart of the pretest objective scores. (The objectives are indicated by the corresponding topics names.).....	139
Figure 5.6	Posttest histogram of the current energy assessment for four <i>M&amp;I</i> sections collectively.....	140
Figure 5.7	Comparison between two sets of data: computer generated data and collected posttest data.....	141
Figure 5.8	Posttest item scores.....	142
Figure 5.9	A bar chart of posttest objective scores. (The objectives are indicated by the corresponding topics names.).....	145
Figure 5.10	A pair-wise comparison between pre- and post-test scores.....	147
Figure 5.11	A scatter plot for the pair-wise pre- and post-test scores.....	148
Figure 5.12	Pre- and post-test histograms and scatter plots for four <i>M&amp;I</i> sections individually.....	149
Figure 5.13	A pair-wise comparison between the pre- and post-test item scores.....	153
Figure 5.14	A 3-D bar chart for Q5.....	154
Figure 5.15	A pair-wise comparison between pre- and post-test objective scores (The objectives are indicated by the corresponding topics names.).....	156
Figure 5.16	Score histogram of the current energy assessment for the <i>M&amp;I</i> (U)TAs.....	158
Figure 5.17	Item scores for the <i>M&amp;I</i> (U)TAs.....	159
Figure 5.18	Score histogram of the current energy assessment for the senior physics major undergraduates.....	160

Figure 5.19	Item scores for the senior physics major undergraduates.....	161
Figure 5.20	Score histogram of the current energy assessment for the first-year physics major graduate students at Purdue.....	162
Figure 5.21	Item scores for the first year graduate students at Purdue.....	163
Figure 5.22	A scree plot showing the eigenvalues of the extracted factors from the factor analysis.....	164
Figure 5.23	A summary of average total scores for four samples of students.....	196

## CHAPTER 1: Introduction

Energy, as one of the fundamental topics in the physics enterprise, has a considerably rich content and a wide range of applications in various physical sciences as well as in our daily lives. Inevitably, the energy concepts, in every sense, occupy a rather important position in the university-level introductory physics courses, both the calculus-based and the algebra-based. As a matter of fact, in many traditional mechanics courses, the energy concepts are, albeit fundamental, often introduced quite late, and consequently are viewed by students as secondary or even tertiary topics in the course. What's worse, some widely adopted physics textbooks, which may originally intend to simplify and sanitize the messy real-world phenomena for good purposes, often result in conveying to students an incomplete or even erroneous picture of energy. In light of this type of issue, research in physics education has started to focus on the content reform of the introductory physics curriculum so as to present students a contemporary view of central topics, such as energy, in the physics subject.

*Matter & Interactions (M&I)*, designed by Chabay and Sherwood (2007), is such an innovative introductory physics curriculum that aims to bring the scientific and unified perspective of the physical fundamentals into the introductory course. Specifically, its careful and systematic approach to the energy topics in the *M&I Modern Mechanics* course—the first semester of a two-semester sequence—indeed differs remarkably from the long-established treatment of the energy topics in the traditional introductory mechanics courses. As a result, the existing assessments on the energy topics that are largely designed for the traditional mechanics courses become inappropriate and even misleading for the students

from the *M&I* mechanics classes. To this end, it is necessary and valuable to develop a new energy assessment that is appropriate for the *M&I* mechanics course.

Motivated by this consideration, the study addressed in this dissertation is devoted to the development of such a new energy assessment as well as to the evaluation of the *M&I* students' performance using this assessment. By doing so, it is the author's goal that this energy assessment can become a useful instrument for the *M&I* mechanics course and thus contribute to the continual development of the *M&I* curriculum. Moreover, it is the author's great hope that, through the discreet circulation of the energy assessment, the incorrect treatment of the energy topics in the traditional mechanics courses will be further manifested, and the precise and scientific approaches employed in the *M&I* curriculum will eventually be adopted by more instructors.

Before outlining the research questions and the organizations of the dissertation, it is informative to highlight briefly some approaches of *M&I* to the energy topics that are often inadequately or incorrectly addressed in the traditional introductory mechanics courses. Moreover, it is also useful to provide readers a brief introduction on assessments as a precursor to the subsequent materials in this dissertation.

### **1.1 Highlights of the energy topics in the *M&I* mechanics course**

One central focus of the *M&I* mechanics course is the proper application of the fundamental principle associated with energy, namely the energy principle  $\Delta E_{\text{system}} = W + Q$ . In traditional introductory mechanics courses, this fundamental principle is often called the “work-energy theorem”, and it only includes work—the mechanical transfer of energy—as

the means to change the energy of a system<sup>1</sup>. Consequently, the “work-energy theorem” is simply expressed as  $\Delta E = W$ . In addition, the traditional treatment of a system of interest often carelessly idealizes the system as a rigid particle-like object without any forms of internal energy. As a result, much lost is the real essence of the relationship between the energy transfer processes and the energy change of a system. What’s more, it becomes almost impossible for students in the traditional mechanics classes to be able to deal with real-world systems, which, in many cases, are composed of multiple particles and are often deformable or rotatable. In the *M&I* mechanics course, the systems of interest are generalized to non-rigid multi-particle objects. In the analysis of such a system, students use a scheme called “point-particle system” to determine the translational kinetic energy for the center of mass of the system, as well as use the “real system” to find out the various forms of internal energy of the system. (See, for example, Chabay & Sherwood, 1999.) To determine the work done on a deformable and/or rotatable system, students in the *M&I* mechanics course learn to take into consideration the actual work done by the individual forces at the contact points where the external forces are applied.

In an attempt to identify the external forces and thus to determine the external work, students first need to specify a system of interest. In the *M&I* mechanics course, one important feature is the careful and precise specification of a system and its surroundings. In so doing, the distinction between work and potential energy is facilitated, and students are able to decide which terms ought to appear on the left or the right side of the energy principle depending on the choice of the system (Chabay & Sherwood, 2004).

---

<sup>1</sup> Typically, in the traditional introductory mechanics courses, the heat concept ( $Q$ ) is not introduced until late in the course where the thermal physics comes to play.

Closely related to the system specification is the definition of potential energy. In the traditional mechanics courses, potential energy is often attributed to a single object. This incorrect notion is completely abandoned in the *M&I* mechanics course. Instead, this course explicitly emphasizes the inseparable association of potential energy with the pair-wise interactions between objects. Consequently, students in the *M&I* mechanics course are aware that one particle does not have any forms of potential energy, and that potential energy only exists in a system that consists of at least two particles.

Another noteworthy aspect regarding energy in the *M&I* mechanics course is the introduction of the relativistic form of particle energy  $E = \frac{mc^2}{\sqrt{1 - v^2/c^2}}$ . As appreciated by the physics experts more than by the novices, this relativistic expression in fact satisfies the relation that connects the spatial derivative of the particle energy ( $\nabla E$ ) and the time derivative of the particle momentum ( $\frac{d\vec{p}}{dt}$ ). In one dimension, this relation can be simply expressed as  $\frac{dE}{dx} = \frac{dp}{dt}$ . Often, it is useful to separate particle energy into rest energy and (relativistic) kinetic energy. Practically, students, when analyzing a particular system, take into consideration the rest energy of the system in case the identity of the system should change. Additionally, students learn to make valid approximations for the kinetic energy of a system through comparing the speed of the system with the speed of light.

## 1.2 A brief introduction to assessment instruments

The energy assessment that the study aims to design is in the multiple-choice format. Readers may wonder if there is any particular reason to select this format rather than others,

and if there are other formats that one can use to design an assessment. In this section, a brief introduction on different formats of assessment is provided to answer these questions.

Primarily, there exist at least three distinctly different test formats, including the multiple-choice test format, the constructed response test format, and the performance test format. (Haladyna, 1997)

Multiple-choice tests are perhaps the most widely used assessment format with which many readers are probably rather familiar. Multiple-choice test questions are easy to grade and less likely to induce subjective grading, as questions in this format are free of responses (Mehrens & Lehmann, 1991). More importantly, data collected from the multiple-choice tests are conveniently amenable to different statistical analyses; thereby both the descriptive and the inferential results can become practically obtainable. Since a multiple-choice test can be efficiently administered among a large number of students simultaneously, the results are often generalizable (Beichner, 1994) and the process is less time demanding. Moreover, a multiple-choice test can cover different levels of cognitive behaviors (Doran, 1980) as well as different levels of knowledge content (Haladyna, 1997); therefore, it is feasible to use a well-designed multiple-choice test to evaluate higher order thinking, such as application of principles.

Nevertheless, the multiple-choice format has an intrinsic disadvantage when it comes to the student reasoning process. A major concern about the multiple-choice questions is that questions in this format potentially encourage random guessing and cannot truly reflect students' reasoning process. (Bork, 1984; Varney, 1984; Sandin, 1985; Scott, 1985; Mehrens & Lehmann, 1991) However, when multiple-choice tests are combined with student

interviews, this concern can be alleviated, as the interviews can reveal in depth student reasoning. (See, for example, Patton 1990)

Constructed response format differs greatly from the multiple-choice format in that the constructed response questions ask students to fill-in-a-blank or generate a short written response to a question, rather than choose from a set of possible answers. The major advantage to the constructed response format is that student responses are less affected by guessing, and clues about students' thought processes can be provided (Powell & Gillespie, 1990). However, the constructed response format lacks objectivity in scoring, so tests in this format are generally less reliable and more difficult to grade.

Similar to the constructed response test format, the performance test format requires students to provide a free response to a question rather than selecting from a list of given choices. Contrary to the constructed response tests, the performance tests take various forms that are not limited to the written responses. Examples of performance tests include exhibitions, demonstrations, written or oral presentations, journals and portfolios (Haladyna, 1999). Although the performance tests provide students great opportunities to present their organization skills, creativity, style, expressiveness and the like, the constructed response format "has serious limitations, not the least of which are lack of objectivity in scoring, judge bias, and inconsistency, and the high cost of scoring" (Haladyna, 1997).

Considering these aforementioned pros and cons of different tests, it is clear why the multiple-choice format has emerged to be the most favorable and practical format for assessment in large scale. Based on this consideration, the energy assessment is designed into the multiple-choice format.

### **1.3 Research questions and organization of the dissertation**

As mentioned above, the goal of this study is to design an appropriate energy assessment for the *M&I* mechanics course so as to evaluate student performance on the relevant energy topics. Specifically, three research questions, as addressed below, are what this study mainly seeks to answer.

- *Can a multiple-choice test be used reliably to assess student performance on the energy topics that are covered in the M&I mechanics course?*
  - (1) *Is the energy assessment a valid test for the M&I mechanics course*
  - (2) *Is the energy assessment a reliable test?*

If the answers to the above questions are affirmative, it is then reasonable to consider the following questions.

- *Do students from the M&I mechanics class achieve the course objectives on the energy topics after the instructions?*

Since student performance on the energy assessment can provide useful insights into the effectiveness of the course instructions, it is then valid to consider:

- *Is the M&I mechanics course effective in promoting student understanding of the energy topics?*

In order to answer these questions, this dissertation is accordingly organized into four major components.

Chapter 2 is devoted to an intensive review of the relevant literature on the energy topics addressed in the introductory physics courses at the university level. Also reviewed is the literature on research-based assessments in the physics education field.

Chapter 3 presents the methodology of the study. Specifically, this chapter provides detailed descriptions on how the energy assessment was developed, validated and administered. Moreover, this chapter describes how the data collected from student written responses were analyzed, and how student reasoning was probed through conducting a small number of student interviews.

Chapter 4 and Chapter 5 essentially lay the foundations for answering these aforementioned research questions. In particular, Chapter 4 reports the detailed results on the energy assessment, including the reliability and discriminator power of the assessment, content/cognition levels and the reasoning steps involved in the individual items. Chapter 5 examines student pre- and post-instructional performance on the energy assessment in large scale, discusses student reasoning through student interviews in small scale, and addresses student achievement on the energy topics after the course instructions.

Chapter 6 summarizes the important findings and provides useful suggestions for the future directions of the research topic.

## CHAPTER 2: Literature Review

This literature review is devoted to the relevant studies on the following three subjects:

- (1) Students' difficulties with energy and energy related topics
- (2) Energy in the *Matter & Interactions (M&I) Modern Mechanics* course
- (3) Assessment tools in physics education research

Section 2.1 reviews students' difficulties with energy and energy related topics. Studies conducted by Vienott (1979), Goldring and Osborne (1994), Singh and Rosengrant (2003), Loverude, Kautz and Heron (2001), Meltzer (2004), Lawson and McDermott (1987), and Pride, Vokos and McDermott (1997) suggest that students often have the following difficulties regarding energy and its related topics: difficulties with the concept of energy, difficulties with the concept of work, confusion among work, heat and internal energy, and inability to apply the energy principle<sup>2</sup>.

Students' difficulties with the concept of energy are examined in subsection 2.1.1. This subsection shows that students often confuse energy with force and power.

Students' difficulties with the concept of work are reviewed in subsection 2.1.2. These difficulties include students' confusion between work and force; students' difficulties with path-independence of work done by the gravitational force; students' difficulties in determining the sign of the work done on an object; and students' inability of relating the work done *on* and *by* a system.

---

<sup>2</sup> In the literature, the energy principle is often called the “work-energy theorem”; thereby it is only limited to the causal relation between work and energy change of a system. As a matter of fact, there exist more ways to change the energy of a system than work. For an example, heat and radiation. For the purpose of consistency, the name of “the energy principle” is used thereafter.

Subsection 2.1.3 reviews students' confusion among work, heat and internal energy, and shows that students often use these concepts interchangeably.

Subsection 2.1.4 examines students' inability to apply the energy principle. In this subsection, analyses on the application of the energy principle to non-rigid body systems are also discussed. Analyses by Sherwood (1983), Sherwood and Bernard (1984), Arons (1989, 1999), Mallinckrodt and Leff (1991, 2002), and Leff and Mallinckrodt (1992) suggest that the energy principle is often misused, particularly in the non-rigid body systems.

Section 2.2 briefly reviews the unique approaches of the *Matter & Interactions Modern Mechanics* course to the topic of energy. *Matter & Interactions* is an innovative curriculum for introductory physics designed by Chabay and Sherwood (2002). This curriculum emphasizes a small number of fundamental principles and focuses on the atomic nature of matters and their interactions. In the *M&I Modern Mechanics* course, the first semester of a two-semester sequence, some unique approaches to the topic of energy are designed to offer students a more scientific view of energy and its place in the unified physics enterprise.

Section 2.3 reviews some standardized assessment instruments in physics education research. This section includes three subsections.

Subsection 2.3.1 examines the validity and reliability of assessment tools. Validity and reliability are two important aspects of a useful assessment tool. Validity refers to the extent to which an assessment tool measures what it purports to measure. Reliability refers to how consistent an assessment tool is in assessing students' knowledge in a certain domain.

Subsection 2.3.2 reviews studies and discussions on how to use and interpret multiple-choice tests in physics education. A study by Henderson (2003) shows that students'

random guessing in multiple-choice test can be identified through checking the answer patterns. Discussions by Heller et al. (Huffman & Heller, 1995a; Heller & Huffman, 1995b) and Hestenes et al. (Hestenes & Halloun, 1995a; Halloun & Hestenes, 1995b) suggest that interpretation of multiple-choice tests through a certain type of statistical analysis (factor analysis) requires great caution.

Subsection 2.3.3 examines the published and unpublished tests on various topics with items involving energy. These tests include the Mechanics Baseline Test (Halloun & Hestenes, 1985), a multiple-choice test of momentum and energy concepts (Singh & Rosengrant, 2003), and the Energy Concept Inventory (Swackhamer & Hestenes, unpublished).

## **2.1 Students' difficulties with energy and energy related topics**

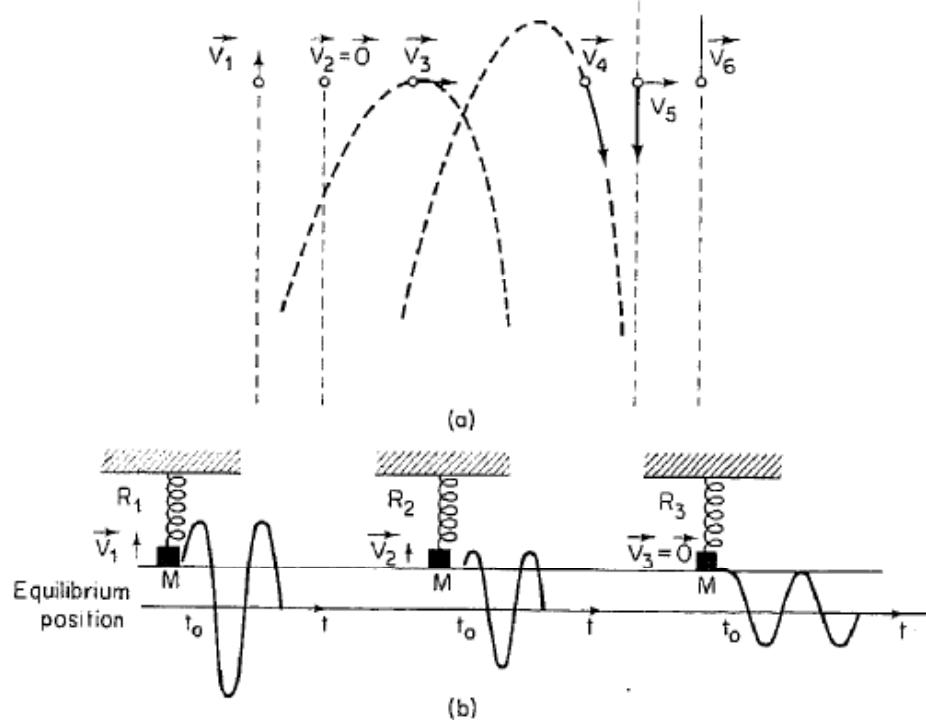
### 2.1.1 Students' difficulties with the concept of energy

A few studies have documented students' difficulties with the energy concept in introductory physics at the college or college equivalent level. These studies by Viennot (1979) and Goldring & Osborne (1994) identified students' confusions of the energy concept in general with force and power.

Viennot (1979) investigated students' spontaneous reasoning in elementary dynamics. The student population comprised mostly of first and second year university students in France, England and Belgium, and some third and last year university students in France and England. Only 29 students (less than 5%) were last year secondary school students in France.

In this study, Viennot showed students several simple systems (see below) with their motion "frozen" at a time instant. In one case, the system is a set of juggler's balls captured

in the air, all at the same height. In another case, the system is a set of spring-mass oscillators suspended from the ceiling, all at the equilibrium position but with different velocity. Students were asked very simple questions, such as “Is there a force?” or “Will an object move or not?”



**Figure 2.1 Two series of systems having identical positions but different motions. (Vienott, 1979)**

The original intention of the researcher is to understand how students think about dynamics given specific situations, and to describe and formulate their thinking. In order to avoid any confound from mathematical difficulties students might have, the researcher only asked very basic conceptual questions, so much so that “one hardly dares to ask it”. To the researcher’s surprise, however, students’ responses revealed great deficiencies in understanding force, motion and energy. One deficiency is students’ confusion of energy and force. Though sometimes students were able to use the energy concept correctly in the case

of oscillators, many students inextricably mixed this concept with force. Some students even explicitly commented that, “The force must be equal to the kinetic energy...” This half vector half scalar notion was characterized by the researcher as “impulse” in ordinary language and as “impetus” in pre-Galilean dynamics.<sup>3</sup> Although from the published paper it is not clear whether this study was conducted after the energy topic was covered in class, the author did mention that students’ spontaneous reasoning as an “intuitive physics” resists physics instruction.

Another study conducted by Goldring and Osborne (1994) identified students’ confusion between energy and power. In this study, the researchers administered an energy questionnaire to 75 students taking advanced-level physics in six London Secondary Schools. (This level is roughly equivalent to first-year introductory physics in the US. See Adkins (1981), for example.) This questionnaire involves both qualitative and quantitative reasoning. In the questionnaire, one set of items relevant to college-level introductory physics requires of students the difference between energy and power. Two examples provided by the authors from this set read as following:

Two identical quantities of water were brought to the boil from room temperature in electrical kettles A and B. The water in kettle A started to boil ten minutes after the kettle was plugged in, and the water in B started to boil twenty minutes after it was plugged in. You may consider heat losses to be negligible. Were the amounts of energy supplied to the two kettles equal? Were the power ratings of the two kettles equal? Explain your answer!

---

<sup>3</sup> “Impetus” is a pure common sense concept. According to Jean Buridan, it is a certain immaterial motive power that sustains a body’s motion until it has been dissipated due to resistance by the medium. (Crombie, 1963) As opposed to “impulse”, there is no accepted scientific definition for “impetus”.

Two power stations, power station A and power station B, generate electrical energy. A operates at 800 MW and B at 500 MW. Assuming that A generated electrical energy for 1 hour, and B for 2 hours, which station generated more electrical energy in the period in which it was active? Tick the correct answer.

- (a) Power station A generated more than power station B.
- (b) Power station A generated less than power station B.
- (c) Both power stations generated the same amount of energy.
- (d) There is not enough information to decide.

Students' responses to this set of questions show that they often confuse energy with power. Unfortunately, the author didn't report the data of correct percentage or students' detailed replies to these questions. As to the first one of the above examples, the authors found that more than one-third of those students, who replied to the first two questions correctly, reasoned that the equality of energy and power was due to the equality of the voltage of the mains.

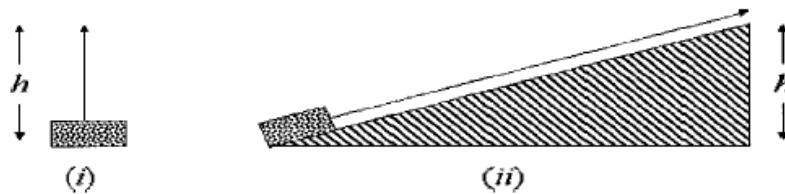
### 2.1.2 Students' difficulties with the concept of work

Work as an energy-related concept often brings students conceptual difficulties. Two studies conducted by Singh and Rosengrant (2003), and Loverude, Kautz and Heron (2001) suggest that students often have difficulties in (1) differentiating between work and force, (2) understanding the path independence of work done by the gravitational force, (3) determining the sign of work done on an object, and (4) relating work done *on* and *by* a system.

Singh and Rosengrant (2003) designed a 25-item multiple-choice test to investigate students' conceptual understanding of momentum and energy. They administered the test to students in both calculus- and algebra-based introductory physics courses in different colleges and universities, and conducted individual interviews to analyze students' reasoning.

In this test, 11 items are energy questions that do not involve momentum; among these there are 5 items intended to probe students' understanding of the work concept. For 2 of these 5 items, the authors provided relatively detailed analyses of or comments on the students' responses. One of these two items reads as follows:

You want to lift a heavy block through a height  $h$  by attaching a string of negligible mass to it and pulling so that it moves at a constant velocity. You have the choice of lifting it either by pulling the string vertically upward or along a frictionless inclined plane (see diagram). Which one of the following statements is true?



**Figure 2.2 A question from the multiple-choice test of energy and momentum concepts. (Singh and Rosengrant, 2003)**

- (a) The magnitude of the tension force in the string is smaller in case (i) than in case (ii)
- (b) The magnitude of the tension force in the string is the same in both cases
- (c) The work done on the block by the tension force is the same in both cases
- (d) The work done on the block by the tension force is smaller in case (ii) than in case (i)
- (e) The work done on the block by the gravitational force is smaller in case (ii) than in case (i)

Students' responses to this item indicate that many students have difficulty in differentiating between work and force. The overall result in the post-instruction test is that only 45% of the students chose the correct answer (c). The most popular alternative choice is (d), which accounts for 33% of the students. This alternative choice states that the work done along the inclined plane is less than the work done along the vertical path. Many students incorrectly reasoned that the smaller amount of work along the inclined plane is due to the smaller magnitude tension force. Student individual interviews confirmed this finding. An

interviewed student who chose (d) noted: *...you do a lot less work in the inclined case than in the vertical case... it is called mechanical advantage or something...* When asked how to calculate the work, he added: *...I don't know if I remember the math... the concept is clear to me though... you definitely do less work when it is inclined.* A second interviewed student explicitly stated that: *in inclined case it would be easier to pull which implies less work for you.* Another student also gave similar explanations: *tension in case (ii) is at an angle... say 30° so that work done will have cos(30°) and it will be smaller than straight up.*

The other item for which the authors provided comments on the students' responses is cited below.

You lift a suitcase from the floor to a table. In addition to the weight of the suitcase, select all of the following factors that determine the work done by the gravitational force on the suitcase.

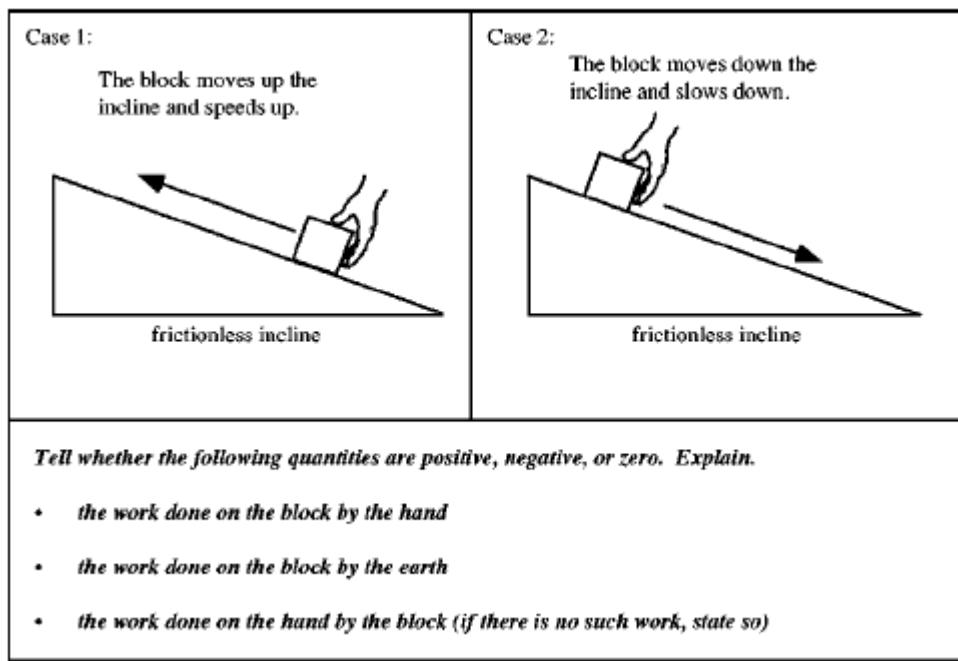
- (1) whether you lift it directly up to the table or along a longer path
  - (2) whether you lift it quickly or slowly
  - (3) the height of the table above the floor
- 
- (a) (1) only
  - (b) (3) only
  - (c) (1) and (3) only
  - (d) (2) and (3) only
  - (e) (1), (2), and (3).

Responses to this item revealed prevalent difficulties among students in understanding the work done by the gravitational force, particularly the path-independence of the work done by the gravitational force. In the post-instruction test, nearly half of the students (47%) chose (c), which states the work done by the gravitational force depends on both the height and the path. Another 13% of the students chose (a), which states the work done by the gravitational force solely depends on the path. Only a small percentage of the students (14%) chose the correct answer (b). The authors attribute this difficulty among

students to the confusion of physics and everyday definitions of work, because in everyday language doing more work means getting more tired.

Loverude, Kautz and Heron (2001) also investigated students' understanding of the concept of work in mechanics. To perform the investigation, the researchers designed some written questions about the work done on objects. They administered the questions to students as an ungraded quiz in introductory algebra-based physics and second-year thermal physics courses at the University of Washington (UW) and introductory calculus-based course at the University of Maryland, College Park (UMCP).

The questions involve a block being pushed by a hand and moving on an inclined plane. Two cases were designed. In both cases the force exerted on the block is upward along the incline. In case 1, the block is moving up the incline. In case 2, the block is moving down the incline. Some versions of the quiz also included a third case, in which the hand pushes down the incline as the block moves down.



**Figure 2.3 A mechanical problem. (Loverude, Kautz, and Heron, 2001)**

Students were asked to determine the sign of the work done (1) on the block by the hand, (2) on the block by the earth, and (3) on the hand by the block (or to state explicitly if there is no such work).

As one can see, in all cases the sign of the work done by the hand to the block can be determined by comparing the direction of the force and the direction of the displacement of the block. Taking into account Newton's third law, one can also determine the sign of the work done on the hand by the block. The sign of the work done on the block by the gravitational force can be found by comparing the direction of the gravitational force and the direction of the displacement of the block.

Students' responses to the first two questions reveal that students have difficulties in determining the sign of work done on a system. In case 1, for which the results were reported, the researchers found that many students who gave the correct answer provided wrong reasoning. One student wrote that the work done on the block by the hand is "positive, because work=(force)(distance) and since the block moves work is done." Many students also defined a coordinate system to aid the determination of the sign of work. Nonetheless, this scheme didn't help students resolve the trouble. One student wrote, "There really is no such thing as negative work since the distance used to calculate work depends on where a person chooses to place the coordinate axis." Another student wrote, "The work done [on the block by the hand] is positive, because I defined my coordinate system to be positive in the up direction."

Students' responses to the last question about the sign of work done on the hand by the block showed that students have difficulty in relating the work done on and by a system. In the UW algebra-based course, around 10% of the students answered zero work done on

the hand by the block. An additional 20% of the students in the course stated that there would be no such work. Other students indicated that the work done on the hand by the block was nonzero, but smaller in absolute value than the work done by the hand. Even those students who seemed to understand Newton's third law also claimed either no work done or zero work done on the hand by the block. One student from the UW algebra-based class wrote, “[There is] no such work, the block may have an equal and opposite force on the hand, but the block is not actually doing any work.”

### 2.1.3 Students' confusions among work, heat and internal energy

Work, heat and internal energy are three closely related quantities. All three quantities are scalar, share the same SI unit — Joules, and play equally important roles in the energy principle:  $\Delta E = W + Q$ . Nevertheless, the distinctions among the three concepts are essential.

Work and heat, unlike internal energy, “are *not* state variables or functions of state” (Arons, 1989). This means that work and heat cannot be employed to describe the conditions or the properties of a system. Rather, they are associated with two independent processes of energy transfer and represent “two ways in which the internal energy of a body may be changed...” (Warren, 1972). Work is a process of mechanical transfer of energy by applying a force over a certain displacement. Heat, on the other hand, is a process of thermal transfer of energy, and can be viewed as “microscopic work” done via interactions at the atomic level. (Chabay & Sherwood, 2007) Both processes can cause a change in internal energy of a system.

Temperature is another related concept that is associated with the average energy of a molecule. However, it is neither energy nor transfer of energy, and it doesn't appear

explicitly in the energy principle:  $\Delta E = W + Q$ . The deeper understanding of temperature lies in a statistically based definition of temperature and is more relevant to the topic of entropy. Therefore, this review specifically excludes the students' understanding of temperature.

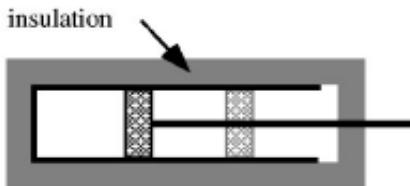
Due to the close relations and subtle distinctions of work, heat and internal energy, students often have difficulties in discriminating among these three concepts. Studies conducted by Loverude, Kautz and Heron (2002), and Meltzer (2004) indicate that students often use these three concepts interchangeably.

Loverude, Kautz and Heron designed interviews and written questions to investigate the students' ability to apply the work concept and the energy principle to an adiabatic compression of air.

Interviews were conducted among 36 students from algebra-based physics courses and a second-year thermal physics course at UW. During the interviews, the students were shown a plastic bicycle pump and were told that the open end of the pump would be sealed while the handle of the pump was rapidly pressed down. They were asked to predict what would happen to the temperature of the air inside the pump and to provide explanations.

A similar written question was also designed and administered to a large number of students from algebra-based physics courses and a second-year thermal physics course at UW and calculus-based physics courses at UMCP. These students took the written question either as a part of the course exam or as an ungraded quiz. There are three different versions of this written question. They read as below:

A cylindrical pump contains one mole of an ideal gas. The piston fits tightly so that no gas escapes, but friction is negligible between the piston and the cylinder walls.



a. Version A

The pump is thermally isolated from its surroundings. The piston is quickly pressed inward as shown.

*What will happen to the temperature of the gas? Explain your reasoning.*

b. Version B

The piston is quickly pressed inward as shown. A sensor in the pump records an increase in temperature.

*How can you account for the increase in internal energy of the gas? Explain your reasoning.*

An ideal gas is contained in a cylinder with a tightly-fitting piston. Several small masses are on the piston.

(Neglect friction between the piston and the cylinder walls.)

The cylinder is placed in an insulating jacket. A large number of masses are added to the piston.

*Tell whether the pressure, temperature, and volume of the gas will increase, decrease, or remain the same. Explain.*



**Figure 2.4 Versions A and B of the bicycle pump problem and the insulated cylinder problem. (Loverude, Kautz, and Heron, 2001)**

As one can see, in this question positive work is done on the system—air in the pump.

Since the push happens very rapidly, it is valid to assume that there is no heat transfer.

Therefore, by applying the energy principle (the first law of thermodynamics), one can

predict the internal energy of the air in the pump increases, which leads to an increase in temperature.

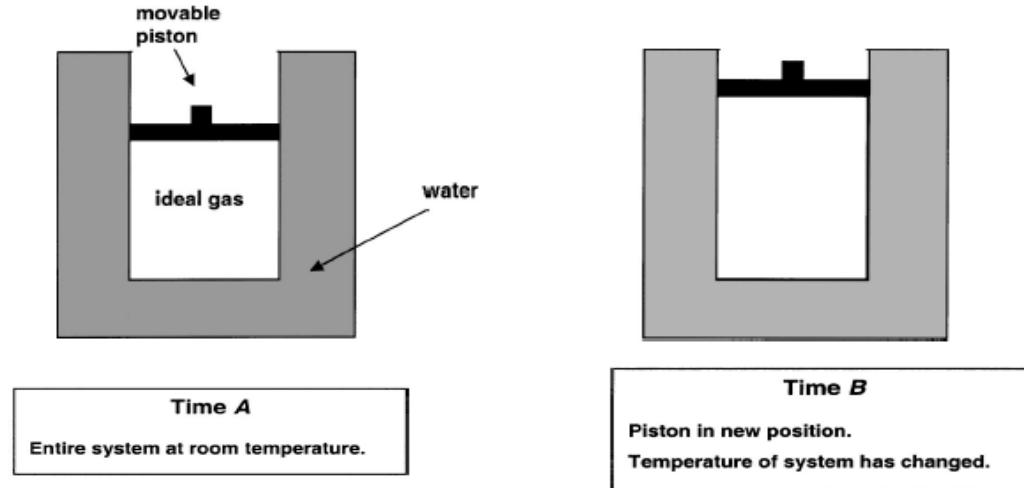
In order to correctly apply the energy principle to this case, students need to be able to distinguish among work, heat and internal energy. However, students' performance in both the interviews and the written question shows that they are unable to discriminate among work, heat and internal energy. They often confuse heat with internal energy, and use work and heat interchangeably.

One interviewed student explicitly expressed her confusion between heat and internal energy, "I am getting confused with how to apply the thing... First of all, if the heat of the system is the same as internal energy, or the temperature, or if it's related to Q." Results from the written question further revealed the prevalence of this confusion among students. In students' responses to the written questions, about half of the students explicitly stated that the insulation would either prevent a change in internal energy or in temperature.

Comparable to the confusion between heat and internal energy is the students' interchangeable use of heat and work. Several interviewed students used "heat" rather than "work" to refer to the process of the piston being pressed inward. In their responses to the written question, many students expressed similar ideas. For example, one student wrote: "The heat caused by pressing the piston inward is gained by the gas, so the temperature increased."

Meltzer reported a similar finding in an investigation of students' reasoning regarding heat, work, and the first law of thermodynamics. In his study, Meltzer designed several questions involving isobaric expansion of a system to elicit students' ideas about work. One question is cited below.

At initial time *A*, the gas, cylinder, and water have all been sitting in a room for a long period of time, and all of them are at room temperature. *Step 1.* We now begin Process #1: The water container is gradually heated, and the piston *very slowly* moves upward. At time *B* the heating of the water stops, and the piston stops moving when it is in the position shown in the diagram below:



**Figure 2.5 Isobaric expansion of a system. (Meltzer, 2004)**

During the process that occurs from time *A* to time *B*, which of the following is true:  
 (a) positive work is done *on* the gas *by* the environment, (b) positive work is done *by* the gas *on* the environment, (c) no *net* work is done on or by the gas.

Meltzer interviewed 32 students with this question. All of the interviewed students were from a second-semester calculus-based physics course at Iowa State University.

Despite the fact that this question does not directly assess students' ability to discriminate between work and heat, reasoning provided by the students indicates that students often have difficulties distinguishing between the two concepts. Many interviewed students used the word "work" when "heat" would have been appropriate. More specifically, many students used "work" to refer to a "heating process" in this isobaric expansion process.

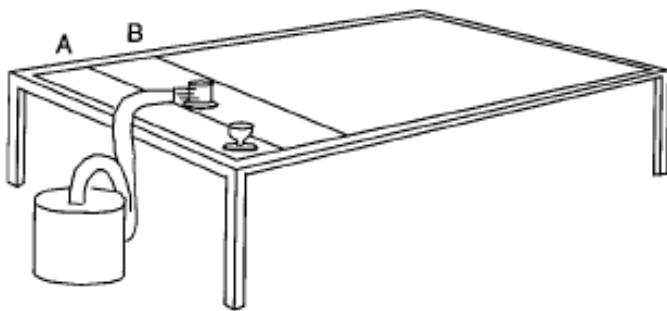
#### 2.1.4 Students' difficulties in applying the energy principle

Two studies conducted by Lawson and McDermott (1986), and Pride, Vokos and McDermott (1997) documented students' difficulties in applying the energy principle in dynamics problems.

Lawson and McDermott (1986) designed individual student interviews to probe students' understanding of the energy principle. The aspect of understanding emphasized in this study is the students' ability to apply the energy principle to the analysis of an actual motion observed from a demonstration.

The researchers interviewed 12 students from an honors calculus-based physics course and 16 students from an algebra-based physics course at UW. During each interview, a student was shown a demonstration of two pucks sliding on a low-friction surface under the same force over the same distance. The student was then asked to compare the final momentum and kinetic energy of the two pucks.

The two pucks in the demonstration are made of brass and plastic respectively, and differ greatly in mass. The brass puck is around ten times heavier than the plastic puck. Both pucks are initially at rest just behind line A, and move along a low-friction glass surface under the same constant force applied between lines A and B (See diagram). The constant force is supplied by a steady stream of air blown through the hose of a reversed vacuum cleaner. Small strips of paper are attached to the end of the hose. When blown by the air, the strips serve as a visual indication of a constant force by maintaining a constant separation between the hose and the puck.



**Figure 2.6** A two-puck demonstration. (Lawson and McDermott, 1987)

Correct comparison of the final kinetic energy of the two pucks requires only qualitative reasoning. Since the two pucks move without friction and do not rotate, the change of the kinetic energy, according to the energy principle, is equal to the external work done on the puck. Since the same constant force is applied on both pucks over the same distance, the change in kinetic energy of both pucks is the same. Considering that the initial kinetic energy of both pucks is zero, one can conclude that the final kinetic energy of both pucks is the same. The key aspect of the analysis is to apply the energy principle to relate the change in kinetic energy with the work done on the puck.

Students' performance on the task reveals their difficulties in applying the energy principle to the analysis of a simple motion. Particularly, none of the 16 students from the algebra-based physics class could apply the energy principle to make a correct comparison before the interviewer's intervention. Even after the interviewer drew the students' attention to the important conditions, such as the same force and distance, and directly asked if the term "work" represented ideas that could be applied to the demonstration at hand, students were still at a loss. Consequently, none of them could provide a satisfactory chain of reasoning to relate the change in kinetic energy with work done on the puck. Many students focused on the definition of kinetic energy and incorrectly argued that speed is a more

important variable than mass in comparing kinetic energy, because speed appears quadratically rather than linearly in the definition. Therefore, they concluded that the plastic puck with greater speed has more kinetic energy than the brass puck with less speed.

Though the 12 students from the honors calculus-based physics class performed better, half of them couldn't apply the energy principle to make a correct comparison without the interviewer's guidance. Similarly, the most common idea invoked was that larger speed means greater kinetic energy. In contrast to the students from the algebra-based physics class, the majority of the honors students were able to make a correct comparison and provide appropriate reasoning after the interviewer's guidance.

Based on this study, Pride, Vokos and McDermott (1997) further extended the scale of the investigation. The researchers presented the same comparison task in written form to 985 students from 11 calculus-based physics courses at UW. The live two-puck demonstration was shown. Students were then asked to make a comparison of the final kinetic energy of the two pucks and write down their reasoning.

Students' written responses showed that only 15% of the students were able to make a correct comparison based on the energy principle. Most students were unable to relate the result of a force acting over a distance to a change in kinetic energy. It seemed that the cause-effect relationship inherent in the energy principle did not occur to most students. As in the previous interview study, students often relied on the definition of kinetic energy to make a comparison and mistakenly reasoned that kinetic energy depends more on speed than mass.

The above two studies on the students' ability to apply the energy principle involve analyses of a rigid-body system that can be legitimately simplified as a point particle. There have been no formal studies reported on how students apply the energy principle in non-rigid

and/or rotating systems. However, careful analyses on the appropriate application of the energy principle suggest that the energy principle is often misused (even by experts) in non-rigid body systems. (For example, see Sherwood, 1983.) More specifically, an equation associated only with center-of-mass quantities is often mistakenly used as a real energy equation in a non-rigid system.

Here is an example provided by Sherwood and Bernard (1984). Suppose a block is pushed at a constant speed along a frictional surface. The applied force of magnitude  $f$  acting through a distance  $d$  does work  $fd$ . The frictional force is equal to  $f$  with opposite direction, since there is no acceleration. It would seem correct that the frictional force does work  $-fd$  on the block. Thus, the total work would be 0 and the energy equation would be  $\Delta(\frac{1}{2}mv^2) = 0 = fd - fd$ . Everything seems correct.

Yet there is an apparent paradox. Since the block gets warmer due to the friction, the increase in thermal energy of the block is nonzero. Where is the energy term for the increased internal energy of the block?

The key to the paradox lies in the difference between two similar equations. One equation relates the translational kinetic energy of the center of mass of an extended body to the net external applied force. This equation looks similar to the real energy equation. It, however, allows us to bypass internal forces, or vibrational, rotational and thermal energy. This equation has been entitled different names, such as “pseudowork-energy” equation and “center of mass equation”. (Penchina, 1976; Sherwood, 1982) As a matter of fact, this equation is the energy equation for a “point-particle system.” The “point-particle system” is defined as a simple point at the center of mass of the system, has the mass of the entire

system, and is subjected to all the forces acting on the system (Chabay & Sherwood, 2007).

The energy equation for the point-particle system is simple: the net force does work on the point-particle system and results in a change in translational kinetic energy. In the above

example,  $\Delta(\frac{1}{2}mv^2) = 0 = fd - fd$  is the energy equation for the point-particle system.

The other equation is the energy equation for the real system. It includes translational, vibrational, and rotational kinetic energy, potential energy, thermal energy and other forms of energy transfer such as heat. The work done on the real system is calculated as the sum of the actual work done by each individual force. In the above example, the energy equation for the real system (block) is:  $\Delta(\frac{1}{2}mv^2) + \Delta E_{\text{thermal}} = fd + W_{\text{friction}}$ . The increase in thermal energy of the block is included. In addition, the absolute value of the negative work done by the frictional force is less than  $fd$ . The reason lies in the microscopic mechanism of friction, particularly the deformability of the contact teeth between the block and the surface (Sherwood & Bernard, 1984). As a result, the displacement of application points of the frictional force actually is less than  $d$ .

For more examples of a general class of situations involving non-rigid systems, refer to Sherwood (1983), Arons (1989, 1999), Leff and Mallinckrodt (1991), Mallinckrodt and Leff (1993, 2002), and Mungan (2005, 2007a, 2007b).

## 2.2 Energy in the *Matter & Interactions Modern Mechanics* course

*Matter & Interactions* (Chabay & Sherwood, 2002b, 2007) is an innovative calculus-based introductory physics curriculum for engineering and science students at the college

level. This curriculum emphasizes a small number of fundamental principles and focuses on the atomic models of matter and its interaction. (Chabay and Sherwood, 1999; 2004b) In the *Modern Mechanics* course, the first semester of a two-semester sequence, the energy concept and its associated fundamental principle (the energy principle) stand out as one of the central topics. In order to introduce students to a scientific view of energy and its place in the physics enterprise, some unique approaches to this topic are designed. These unique approaches include the sequence, content and emphasis of the energy topic.

In contrast to the late appearance of energy in traditional calculus-based introductory mechanics courses, the energy topic is introduced rather early in the *M&I Modern Mechanics* course. In this course, momentum and the momentum principle<sup>4</sup> are introduced as the first central topic at the beginning of the course. Immediately after students have adequate practice on the momentum principle in various situations that involve gravitational and ball-spring-model interatomic interactions, energy and the energy principle enter the course as the second central topic. Four major aspects about energy are covered in the following sequence:

- The concept of energy and the energy principle
- Importance of the choice of system, and its effect on the definition of potential energy
- Quantized energy levels
- The calculation of work done on multiparticle systems

Energy of a particle (particle energy) is introduced in its relativistic form:

$$E = \frac{mc^2}{\sqrt{1 - |\vec{v}|^2/c^2}}, \text{ and is separated into rest energy } mc^2 \text{ and kinetic energy}$$

---

<sup>4</sup> The momentum principle is also called “the impulse-momentum” theorem. For the purpose of consistency, the name of “the momentum principle” is used thereafter.

$\frac{mc^2}{\sqrt{1 - |\vec{v}|^2/c^2}} - mc^2$ . The relativistic form presents an authentic expression of energy.

Moreover, the separation of rest energy and kinetic energy provides a good opportunity to discuss low-speed approximation to the kinetic energy  $\frac{1}{2}mv^2$  and range of validity. (Chabay and Sherwood, 2004a)

At the same time, the energy principle for a particle is introduced:  $\Delta E = W$ , followed by the energy principle for macroscopic systems:  $\Delta E_{\text{system}} = W + Q$ . In the analysis of macroscopic multiparticle systems, both thermal energy and heat are included to be consistent with thermodynamics; thereby it presents students an authentic picture of a unified physics enterprise. (Chabay and Sherwood, 1999) Moreover, the distinction between work  $W$  and heat  $Q$  is explicitly made. Work is defined as mechanical transfer of energy, while heat is defined as thermal transfer of energy due to a temperature difference and can be regarded as “microscopic work” at the atomic level.

In order to be able to apply the energy principle correctly in macroscopic systems, a careful specification of a system and its surroundings is particularly emphasized in the *M&I Modern Mechanics* course. Students are explicitly shown that different specifications of system will result in different energy accounting. (Sherwood and Chabay, 2004) One example given in the textbook is a ball of mass  $m$  falling by a distance  $h$  on an airless Earth. Two different systems can be chosen for analysis. If a chosen system includes both the ball and the Earth, there will be no external work done on the system. The energy equation can be expressed as:

$$\Delta K + \Delta U = \Delta K + (-mgh) = 0$$

Here, the energy terms that appear to the left of the equal sign include both the kinetic energy of the ball (assuming the kinetic energy of the Earth doesn't change) and the potential energy of the system consisting of the ball and the Earth.

If the falling ball itself is chosen to be the system, the Earth does external work on the ball, and the kinetic energy of the ball increases. Thus, the energy equation can be expressed as:

$$\Delta K = W = +mgh$$

In this case, the only energy term to the left of the equal sign is the kinetic energy of the falling ball. Note that there is no potential energy involved. However, students who study the standard approach may double count  $mgh$  by including both the work and potential energy terms, thus adding  $mgh$  twice.

This mistake lies in the incorrect definition of potential energy. Often, potential energy is mistakenly associated with one particle or one single object. In the *M&I Modern Mechanics* course, potential energy is carefully defined as energy associated with pair-wise interactions of particles internal to a system. Therefore, one particle cannot have any kind of potential energy. Similarly, in the above example, the falling ball does not have gravitational potential energy, if the ball itself is chosen to be the system of interest.

In order to present to students the discrete nature of energy in atomic systems, the most immediately accessible aspects of quantum mechanics are introduced in the *M&I Modern Mechanics* course. These aspects include quantized energy levels and transitions between energy levels associated with photon emission or absorption. The discrete electronic energy levels are familiar to students from their previous chemistry courses. Based on this background, the quantization of vibrational energy in the harmonic oscillator is discussed.

Working on systems such as hydrogen atoms and diatomic molecules, students are able to relate physics with chemistry and link the nature of energy at the macroscopic level to the behavior of energy at the atomic level. (Chabay and Sherwood, 1999; 2004a; 2007)

In multiparticle systems, particularly in non-rigid systems, the calculation of work done on the systems is carefully defined in the *M&I Modern Mechanics* course. Students are instructed to identify the point of application of each force and use the displacement of that point to calculate the work done by that force. In addition, kinetic energy in the multiparticle systems is expressed as translational kinetic energy (associated with the center of mass) plus kinetic energy relative to the center of mass.

To sustain a careful and precise treatment of energy in multiparticle systems, particularly to resolve the common confusion between the “pseudowork equation” and the real energy equation, the *M&I Modern Mechanics* course introduces a new concept of “point-particle system.” The “point-particle system” is defined as a system that has the total mass of the real system with all of this mass concentrated at the location of the system’s center of mass, and is subjected at that point to all of the forces that are exerted at various locations on the real system. Working with both the point-particle system and the real system, students write out two energy equations for two different systems rather than two similar equations (the “pseudowork” equation and the energy equation) for one system. Therefore, the effortful attempts to distinguish two similar mathematical expressions can be facilitated by “an easy visual distinction between two different and different-looking systems”. (Chabay and Sherwood, 1999; Sherwood and Chabay, 1999) For the point-particle system the energy equation is fairly simple: the net force does work on the point particle, which results in a change of translational kinetic energy. In contrast, the energy equation for the real system

includes translational, vibrational, and rotational kinetic energy, potential energy, and other forms of energy transfer such as heat. The work done on the real system is calculated as sum of the actual work done by each individual force.

## **2.3 Assessment instruments in physics education research**

### 2.3.1 Validity and reliability

Validity and reliability are two issues that need to be addressed in the construction of a standardized assessment instrument. Validity is generally referred to as the extent to which a test actually measures what it ought to measure. There is no statistical measurement to assess the validity of a test. Rather, it can be established through expert consultation. In contrast, reliability, which is referred to as the internal consistency of a test, can be statistically assessed. In principle, the consistency of a test is largely dependent on the difficulty levels and the discriminatory power of the test items. Five indices (coefficients) are often used to evaluate the difficulty, discrimination, and consistency of a test. These five indices (coefficients) include item difficulty index, item discriminatory index, point bi-serial coefficient, Kuder-Richardson reliability index, and Ferguson's delta.

Validity is a broad topic that encompasses several aspects: face validity, content validity, construct validity, and criterion-related validity. In general, a valid test should measure skills and knowledge that are directly relevant to the stated domain of the test.

Face validity is the apparent appropriateness of a test, and can be determined by a surface level, common sense reading of an instrument. (Kline, 1986) A test would lack face validity if the concepts tested were not related to the subject matter. This usually is not an issue, unless a test is noticeably erratic and irrelevant.

Content validity reflects the coverage of the subject matter in a test. In other words, does a test cover sufficient materials of a specific knowledge domain that it aims to test (Aubrecht II & Aubrecht, 1983)? For example, after a careful examination of several commonly used test banks and introductory physics books, informal interviews and contacts with science teachers, and a pilot study, the author of the Test of Understanding Graphs in Kinematics (TUG-K) included seven topics in the test that sufficiently cover the specific knowledge domain. (See Table 2.1.) As such, the content validity of TUG-K is established.

**Table 2.1 Seven topics covered in the TUG-K. (Beichner, 1994)**

<b>Given</b>	<b>The student will</b>
1. Position-Time Graph	Determine Velocity
2. Velocity-Time Graph	Determine Acceleration
3. Velocity-Time Graph	Determine Displacement
4. Acceleration-Time Graph	Determine Change in Velocity
5. A Kinematics Graph	Select Another Corresponding Graph
6. A Kinematics Graph	Select Textual Description
7. Textual Motion Description	Select Corresponding Graph

Both face validity and content validity can be established, during the designing process, by consulting experts. In physics education research, both types of validity of an instrument can be assessed by physicists' consensus, as was done with many standardized instruments, such as FCI (Hestenes, Wells & Swackhamer, 1992), MBT (Hestenes & Wells, 1992), TUG-K (Beichner, 1994), FMCE (Thornton & Sokoloff, 1998), BEMA (Chabay & Sherwood, 1997; Ding, Chabay, Sherwood & Beichner, 2006), CSEM (Maloney, O'Kuma, Hieggelke & Heuvelen, 2000) and DIRECT (Engelhardt & Beichner, 2001).

Other aspects of validity, including construct validity and criterion-related validity, are not as relevant in physics education research as in psychometric research. Construct

validity refers to the extent to which a test measures a theoretical construct or trait such as creativity, honesty, or intelligence (Kline, 1986). Criterion-related validity refers to the evidence that performance on one assessment instrument can be used to make inferences about performance in a different domain (Kline, 1986).

As opposed to validity, reliability can be statistically assessed. In general, test reliability is referred to as the internal consistency of a test, and is largely affected by the difficulty levels and the discriminatory power of the test items. Five indices (coefficients) are often employed to assess the difficulty, discriminatory power and consistency of a test. These indices (coefficients) include three measures that assess individual test items (item difficulty index, item discriminatory index, point bi-serial coefficient) and two measures that assess the test as a whole (Kuder-Richardson reliability index and Ferguson's delta).

Item difficulty index (P) is a measure of the difficulty of a single test item. It reflects how easy (or difficult) an individual item is. Acceptable item difficulty index values range from 0.3 to 0.9 (Doran, 1980).

Item discrimination index (D) is a measure of the discriminatory power of a test item. In other words, it measures the extent to which a single test item distinguishes students who know the material well from those who do not. The acceptable value for item discriminatory power is typically greater than 0.3 (Doran, 1980).

The point bi-serial coefficient (sometimes referred as the item reliability) is a measure of the consistency of a single test item with the whole test. It reflects the correlation between an individual item and the entire test. The acceptable value for point bi-serial coefficient is greater than 0.2 (Kline, 1986).

Kuder-Richardson test reliability index (KR-20) is a measure of the internal consistency of a test. If a test is reliable, one can have confidence that the same students would get the same score if they took the test more than once. Moreover, large fraction of the variance of test scores on a reliable test is due to a systematic variation in the population of test takers (Ghiselli, Campbell & Zedeck, 1981). A widely accepted KR-20 index value for a reliable test is greater than 0.7 (Doran, 1980).

Ferguson's delta is a measure of the discriminatory power of an entire test. It reflects how broadly the total test scores are distributed over the possible range. If a test is designed to discriminate among students, one would like to see a broad distribution of total scores. The accepted value for Ferguson's delta is greater than 0.9. (Kline, 1986)

The following table collects the published data on the statistical indices (coefficients) for some assessment instruments in physics education research.

**Table 2.2 Statistical evaluations of some standardized physics tests.**

Test	Average Item Difficulty Index	Average Item Discriminatory Index	Average Point Bi-serial Coefficient	KR-20 Reliability Index	Ferguson's delta
<b>MBT</b>	N/A	N/A	N/A	N/A	N/A
<b>FCI</b>	N/A	N/A	N/A	N/A	N/A
<b>FMCE</b>	N/A	N/A	N/A	N/A	N/A
<b>TUG-K</b>	0.40	0.36	0.74	0.83	0.98
<b>BEMA</b>	0.42	0.33	0.43	0.85	0.98
<b>CSEM</b>	[0.1, 0.8] (No average reported)	[0.1, 0.55] (No average reported)	N/A	0.75	N/A
<b>DIRECT (Version 1.1)</b>	0.41	0.23	0.32	0.70	N/A
<b>Energy and Momentum</b>	[0.2, 0.8] (No average reported)	N/A	[0.21, 0.48] (No average reported)	0.75 (Calculus-based) 0.68 (Algebra-based)	N/A
<b>Energy Concept Inventory</b>	N/A	N/A	N/A	N/A	N/A

### 2.3.2 Using a multiple-choice test

One major concern among physics instructors, when using multiple-choice tests, is that the tests can encourage guessing and cannot truly reflect students' thoughts. (Bork, 1984, Varney, 1984, Sandin, 1985, Scott, 1985)

A study conducted by Henderson (2002) shows that random guessing can be identified through checking students' answer patterns. At the University of Minnesota

(UMN), the Force Concept Inventory (FCI) was administered to more than 3500 students in calculus-based physics courses between 1997 and 1999. Henderson examined the students' responses to the FCI, and found several types of answer patterns that may indicate a student is not taking the test seriously. These types of patterns are (1) drawing a picture on the Scantron answer sheet; (2) answering all A's, B's, etc.; (3) leaving six or more blanks; and (4) other patterns (such as ABCDE, EDCBA, AAAAA, BBBBB, etc.) any place in their responses. For the analysis of the test results, individuals showing the guessing patterns can be excluded from the student population to reduce the effect of random guessing. As such, the test results can more closely reflect the true performance of those students who take the test seriously.

Another technique that has been widely used to reduce guessing and probe students' real thoughts on test items is student interviews. This technique is particularly useful in detecting false positives (correct answers based on wrong reasoning) in multiple-choice tests. In physics education research, many standardized assessment instruments are multiple-choice tests and have been accepted as powerful diagnostic tools when combined with small-scale student interviews.

Though the usefulness of multiple-choice tests in physics education research has been well established, how to interpret test results through advanced statistical analyses, such as factor analysis, still remains debatable. Discussions on the FCI by Heller et al. (Huffman & Heller, 1995; Heller & Huffman, 1995) and Hestenes et al. (Hestenes & Halloun, 1995; Halloun & Hestenes, 1995) suggest that interpretation of multiple-choice tests through factor analysis requires great caution.

FCI is a multiple-choice test designed by Hestenes, Wells and Swackhamer (1992) to assess students' Newtonian and non-Newtonian conceptions of force. According to the authors, the questions on the FCI can be grouped into the six conceptual dimensions of kinematics, Newton's first law, Newton's second law, Newton's third law, superposition principle, and kinds of forces.

Factor analysis is an advanced statistical technique used to determine how the items on a test are related through analyzing the correlations, variance and covariance among all items. Items that measure the same concept or idea will be grouped together. This concept or idea is called a "factor" or a "dimension"<sup>5</sup> (Kim & Mueller 1978).

Huffman and Heller (1995) administered the FCI to 145 high-school students in physics classes and 750 university students in calculus-based physics classes. After performing a factor analysis on FCI scores, Huffman and Heller found that responses from high school students yielded only two factors and responses from university students yielded only one factor. In addition, none of these factors completely matched any of the six dimensions proposed by the FCI authors. In conclusion, Huffman and Heller stated that the inventory "measures bits and pieces of students' knowledge that do not necessarily form a coherent concept." They also called for cautions for using the FCI.

In response to Huffman and Heller, Hestenes and Halloun (1995) contended that the proposed six conceptual dimensions were not meant to describe student concepts but to provide a Newtonian standard against which student concepts could be compared. They also

---

<sup>5</sup> As a matter of fact, many physics topics are related. Testing one can inevitably involve another. For example, to be able to answer a question about acceleration, one must presumably understand velocity. Therefore, acceleration and velocity may not impose two independent factors or dimensions. Performing factor analysis via unorthogonal method can resolve this issue. However, no relevant discussions have been published.

stated that the clustering of students' responses on only one or two factors "justifies interpreting the FCI score as a measure of the disparity between student concepts and the Newtonian force concept" (Hestenes & Halloun, 1995).

The discussions continued and became intense. In the end, both sides reached the same essential point that the factor analysis of students' responses to the FCI reveals the force concept only *from the students' point of view*. Hence the factor analysis results do not address the validity issue of the FCI and should be interpreted with caution. (Heller & Huffman, 1995; Hestenes & Halloun, 1995)

### 2.3.3 Assessment instruments on energy

In physics education research, a few assessment instruments on energy have been designed. A multiple-choice test of energy and momentum concepts by Singh and Rosengrant (2002) and the Energy Concept Inventory (ECI) by Swackhamer and Hestenes (unpublished) are two major tests specifically designed to probe students' understanding of energy and energy related concepts. The Mechanics Baseline Test (MBT) by Hestenes and Wells (1992) is another test that covers the energy concept, in which a few items aim to assess students' understanding of the conservation of energy and the energy principle.

A multiple-choice test of energy and momentum concepts by Singh and Rosengrant (2002) is a 25-item research-based conceptual test. This test is designed for university students in either calculus-based or algebra-based physics classes, and is mainly suitable for traditional mechanics courses. In this test, 11 items are energy questions that do not involve momentum. These questions cover kinetic energy, gravitational potential energy, work, the conservation of mechanical energy, and the energy principle. According to the authors, when

combined with student interviews, this test is useful in detecting students' conceptual difficulties with energy and momentum concepts.

However, several energy items in the test reflect some fundamental errors in the treatment of energy in standard physics textbooks, particularly in the discussion of potential energy and work done by frictional forces.

One item about the conservation of mechanical energy is quoted and analyzed below. It reflects the common misconception about potential energy being associated with a single object.

Three bicycles approach a hill as described below:

- (1) Cyclist 1 stops pedaling at the bottom of the hill, and her bicycle coasts up the hill.
- (2) Cyclist 2 pedals so that her bicycle goes up the hill at a constant speed.
- (3) Cyclist 3 pedals harder, so that her bicycle accelerates up the hill.

Ignoring the retarding effects of friction, select all the cases in which the total mechanical energy of the cyclist and bicycle is conserved.

- (a) (1) only
- (b) (2) only
- (c) (1) and (2) only
- (d) (2) and (3) only
- (e) (1), (2), and (3)

According to the authors, the correct answer is (a); the mechanical energy of the cyclist would be conserved if she stopped pedaling. Note that the question explicitly asks about the total mechanical energy “of the cyclist and bicycle”. Hence, the chosen system for consideration should be “the cyclist and bicycle”. This means the mechanical energy of “the cyclist and bicycle” only involves kinetic energy, and does not include the gravitational potential energy<sup>6</sup>. If “the cyclist and bicycle” maintains the same speed, then the kinetic

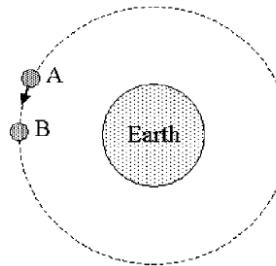
---

<sup>6</sup> The gravitational potential energy of “the cyclist and bicycle” system, due to the interactions between the cyclist and bicycle, is negligibly small.

energy of the system will be constant<sup>7</sup>. Therefore, the correct answer is (b), not (a). The authors who choose (a) for the correct answer obviously take into consideration the gravitational potential energy associated with the interactions between the earth and “the cyclist and bicycle”. As we know, within “the cyclist and bicycle” system, there are no interactions between the earth and “the cyclist and bicycle”. Consequently, considering the gravitational potential energy associated with the interactions between the earth and “the cyclist and bicycle” is meaningless.

A second controversial example that reflects the same common misconception is a question about a satellite moving around the earth. It reads as follows.

A satellite is moving around Earth in a circular orbit at a constant speed (see diagram). The only force that acts on the satellite is Earth’s gravitational force which points directly toward earth’s center. Which of the following statements is true as the satellite moves from point A to point B in the orbit?



**Figure 2.7 A question from the multiple-choice test of energy and momentum concepts. (Singh and Rosengrant, 2003)**

- (a) The gravitational potential energy of the satellite decreases as it moves from A to B.
- (b) The work done on the satellite by the gravitational force is negative for the motion from A to B.
- (c) The work done on the satellite by the gravitational force is zero for the motion from A to B.
- (d) The velocity of the satellite remains unchanged as it moves from A to B.
- (e) None of the above.

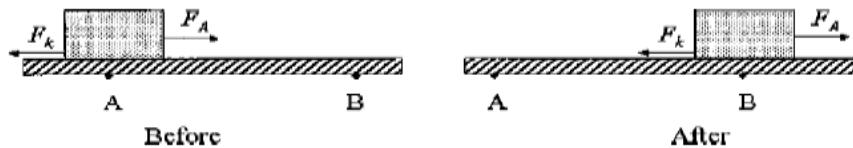
---

<sup>7</sup> Precisely speaking, the word “conservation” is reserved for quantities that are constant for a system and its surroundings. For isolated systems we say that the energy of the system is a “constant”, and the energy is called “a constant of the motion”. (Bauman, 1992)

Although choice (a) is not the correct answer, the flawed statement about “the gravitational potential energy of the satellite” reflects the misunderstanding of potential energy to be associated with one single object.

Another problematic item, which involves the work done by frictional forces, is cited and analyzed below.

Using a rope of negligible mass, you pull a box along a horizontal surface with a constant horizontal force  $F_A$ . The box moves at a constant velocity from position A to position B. The force of friction  $F_k$  cannot be neglected. Which one of the following statements concerning the motion of the box from A to B is true?



**Figure 2.8 A question from the multiple-choice test of energy and momentum concepts. (Singh and Rosengrant, 2003)**

- (a) The work done on the box by the gravitational force is non-zero
- (b) The work done on the box by  $F_k$  is positive.
- (c) The total work done on the box by the net force is nonzero
- (d) The magnitude of the work done on the box by  $F_A$  is equal to the magnitude of the work done by  $F_k$ .
- (f) The magnitude of  $F_A$  is greater than the magnitude of  $F_k$ .

This question actually is more complicated than it seems to be. It involves a microscopic view of friction and the work done on a non-rigid-body system. As analyzed by Sherwood and Bernard (1984), the magnitude of work done on the box by the frictional force  $F_k$  is less than the magnitude of work done by the force  $F_A$ ; the total work done on the block is positive. This is why the internal energy of the block increases. So, the correct answer to this item should be (c), although a more careful wording would be “The total work done on the box by the two forces is nonzero.” Nonetheless, the authors considered (d) being the correct answer. Clearly, they only considered the motion of the center of mass of the box and

oversimplified the box as a point particle. Since friction involves an interaction between deformable surfaces, treating objects subject to friction as point particles with no internal degrees of freedom is inherently contradictory, and can only lead to paradoxical results.

The Energy Concept Inventory (ECI) developed by Swackhamer and Hestenes (unpublished) is another conceptual instrument intended to assess students' qualitative understanding of energy concepts. It is aimed to discriminate scientific views of energy from the common sense alternatives. (Swackhamer & Dukerich, 2003) The main aspects of energy tested in the ECI include energy storage and energy transfer, which are essential to chemistry and biology as well as physics. In order to be useful in comparative studies of students' knowledge and learning in different grade levels, two versions of ECI have been designed; one version with 25 items is suitable for 8th & 9th grade science students, and the other with 35 items for 10th, 11th, & 12th grade science students.<sup>8</sup> The version for 10<sup>th</sup> to 12<sup>th</sup> grade levels covers additional topics, such as mass-energy conversion in atomic reactions and graphical representation of energy. Though the ECI can be used as a diagnostic test for different grades at the pre-college level, it may not be suitable for certain physics course evaluations at higher levels, such as calculus-based mechanics courses at the university level, which includes more topics than just energy conservation.

Two examples of the ECI are presented here. The first item appears in both versions of ECI. The second item appears only in the version for 10<sup>th</sup>-12<sup>th</sup> grade science students. As one can see, these items are aimed at the most basic everyday phenomena that are staples of school science instruction about energy.

---

<sup>8</sup> See the "Modeling Instruction Program" website at the Arizona State University:  
<http://modeling.asu.edu/MNS/MNS.html>.

When a candle burns the energy released

- comes mainly from the wax and air
- comes mainly from the burning wick
- is produced mainly by the fire
- comes mainly from the wax

A book slides across a table and comes to rest. Which of the following graphs best describes the total energy of the book, table, and surroundings during this process?

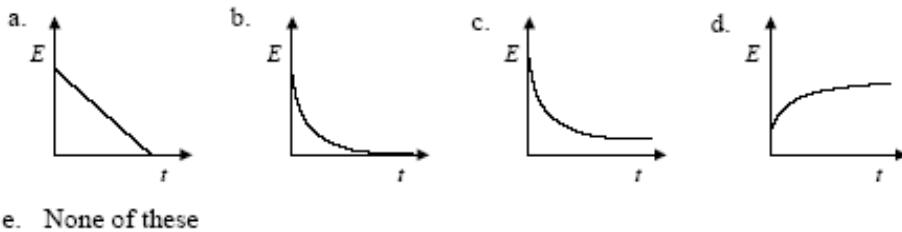


Figure 2.9 Choices for a question in the ECI. (Hestenes and Swackhamer, unpublished)

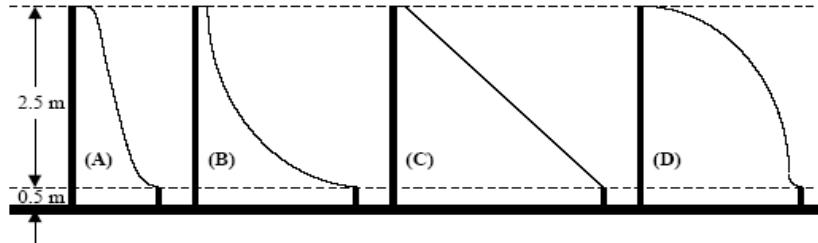
The Mechanics Baseline Test (MBT) designed by Hestenes and Wells (1992) is a 26-item multiple-choice test aimed to assess students' understanding of the most basic concepts in mechanics. The topics covered in the test include kinematics, Newton's laws, the energy principle, the momentum principle, conservation of energy, conservation of momentum, and kinds of forces. In contrast to the FCI, where questions are designed to be meaningful to students without formal training in mechanics, the MBT emphasizes concepts that cannot be grasped without formal knowledge about mechanics (Hestenes & Well, 1992). In addition, many items in the MBT are quantitative questions that require simple calculations. Therefore, the MBT is mostly appropriate for post-instruction use. As is intended by the authors, the MBT is universal in the sense that all the tested concepts should be addressed in any introductory physics class. (Hestens & Well, 1992)

In the MBT, only 3 items are energy questions; two are related to the conservation of energy, and the other one is related to the energy principle. The item that involves the energy

principle is borrowed from the two-puck research conducted by McDermott and Lawson (1987).

One example that involves the conservation of energy is cited below.

A young girl wishes to select one of the frictionless playground slides illustrated below to give her the greatest possible speed when she reaches the bottom of the slide. Which of the slides illustrated in the diagram above should she choose?

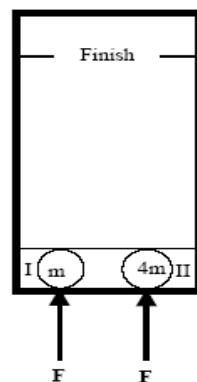


**Figure 2.10 A question from MBT. (Hestenes and Wells, 1992)**

- (A) A
- (B) B
- (C) C
- (D) D
- (E) It doesn't matter; her speed would be the same for each.

The multiple-choice version of two-puck question in the MBT reads as follows.

The diagram depicts two pucks on a frictionless table. Puck II is four times as massive as puck I. Starting from rest, the pucks are pushed across the table by two equal forces. Which puck will have the greater kinetic energy upon reaching the finish line?



**Figure 2.11 A two-puck question from MBT. (Hestenes and Wells, 1992)**

- (A) I
- (B) II
- (C) They both have the same amount.
- (D) Too little information to answer.

As one can see, the above reviewed tests are either problematic, or inadequate for higher-level use, or insufficient as a complete assessment instrument on the energy topic. In order to better evaluate students' understanding of the energy topic, particularly in the calculus-based introductory mechanics courses at the college-level, research in developing an appropriate assessment instrument is worth conducting.

As a part of an innovative calculus-based introductory physics curriculum, the *M&I Modern Mechanics* course presents to students a scientific view of energy through some unique approaches including the sequence, content and emphasis. Evaluations of students' understanding of energy in the *M&I Modern Mechanics* course can assess the effectiveness of these unique approaches; thus providing valuable feedback on the course development. To that end, an assessment instrument appropriate for the *M&I Modern Mechanics* course is desirable.

## **CHAPTER 3: Methodology**

The purpose of the study is to evaluate student understanding of the energy topics that are covered in the *Matter & Interactions* modern mechanics (Chabay & Sherwood, 2006) course through a new energy assessment. In order to fulfill this purpose, it is essential that a valid and reliable energy assessment be developed to match the goal of the *M&I* mechanics course. This chapter details the procedures of how the energy assessment was designed, how its validity was established, and how its reliability and discriminatory power was tested. In addition, this chapter describes the academic background of the student participants who took different versions of the energy assessment, as well as the detailed test administration settings.

Further, this study employs both quantitative and qualitative research methods to evaluate student understanding of energy. In an attempt to evaluate a large number of students, quantitative analyses were performed using the data from the energy assessment administered either on paper or computer. To explore student reasoning qualitatively, a small number of student interviews were conducted to uncover students' explanations for their answers. This chapter outlines both of the two methods and their places in this study.

### **3.1 Development of the energy assessment**

#### 3.1.1 Test Objectives

The development of the energy assessment started with the construction of test objectives. These test objectives were derived directly from the energy topics that are covered in the *M&I* mechanics course, including those that are unique to the course.

Figure 3.1 displays the hierarchical structure of the energy topics in the *M&I* mechanics course, which also serves as a blueprint for the test objectives.

In Figure 3.1 there are five major components that represent the main topics on energy in the *M&I* mechanics course: the energy principle, forms of energy, appropriate systems, means to change the energy of a system, and the discrete nature of energy at the atomic level. Accordingly, there are five major objectives to address these topics respectively: application of the energy principle, identification and determination of different forms of energy, specification of appropriate systems, determination of and differentiation between  $W$  and  $Q$ , and calculation of absorption/emission spectrum. Among these five objectives, application of the energy principle plays the fundamental role; therefore taking the top position in the hierarchical structure. For some other objectives, there exist lower level sub-objectives. For instance, under the objective “identification and determination of different forms of energy”, there are three sub-objectives, each of which pertains to one of the three basic forms of energy—rest energy, kinetic energy ( $K$ ) and potential energy ( $U$ ).

The rationale of the test objectives originated from one of the goals for the *M&I* mechanics course, that is, students are able to apply the energy principle ( $\Delta E_{\text{sys}} = W_{\text{ext}} + Q$ ) to describe and predict a wide range of physical phenomena. In order to achieve this goal, students need to know how to specify a system, what kinds of energy are included in the system ( $E_{\text{sys}}$ ), and what causes the change in the total energy of a system ( $W_{\text{ext}}$  and  $Q$ ). Additionally, the discrete nature of energy at the atomic level also plays an important role in the *M&I* mechanics course. Particularly, in the context of absorption or emission spectra, calculating the differences between the discrete energy levels is just a special case of

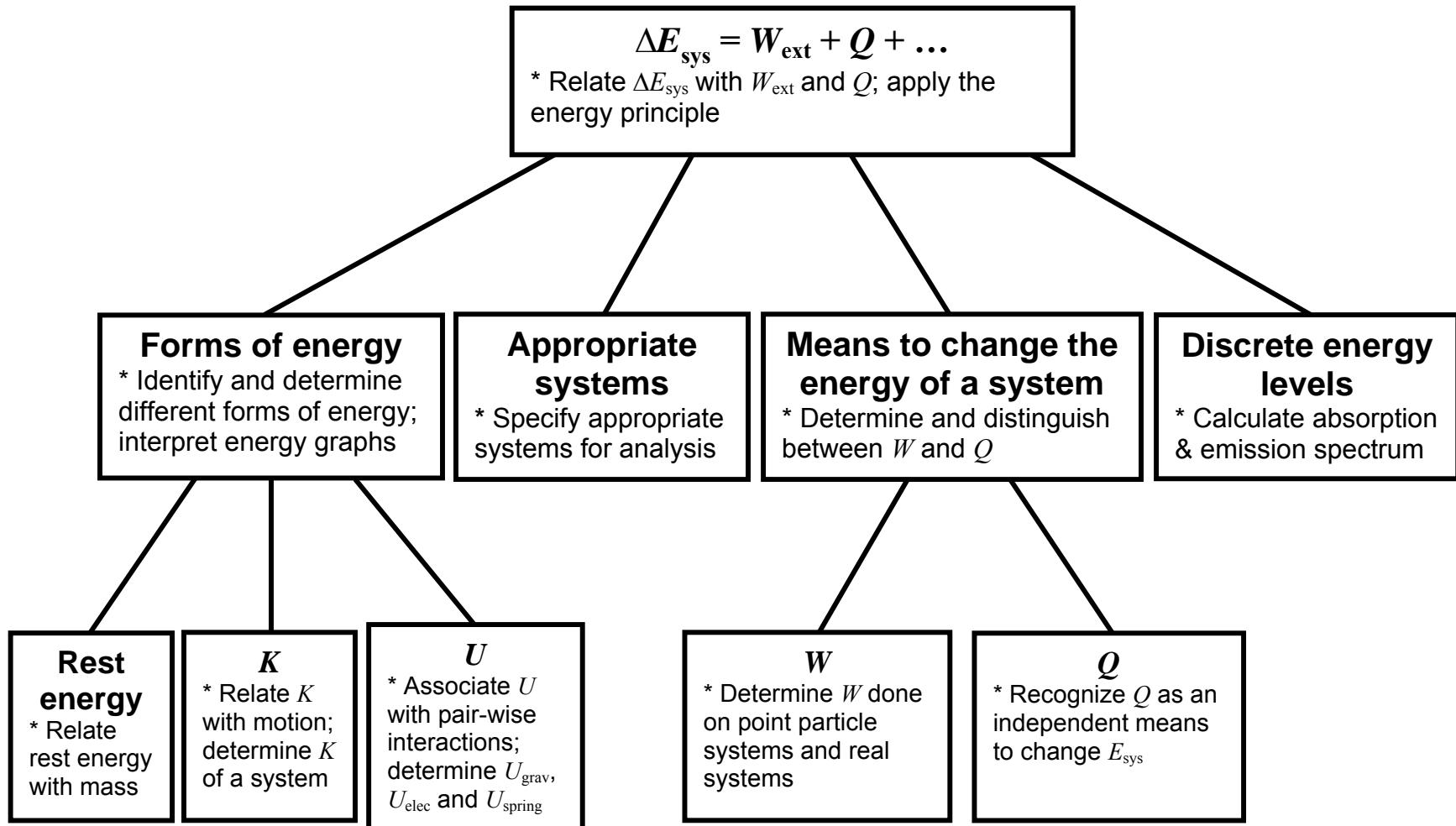


Figure 3.1 Hierarchical structure of the test objectives.

applying the energy principle. Therefore, the objective "calculation of absorption/emission spectrum" is classified under the top objective—application of energy principle—as well.

In constructing Figure 3.1, a panel of four experts participated in many discussions on the test topics covered in the test objectives. These four experts consisted of the two authors of the *M&I* curriculum and two physicists with many years of experience teaching the *M&I* courses in the SCALE-UP (Student Centered Activities in Large Enrollment University Physics; Beichner *et al.* 2000) classroom settings. (See Table 3.1.) The current version of test objectives resulted from several revisions based on the feedback provided by the panel of experts.

**Table 3.1 Four panel experts.**

<b><u>Panel Expert</u></b>	<b><u>Affiliation</u></b>	<b><u>Background</u></b>
Beichner, Robert	NCSU	SCALE-UP designer / experienced <i>M&amp;I</i> instructor
Chabay, Ruth	NCSU	<i>M&amp;I</i> author
Risley, John	NCSU	Experienced <i>M&amp;I</i> instructor
Sherwood, Bruce	NCSU	<i>M&amp;I</i> author

One major revision to Figure 3.1 was the replacement of particle energy with rest energy under the component "forms of energy". As opposed to the current version of Figure 3.1, the previous version included particle energy instead of rest energy as one of the basic energy forms. The reason that particle energy was replaced lies in the fact that particle energy can be separated into rest energy and kinetic energy. As one of the *M&I* authors pointed out, although the concept of particle energy is rather important, it is *not* one of the basic energy forms. Since the only basic forms of energy are related to mass, motion and interaction, it is

reasonable to include only rest energy, kinetic energy and potential energy in the test objectives. In addition, testing the concept of particle energy would involve the concept of rest energy. Therefore, it is useful to address rest energy directly in the assessment. Based on this argument, rest energy takes its place in the current version of Figure 3.1.

Another major revision to Figure 3.1 pertained to thermal transfer of energy (heat,  $Q$ ), which was initially absent from the test objectives. Most panel experts suggested that thermal transfer of energy ought to be included in the energy assessment, as it represents an independent means of changing the total energy of a system. This suggestion gave rise to other topics, such as radiation that can also change the energy of a system. Nevertheless, work and thermal transfer of energy are the two topics that are covered in detail in the *M&I* mechanics course. Therefore, it is reasonable to include only work and thermal transfer of energy under the component “means to change the energy of a system”.

Also missing from the previous version of Figure 3.1 was the component “discrete nature of energy levels in atomic systems”. The panel experts recommended that this topic should be added to the test objectives, as it reflects one of the unique features for the *M&I* mechanics course. Based on this suggestion, the discrete nature of energy levels—calculation of absorption and emission spectra—was added to the test objectives.

### 3.1.2 Item creation and review

After the panel experts approved the revised test objectives, the next step was to write test items. Based on the test objectives, a variable number of test items were created for each objective or sub-objective. Take potential energy  $U$  for instance: one item was designed purposefully to test if students can associate  $U$  with pair-wise interactions. In addition,

several other items were created to address the different kinds of potential energy individually, which include gravitational potential energy, electric potential energy, and spring potential energy. At a higher level of “identification and determination of different forms of energy”, test items were created to address the combinations of different energy forms, for example,  $K+U$  graphs and bound or unbound states of a system. At the top level of “application of the energy principle”, quite a few items were generated to test the relationship of the energy change in a system with the work and heat process, as well as to test the energy flow within a closed system as a special case for  $W_{\text{ext}}+Q=0$ .

For each item, alternative choices (distracters) were designed to formulate the energy assessment into a standardized multiple-choice test. Mainly, there are two commonly used methods to write the distracters. First, one can draw on students’ common errors to write the distracters, which often require empirical observations through teaching or research. The other method is to exhaust all the possibilities, if possible, for a question scenario. For example, one item in the energy assessment asks for the sign of gravitational potential energy. Since the possible choice is either positive, zero, negative or not enough information to determine, it is valid to include all of these possibilities in the choices.

In designing the energy assessment, both methods were used to write the distracters. For items where there exist an infinite number of possibilities, distracters were created based on students’ common errors. These students’ common errors became available to the author through two avenues. The first and main avenue was the author’s own teaching experience as well as many discussions with the panel experts, particularly with the *M&I* authors. The second avenue was soliciting written explanations from students using semi open-ended questions. Semi open-ended questions are multiple-choice questions, to which students can

provide their own answers if they find none of the given choices is correct. In this study, the previously designed distracters were rather inclusive, and no new distracters were added using the semi open-ended questions. The semi open-ended questions used in this study essentially confirmed that the previously designed distracters had adequate coverage of student common errors.

For items where there exist a definite number of possibilities, each possibility was converted into a choice. If there exist a definite yet large number of possibilities, then the panel experts were consulted for selection. For example, in the current version of the energy test, one question (Q 6b) asks for how the gravitational potential energy ( $U$ ) and kinetic energy ( $K$ ) of a two-asteroid system change when two asteroids are moving farther apart. The possible answers to this question are of a definite but large number (nine possible choices). The panel experts suggested not to include the two possible distracters: “ $U$  remained constant,  $K$  increased” and “ $U$  remained constant,  $K$  decreased,” as it would be obvious to students that  $U$  must have changed due to the changing distance between the two asteroids as they are moving apart. As for the choice “both  $U$  and  $K$  remain constant”, it stays in the item because it reflects the confusion between  $U+K$  remaining constant and  $U$  &  $K$  each remaining constant.

Upon creation, the test items were submitted to the panel experts for review. Discussions with the panel experts on the test items then followed and permeated the past three years. The purpose of the expert review was multi-fold. First, the panel experts would check if these questions were physically correct, including diagrams, graphs, symbols and numbers. Second, the panel would check if the test items matched the test objectives and the goal of the *M&I* mechanics course. Third, the panel would check if the distracters reflected

their observations of the student common errors. Lastly but not least importantly, the panel experts would identify and correct any typographical or grammatical errors in the items.

Feedback from the panel experts was incorporated into the item revisions, where test items were modified, added or removed. After revision, test items were once again submitted to the panel experts for review, followed by another revision and another review; thereby the review-revision-review cycle started. As for the major revisions, Section 3.1.3 reports the details.

By the fall of 2006, there were 37 test items that survived the expert review and were ready for faculty review. The purpose of faculty review was to solicit opinions on the test items from faculty members who had previously taught the *M&I* mechanics course. Since these faculty members were familiar with the course materials, their feedback on the test items would be valuable for the test improvement. Moreover, these faculty members did not directly participate in designing the test items; thus their perspectives on the test may be different. As a result, it might be easier for them to catch some subtleties that were missed in the previous revisions.

Prior to the faculty review, a list was generated of twelve faculty members who had taught the *M&I* mechanics course at North Carolina State University (NCSU). Also on the list were one faculty member at Purdue and one at Georgia Tech, both of whom had previously taught the *M&I* mechanics course. A solicitation request was sent to all these faculty members on the list. Subsequently, ten<sup>9</sup> faculty members expressed their willingness to review the test items and thus became faculty reviewers. (See Table 3.2.)

---

<sup>9</sup> Among the ten faculty members, one also served as a panel expert.

**Table 3.2 Ten faculty reviewers.**

<b><u>Faculty Reviewer</u></b>	<b><u>Affiliation</u></b>	<b><u>Faculty Reviewer</u></b>	<b><u>Affiliation</u></b>
Beichner, Robert*	NCSU	Huffman, Paul	NCSU
Brown, David	NCSU	Kohlmyer, Matthew	Georgia Tech.
Clark, Laura	NCSU	Pearl, Thomas	NCSU
Daniels, Karen	NCSU	Ramakrishnan, Prabha*	NCSU
Haugan, Mark	Purdue	Weiniger, Keith	NCSU

\* Faculty reviewers marked with asterisk also served as revision reviewers (described later).

The ten faculty reviewers were asked independently to rate each item as either “good”, “neutral”, or “poor”. In order to be rated as “good”, an item had to be physically correct, unambiguously stated, covering the student common errors in the choices, and reflecting the goals of the course. If a faculty reviewer labeled any item as “neutral” or “poor”, (s)he would also need to specify the reason why it was labeled as such. For the faculty’s convenience, several major reasons were provided for them to choose from. These reasons were: “not covered”, “not relevant”, “not important”, “too hard”, “too easy”, and “other (please specify)”. In addition to comments on the individual test items, the faculty reviewers were also asked to comment on the entire test, i.e. report any missing important concepts, inadequate coverage or over coverage of certain topics. (See Appendix A.)

Ratings from the faculty reviewers were then analyzed. To each item, a number was assigned as a score to indicate each reviewer’s rating. A number of 1 was employed to

indicate a rating of “good”, 0 for “neutral”, and –1 for “poor”. As such, the overall ratings for each item were readily available through the average scores.

Except for two items, the average score for each question was 0.8 or above, which suggested that the majority of the items were fairly acceptable questions. As to one of the two items that received low ratings, some faculty members commented that it was too easy. The other item whose average score was below 0.8 had wording issues. (See details in Section 3.1.3.) As to the entire test, all the faculty reviewers thought it was a comprehensive test and reflected the main goals of the *M&I* mechanics course. No missing concepts or inadequate coverage of certain topics were reported.

Feedback provided by the faculty was incorporated into the item revisions, where test items were either modified or eliminated. As for the revisions, Section 3.1.3 reports the details. Upon revision, two faculty members were invited to review the revised items. These two faculty members, referred to as revision reviewers, were made up of one panel expert and one senior faculty member. (See Table 3.2.) Each of the two revision reviewers independently examined the revised items against the faculty reviewers’ comments, and assessed if the revisions resolved the issues or concerns raised by the faculty reviewers. After examining the revisions, both revision reviewers thought that the revisions improved the test items.

At this point, there remained 33 test items. These 33 items finally form the current version of the energy assessment. (See Appendix B.) In brief, Figure 3.2 summarizes the development process of the energy assessment.

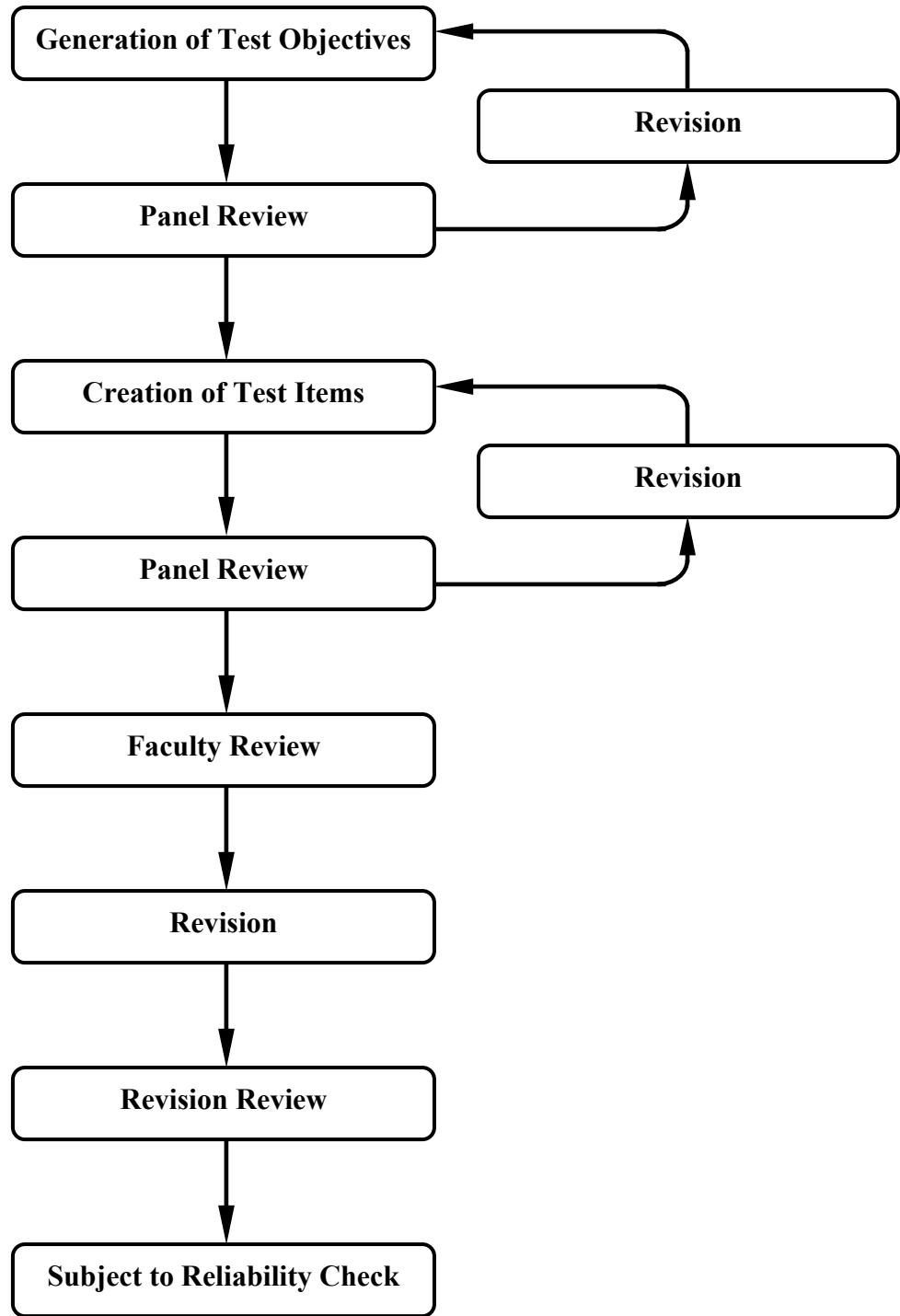


Figure 3.2 Development of the energy assessment.

### 3.1.3 Item revision

Item revisions that took place during the expert panel review and faculty review deserve a whole book by themselves. It is definitely an agonizing process to decide what to be reported in this subsection. Since it is by no means possible to describe every detail, recorded in this subsection are the test items that have gone through major revisions and still remain in the current energy assessment, along with some comments on these revisions. Minor revisions are omitted, such as typographical or grammatical corrections. Also omitted are items that were eliminated during the panel review or faculty review.

*1. An electron was initially traveling at a speed of  $1.4 \times 10^8$  m/s. Some time later, the same electron is observed to be traveling twice as fast. What is the kinetic energy of the electron at this time?*

- (a) *It is the same as the initial kinetic energy, because it's the same electron*
- (b) *It is twice the initial kinetic energy, because it's traveling twice as fast*
- (c) *It is four times the initial kinetic energy, because kinetic energy is proportional to speed squared*
- (d) *None of the above, because its speed is close to  $3.0 \times 10^8$  m/s, the speed of light*

**Figure 3.3 Item 1 in the current energy assessment.**

Initially the above question was not one of the test items in the energy assessment; instead it was in the form as shown in Figure 3.4. As one expert commented during the panel review, the original item was heavily formula-based. In other words, students need to memorize the exact relativistic kinetic energy formula to be able to answer the original item correctly. Based on this comment, the original item was revised into the current form. Although the change seems significant, the current item addresses the same topic on validity of approximation and yet does not involve the exact formula.

*Q. An electron of mass  $m_e$  has speed  $v_e$  close to the speed of light. What is the kinetic energy of this fast electron?*

$$(a) \frac{1}{2} m_e v_e^2$$

$$(b) \frac{1}{2} m_e v_e^2 + m_e c^2$$

$$(c) \frac{\frac{1}{2} m_e v_e^2}{\sqrt{1 - \left(\frac{v_e}{c}\right)^2}}$$

$$(d) \frac{m_e c^2}{\sqrt{1 - \left(\frac{v_e}{c}\right)^2}}$$

$$(e) \frac{m_e v_e^2}{\sqrt{1 - \left(\frac{v_e}{c}\right)^2}}$$

$$(f) \frac{m_e c^2}{\sqrt{1 - \left(\frac{v_e}{c}\right)^2}} - m_e c^2$$

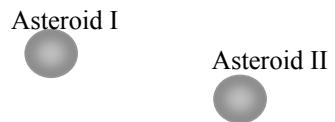
$$(g) \frac{m_e c^2}{\sqrt{1 - \left(\frac{v_e}{c}\right)^2}} + \frac{1}{2} m_e v_e^2$$

**Figure 3.4 The original form of item 1.**

*At a location in space far from any other objects, two moving asteroids (asteroid I and asteroid II) pass each other. Consider the two asteroids as a system, and answer the following two questions.*

*6a. At the instant shown in the diagram, the gravitational potential energy of the system is  $U_1$ . Which of the following is true?*

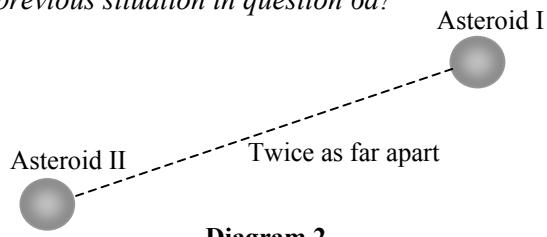
- (a)  $U_1 > 0$
- (b)  $U_1 = 0$
- (c)  $U_1 < 0$
- (d) Not enough information to determine



**Diagram 1**

*6b. At a later time, the asteroids are twice as far apart, as shown in diagram 2. Consider the two asteroids as a system. Which of the following statements about the gravitational potential energy  $U$  and kinetic energy  $K$  is true, compared to the previous situation in question 6a?*

- (a)  $U$  increased,  $K$  decreased
- (b)  $U$  increased,  $K$  increased
- (c)  $U$  increased,  $K$  remained constant
- (d)  $U$  decreased,  $K$  increased
- (e)  $U$  decreased,  $K$  decreased
- (f)  $U$  decreased,  $K$  remained constant
- (g) Both  $U$  and  $K$  remained constant

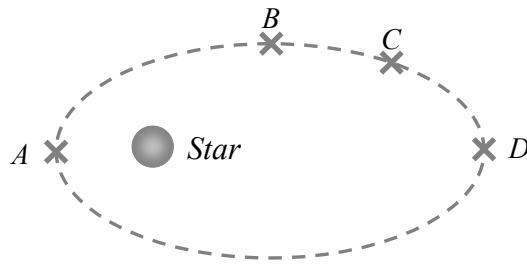


**Diagram 2**

**Figure 3.5 Item 6a and 6b in the current energy assessment.**

As for the above two items (see Figure 3.5), they were revised from their original version that is shown in Figure 3.6. As one can see, the current version differs greatly from its original version in two respects. First, unlike the current version where it directly asks for the sign of gravitational potential energy ( $U_g$ ), the original version asked students to compare the values of  $U_g$  at different locations. In order to answer the original item correctly, students need to take into consideration both the distance dependence and the sign of  $U_g$ , which was, as commented by one of the panel experts, unnecessarily confusing and tricky. Second, the original version only asked students to find locations with the extreme values for  $K$  and  $U_g$ , but it did not ask for the information on the trend of change in  $K$  or  $U_g$ , which was regarded as rather static by one of the panel experts. Therefore, revisions were made to the original items. As opposed to the original items, the current two items disentangle the complications from the distance dependence and the sign of  $U_g$ , as well as address the trend of change in both  $K$  and  $U_g$ .

*Q. The following diagram shows a top view of an asteroid going around a star along the dotted path.*



*At which of the following points does the star-asteroid system have the greatest gravitational potential energy?*

- |             |                             |             |
|-------------|-----------------------------|-------------|
| (a) Point A | (b) Point B                 | (c) Point C |
| (d) Point D | (e) Same value at all point |             |

*At which of the following points does the asteroid have the greatest kinetic energy?*

- |             |                             |             |
|-------------|-----------------------------|-------------|
| (a) Point A | (b) Point B                 | (c) Point C |
| (d) Point D | (e) Same value at all point |             |

**Figure 3.6 The original form of item 6a and 6b.**

4. Consider three moving balls.

Ball 1 moves with a constant speed in a constant direction.

Ball 2 moves with a constant speed, but its direction changes.

Ball 3 moves in a constant direction, but its speed changes.

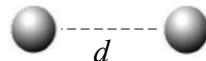
Which of these balls has (have) a constant kinetic energy?

- (a) Ball 1
- (b) Ball 1 and ball 2
- (c) Ball 2 and ball 3
- (d) Ball 1 and ball 3
- (e) Ball 1, ball 2, and ball 3
- (f) None of the above

Figure 3.7 Item 4 in the current energy assessment.

In the above item (see Figure 3.7), choice (a) was not originally one of the distractors. It was during the faculty review that choice (a) was added based on the following consideration. Students may think only a constant speed in a constant direction will result in a constant kinetic energy, in which case some students may choose “Ball 1 only” as their answer. Therefore, a distracter “Ball 1” was added. Correspondingly, the question was rephrased from “which of these balls have a constant kinetic energy” to “Which of these balls has (have) a constant kinetic energy?”

5. When two protons are a distance  $d$  apart, the electric potential energy of the two-proton system is  $U$ .



Now another two protons are taken into account. These four protons form a tetrahedron with a distance  $d$  between any two of them. The electric potential energy of the four-proton system is:

- (a)  $U$
- (b)  $2U$
- (c)  $3U$
- (d)  $4U$
- (e)  $8U$
- (f)  $12U$

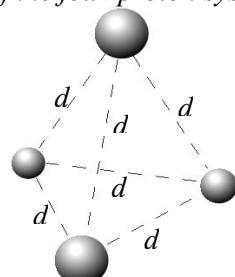
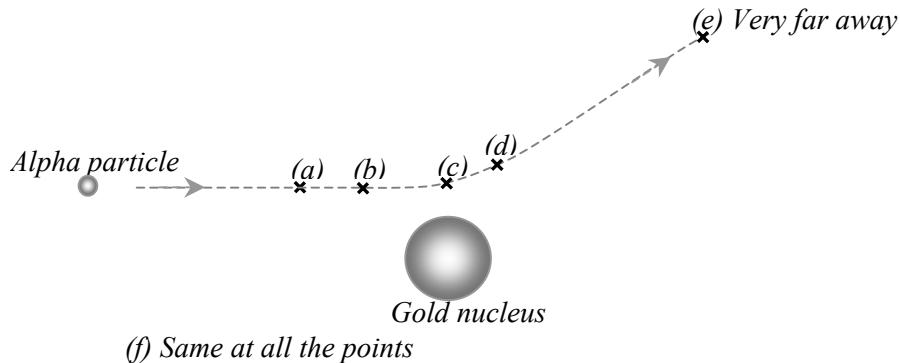


Figure 3.8 Item 5 in the current energy assessment.

The choices (c) and (f) in the above item (see Figure 3.8), which were not originally the distracters, were added during the faculty review. The justification was that, as argued by a few faculty reviewers, some students might mistakenly divide or multiply the number of pair-wise interactions (6 in this case) by 2. Although such mistakes were not detected previously using the semi open-ended questions, it is possible that some students may make these mistakes. Therefore, this suggestion was adopted.

*7. An alpha particle (charge  $+2e$ ) is deflected by a heavy gold nucleus (charge  $+79e$ ) that remains nearly stationary during the collision. The following diagram depicts the trajectory of the alpha particle. Consider a system that consists of both the alpha particle and the gold nucleus. At which of the following marked points does the system have the greatest electric potential energy?*



**Figure 3.9 Item 7 in the current energy assessment.**

Initially, point (e) in the above item (see Figure 3.9) was not located at the far end of the path, nor was it marked as “very far away”. Instead, point (e) was placed a bit farther away from the gold nucleus than point (d) along the path. During the faculty review, one reviewer commented that a point far away from the gold nucleus ought to be a choice, since some students may confuse it with gravitational potential energy and think the electric potential energy of the system is greatest when the alpha particle is far away from the gold nucleus. Though such mistake was not observed through the semi open-ended version of this

question, it was a valid suggestion. Based on this comment, choice (e) was changed to be a point very far away from the gold nucleus.

8. Spring 1 and spring 2 are made of different materials but have the same stiffness. Spring 1 has longer relaxed length than spring 2. After both springs are compressed by 2 cm, the spring potential energy of spring 1 is changed by  $\Delta U_1$  and the spring potential energy of spring 2 is changed by  $\Delta U_2$ . Which of the following statements is true?

- (a)  $\Delta U_1 > \Delta U_2$
- (b)  $\Delta U_1 = \Delta U_2$
- (c)  $\Delta U_1 < \Delta U_2$
- (d) Not enough information to determine

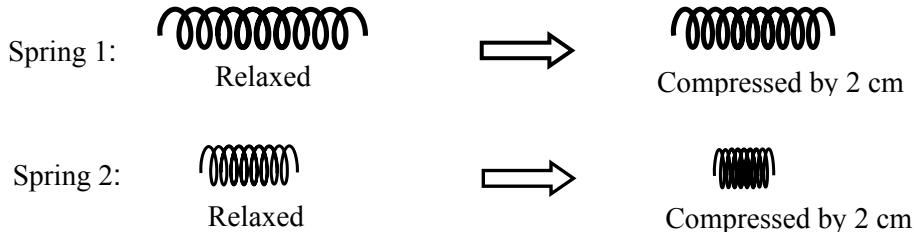


Figure 3.10 Item 8 in the current energy assessment.

Prior to revision, the question statement of the above item (see Figure 3.10) read as follows: "Spring 1 and spring 2 have the same stiffness, but different relaxed length  $L_1$  and  $L_2$  with  $L_1 > L_2$ . Both springs are being compressed by 2 cm..." Accordingly, notations such as " $L_1$ " and " $L_1-2\text{cm}$ " were used previously in the diagram to indicate the initial relaxed length and the final compressed length of the springs. During the faculty review, one reviewer expressed her concerns about the notation " $L_1-2\text{cm}$ " in the diagram, which she thought was unusual. Related to the issue of notations were the many symbols involved in this item, which may also introduce unnecessary difficulties to students. Based on this consideration, the question statement was revised into the following version: "spring 1 and spring 2 have the same stiffness, but spring 1 has longer relaxed length than spring 2. After

*both springs are compressed by 2 cm...*" Correspondingly, the textual expressions "relaxed" and "compressed by 2cm" were used for the captions in the diagram.

Very recently, one of the *M&I* authors identified a subtlety regarding the diagram in the above item. Originally, the diagram showed no difference in the wire thickness between the two springs. Consequently, it would be possible that some students may spontaneously assume, by looking at the diagram, that spring 2 has a greater stiffness than spring 1 because of the fact that a shorter spring has a greater stiffness than a longer spring of the same material. Although the question statement clearly mentioned that the two springs "have the same stiffness", it would be helpful to redraw the diagram with different wire thickness to avoid such an unnecessary mistake. Based on this consideration, the diagram was redrawn; the current diagram now shows a noticeable difference in the wire thickness. In addition, the question statement was once again revised, and now the current item explicitly mentions that the two springs are made of different materials.

17. In a physics lab, a ball attached to a low-mass spring is oscillating up and down. A student chooses a system for analysis. At a certain time, the student finds the system's kinetic energy  $K$  is 0.3 J, the system's spring potential energy  $U_{\text{spring}}$  is 0.4 J, and the system's gravitational potential energy  $U_{\text{grav}}$  is -0.2 J. What was the student's choice of system?

- (a) Ball
- (b) Ball and spring
- (c) Ball and Earth
- (d) Spring and Earth
- (e) Ball, spring, and Earth

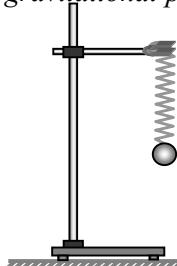


Figure 3.11 Item 17 in the current energy assessment.

The above item (see Figure 3.11) originated from a rather similar question shown in Figure 3.12. As one can see, both items address the same topic—system specification given

the kinds of energy involved. However, the original item gave zero kinetic energy and zero spring potential energy, which some students may falsely interpret as non-existing kinetic energy or spring potential energy in the system. To this end, some panel experts suggested to redesign the question scenario into the current form with the values of energy all being nonzero. As such, the unnecessary confusions would be avoided.

*A ball is launched vertically up from a low-mass compressed spring. When the ball reaches its maximum height, you claim the system's kinetic energy is 0 ( $K = 0$ ), the spring potential energy is 0 ( $U_{\text{spring}} = 0$ ), and the gravitational potential energy is nonzero ( $U_{\text{grav}} \neq 0$ ). What was your choice of system?*

- (a) Ball
- (b) Ball and spring
- (c) Ball and Earth
- (d) Spring and Earth
- (e) Ball, spring, and Earth

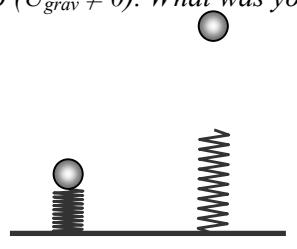


Figure 3.12 The original form of item 17.

*14. Consider a falling stone to be a system and answer the following question. At the instant when the stone is still falling down and has not yet hit the ground, which of the following forms of energy does the system of the falling stone have?*

- (a) Rest energy only
- (b) Rest energy and kinetic energy
- (c) Rest energy and gravitational potential energy
- (d) Rest energy, kinetic energy and gravitational potential energy
- (e) None of the above

Figure 3.13 Item 14 in the current energy assessment.

Previously, the above item (see Figure 3.13) read as follows: “*Consider a falling stone to be a system and answer the following question. At the instant when the stone is several meters above the ground, which of the following forms of energy does the system of the falling stone have?*” During the faculty review, the original item received an average

rating lower than 0.8. In fact, the average rating was 0.7, which was not alarmingly low. The main concern raised by one faculty reviewer was the wording issue about “falling stone” and “several meters above the ground” in the original statement. This reviewer thought that a falling stone could be momentarily at rest. Therefore, he could not decide which choice, (a) or (b), was the correct answer. As a matter of fact, the wording “falling stone” implies that the motion direction of the stone is downward. The momentary zero speed could not occur unless the ball had an initial upward motion, which was neither stated nor implied in the original item. Nonetheless, the original statement was modified based on the reviewer’s suggestion in order to avoid such possible misunderstanding. Now the item reads as above.

*18. You pull one end of a spring over a small distance  $d$  with a force of magnitude  $F$  to the left. At the same time, you pull the other end of the spring over a small distance  $d$  with a force of the same magnitude  $F$  to the right. Consider the spring as a real system. The total work done on the spring is:*

- (a)  $-2Fd$       (b)  $-Fd$       (c) 0      (d)  $Fd$       (e)  $2Fd$       (f) None of the above

**Figure 3.14 Item 18 in the current energy assessment.**

Two noteworthy revisions occurred to the above item (see Figure 3.14). The first revision pertained to the phrase “real system”, which did not appear in the previous version of this item. The previous version of this item asked students only to “consider the spring as a system” without specifying it as “a real system”. To this respect, some panel experts recommended that the statement should explicitly indicate “the spring as a real system” to emphasize its difference from a point particle system. This suggestion was then adopted.

The second major revision to the above item took place during the faculty review. Prior to the faculty review, the above item read as follows: “*You pull one end of a spring with*

*a force of magnitude  $F$  to the left, making a small displacement of magnitude  $d$  to the left. At the same time, you pull the other end of the spring with a force of the same magnitude  $F$  to the right, making a small displacement of the same magnitude  $d$  to the right. Consider the spring as a real system..."* Because of the question statement, one faculty reviewer rated this item as a “poor” question. The reason she provided was that the statement “making a small displacement of magnitude  $d$ ” was confusing, for it didn’t specify whose displacement it was, the center of mass or the end of the spring. Based on this comment, the statement was revised into the current version as above.

22. A box of mass  $M$  is moving at speed  $v$  toward a person along a surface of negligible friction. A person leans against a wall with both arms stretched and applies a pushing force to stop the box. Finally the person brings the box to a full stop. There is no temperature change in the box at any time. We can conclude that during the process:

- (a) The amount of work done on the box by the person was  $+\frac{1}{2}Mv^2$
- (b) The amount of work done on the box by the person was positive, but there is not enough information to determine the value
- (c) The amount of work done on the box by the person was 0
- (d) The amount of work done on the box by the person was  $-\frac{1}{2}Mv^2$
- (e) The amount of work done on the box by the person was negative, but there is not enough information to determine the value

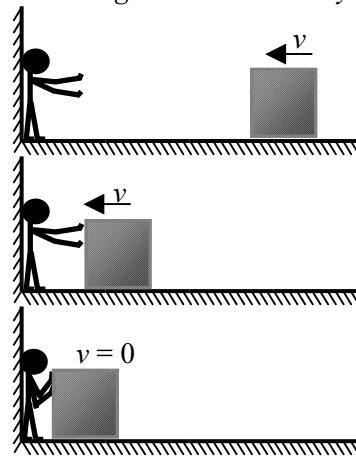


Figure 3.15 Item 22 in the current energy assessment.

As for the above item (see Figure 3.15), the original scenario was a box moving toward the right along the surface and being stopped by a person. In an early stage of the study, a semi open-ended question of the original item was used to solicit student reasoning. It was found that quite a few students answered negative work simply because the person

exerted a force in the  $-x$  direction to stop the box. Therefore, some students ended up choosing the correct answer based on the wrong reasoning. During the panel review, experts were consulted for this issue. Some experts recommended that the motion direction be changed from moving toward the right ( $+x$  direction) to moving toward the left ( $-x$  direction). The item was then revised, and the diagram in the item now shows a box moving toward the left.

*24. You warm up some water on a stove. The transfer of energy due to a temperature difference from the fire to the water is 300 J. At the same time, you stir the water with a beater, doing 200 J work on the water. After the stove is turned off, the beater is removed, and the water stops swirling, the thermal energy of the water is increased by:*

- |           |           |           |
|-----------|-----------|-----------|
| (a) 0     | (b) 100 J | (c) 200 J |
| (d) 300 J | (e) 500 J |           |

**Figure 3.16 Item 24 in the current energy assessment.**

As for the above item (see Figure 3.16), there was initially not a statement that “*the stove is turned off, the beater is removed, and the water stops swirling.*” It was during the expert panel review that this statement was added. As one panel expert commented, this statement marks the end of the work and heat process, as well as it indicates that the work and heat process only causes an increase in thermal energy. Therefore this statement is necessary.

During the subsequent faculty review, the above item received an average rating lower than 0.8. What resulted in low ratings for the above item was the concern that it is an easy question. As many physics education researchers are well aware, students often fail to answer questions that instructors regard as easy. Questions in the FCI, for example, were once despised, in that many physics instructors thought these questions were too easy for

their students. Nonetheless, the results are very disappointing; students' poor performance on the FCI is both pervasive and enduring. Because of the intriguing results, many physics instructors have started paying a closer attention to easy questions; questions that are seemingly easy yet diagnostic are thus adopted in many physics tests. As to the above question, though it is relatively easy as opposed to others in the assessment, it addresses the important concept—heat—that is not covered elsewhere in the assessment. Additionally, easy as it seems, the above question requires the difference and relation between work and heat. Therefore, students' performance on this question may not be as good as what the instructors expect. Based on these arguments, this item remains in the test.

25. *A clay ball moving to the right has kinetic energy of 10 J. An identical clay ball is moving to the left at the same speed. The two balls smash into each other and both come to a stop. What happened to the energy of the two clay-ball system?*

- (a) *The kinetic energy of the system did not change*
- (b) *The kinetic energy changed into thermal energy*
- (c) *The total energy of the system decreased by an amount 20 J*
- (d) *The initial kinetic energy of the system was zero, so there was no change in energy*

**Figure 3.17 Item 25 in the current energy assessment.**

Initially, there was one more distracter in the above item—“(e) all of the above”. (See Figure 3.17.) During the faculty review, several reviewers pointed out that this distracter “(e) all of the above” is logically flawed, for choices (a) and (d) are mutually exclusive with choice (c). Therefore, the alternative choice (e) was eliminated, which leaves the current item with the above four choices.

As a caveat to the aforementioned revisions, particular attentions were paid to some interesting issues about the test items during the development process. These interesting

issues include symbolic/numerical representations of physics quantities, diagrams, reasoning steps, and content/cognition levels of the test items. Details on these issues are addressed in Chapter 4.

### 3.1.4 Two versions of the energy assessment

The current version of the energy assessment is a 33-item multiple-choice test that covers the energy topics discussed in the *M&I* mechanics course, including application of the energy principle, forms of energy, system specification, work and heat process, and discrete energy levels in atomic scale. The target audience of the assessment is students who have taken the *M&I* mechanics course or students who have learned from other advanced physics classes the energy topics comparable to those covered in the *M&I* course. For the latter, examples include senior physics major undergraduate students and physics major graduate students. Typically, the time length for a student to complete the assessment varies from 30 minutes to an hour.

Aside from the current version, a pilot version of the energy assessment was compiled in the early stage of the study. (See Appendix C.) As a predecessor of the current energy assessment, the pilot version shares similarities with the current version in test format, content, and target audience. More specifically, the pilot version of the energy assessment is also a multiple-choice test that covers the energy topics discussed in the *M&I* mechanics course, and it is designed for the same audience as the current version.

However, the pilot version is by no means a comprehensive test, for it lacks an adequate coverage of the energy topics. Topics that are not addressed in the pilot version include rest energy, thermal transfer of energy and discrete energy levels. In addition, as new

items were added to expand the coverage, several old items were removed from the pilot version during the development process. As a result, the current version differs from the pilot version in length as well. Compared to the pilot version where there are 28 items, the current version has 33 items. (See Table 3.3.)

**Table 3.3 Comparison between the pilot version and current version of the energy assessment.**

	<b>Pilot Version</b>	<b>Current Version</b>
<b>Format</b>	Multiple-choice test	Multiple-choice test
<b>Content</b>	Inadequate coverage of the energy topics	Adequate coverage of the energy topics
<b>Target audience</b>	<i>M&amp;I</i> students, or similar	<i>M&amp;I</i> students, or similar
<b>Number of items</b>	28 items	33 items
<b>Test time length</b>	30~60 minutes	30~60 minutes

One footnote to the pilot version is that it is, after all, a product of the pilot study in early stage, and is not meant to be a readily useable assessment tool. Immature as it is compared to the current version, the pilot version is every bit a milestone in the development of the energy assessment.

### **3.2 Validity, reliability and discriminatory power of the energy assessment**

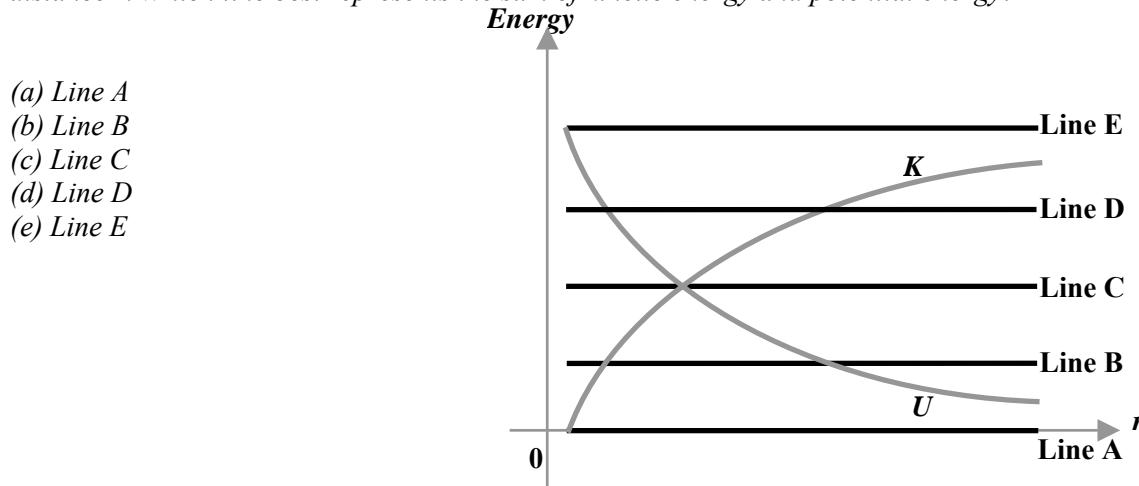
#### 3.2.1 Validity of the energy assessment

Two aspects of validity are particularly relevant and were thus established for the current version of the energy assessment. These two aspects of validity are “face validity” and “content validity”, both of which were assessed by the panel experts and faculty reviewers during the development process.

In an attempt to establish the face validity of the energy assessment, panel experts scrutinized every question to ensure that the items were highly related to the subject matter—energy. For some items that were only remotely related to the subject matter in physics, say more closely related to math, panel experts would suggest eliminating these items. For an example, the item shown in Figure 3.18 was one of the items in the pilot version, which was subsequently removed due to its low relevancy to physics. Although this item mentioned kinetic energy and potential energy in the context, students would not necessarily need physics to answer this question correctly. In fact, as one of the panel experts pointed out, this question may not really test physics.

After the panel review, the face validity of the energy assessment was essentially established, and the subsequent faculty review further strengthened it.

*The following figure plots kinetic energy  $K$  and potential energy  $U$  of a two-particle system versus distance  $r$ . Which line best represents the sum of kinetic energy and potential energy?*



**Figure 3.18 An item that was removed during the expert panel review due to its low relevancy to physics.**

As for the content validity—sufficient coverage of the subject matter—of the energy assessment, much effort was devoted to the test objectives as described in Section 3.1.1, as

well as to the items as described in Section 3.1.2. As such, the current energy assessment covers all of the following five topics regarding energy: application of the energy principle, forms of energy, system specification, work and heat process, and discrete energy levels. In fact, the content validity—adequate coverage of the energy topics—of the current energy assessment was very much established during the expert panel review, so much so that all the reviewers in the subsequent faculty review agreed that the test had an adequate coverage of the energy topics discussed in the *M&I* mechanics course.

### 3.2.2 Reliability and discriminatory power of the energy assessment

The reliability and discriminatory power of the energy assessment was evaluated through five indices (coefficients): item difficulty index ( $P$ ), item discriminatory index ( $D$ ), item point bi-serial coefficient ( $r_{\text{ppi}}$ ), KR-20 reliability index ( $r_{\text{test}}$ ) and Ferguson's delta ( $\delta$ ). This subsection describes the means that were used to calculate these indices (coefficients), whereas Chapter 4 reports the results. As for more information on these five indices (coefficients), Ding *et al.* (Ding, Chabay, Sherwood & Beichner, 2006) provided a detailed “tutorial” on the definition, calculation and application of these statistics.

To obtain the item difficulty index ( $P$ ), a proportion of correct responses for each item was calculated as  $P = N_1/N$ , where  $N_1$  is the number of correct responses to a particular item, and  $N$  is the total number of students who took the assessment. Ideally, test items with difficulty index around 0.5 have the highest reliability and discriminatory power (Ghiselli, Campbell & Zedeck, 1981). Practically, values of item difficulty index ranging from 0.3 to 0.9 are acceptable (Doran, 1980). In this study, the criterion  $P \in [0.3, 0.9]$  was used to check the average item difficulty for the energy assessment.

In order to obtain the item discriminatory index ( $D$ ), students were grouped into four quartiles according to their total scores on the energy assessment. In the top quartile and the bottom quartile, the numbers of students who answered the item correctly were counted, namely  $N_H$  and  $N_L$ . Also counted was the total number of students ( $N$ ) who took the energy assessment. The item discriminatory index ( $D$ ) was then calculated as follows:

$$D = \frac{N_H - N_L}{N/4}$$

A commonly used criterion for the item discriminatory index is  $D \geq 0.3$ , which was adopted in this study to check the average item discrimination of the energy assessment (Doran, 1980).

To examine the item point bi-serial coefficient ( $r_{pbi}$ ), which is a measure of the item reliability, the correlations between the individual item scores and the total scores were calculated as follows:

$$r_{pbi} = \frac{\bar{X}_1 - \bar{X}_0}{\sigma_X} \sqrt{P(1-P)}$$

Here  $\bar{X}_1$  is the average total score for those students who correctly answered a particular test item,  $\bar{X}_0$  is the average total score for students who incorrectly answered the item,  $\sigma_X$  is the standard deviation of the total score for all the students who took the energy assessment, and  $P$  is the difficulty index for this item. Normally, a satisfactory point bi-serial coefficient is higher than 0.2 (Kline, 1986). This standard was adopted to check the average item reliability for the energy assessment.

In an attempt to evaluate the reliability ( $r_{test}$ ) of the entire energy assessment, the Kuder-Richardson KR-20 index was calculated as follows:

$$r_{test} = \frac{K}{K-1} \left( 1 - \frac{\sum P(1-P)}{\sigma_x^2} \right)$$

$K$  is the number of the test items,  $\sigma_{xi}$  is the standard deviation of the  $i^{\text{th}}$  item score, and  $\sigma_x$  is the standard deviation of the total score. A widely accepted criterion is that tests with reliability index higher than 0.7 are reliable for group measurement (Doran, 1980). This criterion was used to check if the energy assessment is a reliable test.

Ferguson's delta ( $\delta$ ), which indicates the discriminatory power of the entire test, was calculated for the energy assessment as follows:

$$\delta = \frac{N^2 - \sum f_i^2}{N^2 - N^2 / (K+1)}$$

Here,  $N$  is the total number of students who took the test,  $K$  is the number of test items, and  $f_i$  is the number of students whose total score was  $i$ . A criterion  $\delta > 0.90$  was used in this study to check if the energy assessment provides a good discrimination (Kline, 1986).

### 3.3 Student participants

During the development of the assessment, students at different academic levels and from different institutions participated in taking the two versions of the energy assessment; thus providing valuable data for the assessment validation. For future reference convenience, the student participants are grouped into five samples based on assessment version, student academic level and institution. Consequently, this section is divided into five subsections with each subsection describing one sample of participants.

### 3.3.1 Sample 1

Students who participated in taking the pilot version of the energy assessment were from the *M&I* mechanics class in SCALE-UP settings in the 2005 spring semester. The SCALE-UP *M&I* mechanics course provided two 100-minute long sessions and one 50-minute long session every week with labs incorporated into the class. Students enrolled in this course were either science or engineering major undergraduates who were in their freshman or sophomore year at NCSU. The pilot version of the energy assessment was administered, upon permission from the class instructor, to 77 students who attended the SCALE-UP *M&I* mechanics class on the day the assessment was scheduled.

### 3.3.2 Sample 2

Participants who took the current version of the energy assessment consisted of students at different levels of academic study and from different institutions. Primarily, the majority participants were students who attended the *M&I* mechanics course in lecture hall settings in the 2006 fall semester at NCSU. Similar to the students in the SCALE-UP class, students in the *M&I* mechanics class in lecture hall settings were mainly freshmen or sophomore in engineering or science majors. Four parallel sections taught by three lecturers were offered in the 2006 fall semester: Sec001, Sec002, Sec003 and Sec004. Students who were enrolled in any one of these sections met three times every week for a 50-minute long lecture. In addition, students attended a 110-minute long lab every week, where they performed hands-on experiments, wrote computer-modeling programs using VPython (<http://vpython.org>), and solved physics problems on whiteboards.

Upon permission from the course instructors, the energy assessment was administered as both a pre-test and a post-test to students in the four sections in the fall semester of 2006. In total, 319 students took the pre-test and 308 students took the post-test. In particular, 262 students completed both the pre-test and post-test. Table 3.4 shows the distribution of the students across the four sections.

**Table 3.4 Student participants from the four sections of the M&I mechanics course in lecture hall settings in the 2006 fall semester at NCSU.**

	<u>Pre-test</u>	<u>Post-test</u>	<u>Both pre and pos-test</u>
<u>Section 1</u>	91	78	74
<u>Section 2</u>	79	75	66
<u>Section 3</u>	81	74	65
<u>Section 4</u>	68	76	57
<u>Anonymous</u>	N/A	5	N/A
<u>Total</u>	319	308	262

### 3.3.3 Sample 3

In addition to the aforementioned participants, a relatively small number of student participants from higher levels of academic study completed the current version of the energy assessment as well. Among them were five *M&I* graduate teaching assistants (TA) and eleven undergraduate teaching assistants (UTA) in the 2006 fall semester at NCSU.

These five TAs were physics major graduate student lab instructors assigned to the *M&I* course, who either had experience teaching the *M&I* mechanics labs or had reportedly read through the relevant materials on momentum and energy in the *M&I* textbook by the time they took the energy assessment. The eleven UTAs were undergraduate student lab assistants who previously took the *M&I* courses and excelled in the course materials and were thus familiar with the energy topics covered in the *M&I* mechanics course.

### 3.3.4 Sample 4

Also among those who completed the current energy assessment were twenty-seven senior physics major undergraduates and nine physics major graduate students in a quantum physics class (PY401/501) offered in the fall semester of 2006 at NCSU.

Of those twenty-seven undergraduates, only one previously attended the *M&I* mechanics class. Nevertheless, these senior physics major undergraduates learned from other advanced physics classes the energy topics comparable to those discussed in the *M&I* mechanics course.

Of the nine graduate students who attended the quantum physics class as a remedial course in their early years of graduate study, four students were also *M&I* TAs. These four *M&I* TAs were not included in the aforementioned five TAs in sample 3. Similar to the participants in sample 3, these four TAs from the quantum physics class either had experience teaching the *M&I* mechanics labs or had read through the course materials pertaining to the energy topics before taking the energy assessment. (Therefore, these four TAs are grouped with the five TAs into sample 3 for data analysis in Chapter 5.) The other

five graduate students in the quantum physics class were also supposedly familiar with the relevant energy topics based on their level of academic study.

### 3.3.5 Sample 5

Moreover, the current version of the energy assessment was administered among twenty-nine graduate students at Purdue University in the 2006 fall semester. These students were first-year physics major graduate students and had been reportedly reading the course materials up to Chapter 5 in the 1<sup>st</sup> edition of *M&I* by the time of taking the energy assessment (through personal communication with Mark Haugan).

In brief, these five participant samples mentioned above are summarized in Table 3.5.

**Table 3.5 Five samples of participants who took one of the two versions of the energy assessment.**

	<u>Version</u>	<u>Participants</u>	<u>No. of Participants</u>	<u>Institution</u>	<u>Time</u>
<b>Sample 1</b>	Pilot	Students from <i>M&amp;I</i> SCALE-UP class	77	NCSU	2005S
<b>Sample 2</b>	Current	Students from <i>M&amp;I</i> lecture hall class	Pre-test: 319 Post-test: 308	NCSU	2006F
<b>Sample 3</b>	Current	<i>M&amp;I</i> (U)TAs	TA: 5 UTA: 11	NCSU	2006F
<b>Sample 4</b>	Current	Students from PY401	Undergrad. : 27 Grad. ( <i>M&amp;I</i> TA): 4* Grad. (Other): 5	NCSU	2006F
<b>Sample 5</b>	Current	First-year physics graduate students	29	Purdue	2006F

\*: These four *M&I* TAs were not included in sample 3.

### **3.4 Test Settings and procedures**

It was in different test settings and through different procedures that these five samples of participants completed either one of the two versions of the energy assessment. As a result, this section consists of five subsections, each of which details the test settings and procedures for one sample of participants.

#### **3.4.1 Test settings and procedures for sample 1**

In the spring semester of 2005, the pilot version of the energy assessment was administered as an in-class test to 77 students in the SCALE-UP class. This test was given after students studied in class all the relevant materials on energy. To reduce the possible effects posed by the test, the class instructor assured students, prior to the test, that the test was for research purposes and the results would not affect their course grades. However, students were required to complete the test independently without reference to textbooks or discussions with other fellow students. Additionally, the class instructor encouraged students to reduce guessing to minimum.

This test was administered on paper. For the students' convenience, pencils and scratch paper were furnished. As for the test period, a sufficient time length—50 minutes—was given to students to answer all the questions. Most students took approximately 30 minutes to complete the test. Upon completion, students turned in the test papers and scratch paper. Students' responses were then manually entered into computer for analysis.

### 3.4.2 Test settings and procedures for sample 2

In the fall semester of 2006, the current version of the energy assessment was administered as both a pre-test and a post-test to students in the four parallel sections of the *M&I* mechanics course in lecture hall settings at NCSU. The pre-test was conducted on WebAssign (Risley, 1999) as an online homework assignment at the beginning of the semester, whereas the post-test was given as an in-class pencil-and-paper test at the end of the semester.

The advantage of using WebAssign to administer the current version of the energy assessment is two-fold. First, giving the questions to students as a homework assignment avoids using the lecture or lab time, particularly the precious time every course instructor is frugal with at the beginning of each semester. Second, WebAssign automatically records student responses and makes them instantly available for analysis; thus greatly facilitating the data collection process. However, giving a test as a homework assignment may not yield a well-controlled condition. As to this, the following measures were taken to alleviate the concern.

First, instructions were provided to students before they took the test. The instructions explicitly advised students not to refer to any notes, textbooks, or discuss with others. Additionally, students were assured that they would not be penalized for wrong answers. Indeed, upon completion, students would get an insignificant amount of extra credit points for homework. Nevertheless, it was emphasized that students should take the assignment seriously.

Secondly, there was a time limit for this assignment; after the time limit was passed WebAssign would not accept students' submissions. Students were instructed not to open the

assignment unless they were ready to dedicate up to 60 minutes to the questions. Students were not allowed to peek at the questions then come back later to complete them. Once they opened the assignment, they had no other choices but proceed to answer the questions in order to earn the credit.

Third, only one submission was allowed for the entire assignment. As such, students were forced to work continuously on the test without the distractions from partial submissions. Within the time limit, students could make as many changes as they wanted. But once they submitted the assignment, they could not change their answers any more. This one-submission requirement, together with other aforementioned measures, simulated the real test environment.

Since it was planned to administer the same test again at the end of the fall semester as a post-test, it was necessary to protect the test questions from being studied. To this respect, the assignment was purposefully scheduled in such a way that students would not receive any kind of feedback on their performances, not even a score or check mark. In addition to that, this assignment was scheduled to be invisible immediately after it was due. Consequently, students could not see the assignment nor the questions in the assignment after the due day.

With the above measures, the current version of the energy assessment was administered on WebAssign as a homework assignment at the beginning of the fall semester of 2006—toward the end of the second week—before any relevant topics on energy were covered in the lectures.

As opposed to the pre-test, the post-test was administered as an in-class pencil-and-paper test at the end—in the second to last week—of the fall semester of 2006. Prior to the

post-test, the three instructors from the four parallel sections of the *M&I* mechanics course had many careful discussions on when to administer the energy assessment. Finally, all the instructors agreed to conduct the energy assessment in four sections within one day. Of the four sections, three were consecutive, and the fourth started ninety minutes after the third section.

In each section, the instructor first made an announcement describing the research purpose of the energy assessment. After confirming the fact that the test score would not affect the final grade of the participants, each instructor reminded the students to try their best in answering the questions. Students were given over 45 minutes to take the assessment; most of the students completed the questions within 30 minutes. In taking the assessment, students were not allowed to refer to their textbooks or notes, nor were they to discuss with others. Scratch paper and pencils were furnished for the students' convenience; optical scan sheets were provided for students to record their answers. Students could mark anything on the test papers, but only their names and answers on the op-scan sheets.

Upon completion of the test, students turned in both the test papers and op-scan sheets in order to secure the test questions. The op-scan sheets were then sent to the computer center to scan.

### 3.4.3 Test settings and procedures for sample 3

In the sixth week of the 2006 fall semester, the current energy assessment was administered among five *M&I* TAs and eleven UTAs on WebAssign at the weekly (U)TA meeting.

A bogus session was first created for the (U)TAs on WebAssign, where each (U)TA could only have access to the posted assignments as a student. Consequently, they could see the assignments without having access to the answers. The current energy assessment was then posted and scheduled for this bogus session on WebAssign. To secure the questions, the assessment was password protected and had a time limit; (U)TAs who participated in taking the assessment were given up to 60 minutes to complete all the questions and submit their answers to WebAssign. Only one submission was allowed for the entire test. Upon submission, the WebAssign window would refresh and provide no feedback.

All the (U)TAs who took the assessment were aware that it was a research-based test, and that their voluntary participation would help validate the energy assessment.

#### 3.4.4 Test settings and procedures for sample 4

In the same week—the sixth week of the 2006 fall semester, the professor in charge of the quantum physics class (PY401) at NCSU administered, upon the author’s request, the current version of the energy assessment as an in-class pencil-and-paper test. On the day the assessment was scheduled, twenty-seven senior physics major undergraduate students and nine physics major graduate students attended the class and thus completed the assessment. Students in the class were informed that this was a research-based test and the results would not affect their grades. However, they were asked to answer the questions seriously without reference to notebooks or discussions with others. Students were given up to 50 minutes to take the assessment, and most of them completed the test within approximately 30 minutes. After the test, both the test papers and the answer sheets were collected, and the results were not provided to students.

### 3.4.5 Test settings and procedures for sample 5

In the eighth week of the 2006 fall semester, the current version of the energy assessment was administered among twenty-nine first-year physics major graduate students at Purdue University. Mark Haugan, who is an associate professor in the physics department at Purdue, conducted, at the author's request, the current energy assessment as a pencil-and-paper test. Prior to the test, students were informed that this was a research-based test. However, independent completion of the assessment was required. To encourage the students seriously to answer the questions, a 50-minute long test period was provided. Some students completed the test "within half an hour, and a few took almost 50 minutes", as Haugan reported. Upon completion, students' test papers were collected and mailed to the author for analysis.

In brief, Table 3.6 summarizes the distinctive features of the test settings and procedures for each sample.

**Table 3.6 Test settings and administration formats for the five samples of participants.**

<b><u>Sample</u></b>	<b><u>Version</u></b>	<b><u>Occurrence</u></b>	<b><u>Administration Format</u></b>			<b><u>Time</u></b>
<b>Sample 1</b>	Pilot	Post-instruction	In-class	Paper-based	Test	50min
<b>Sample 2</b>	Current	Pre-instruction	After-class	WebAssign	Homework	60min
		Post-instruction	In-class	Paper-based	Test	50min
<b>Sample 3</b>	Current	N/A	TA-meeting	WebAssign	Test	60min
<b>Sample 4</b>	Current	N/A	In-class	Paper-based	Test	50min
<b>Sample 5</b>	Current	N/A	In-class	Paper-based	Test	50min

### **3.5 Quantitative and qualitative analyses**

To evaluate student understanding of energy through the current energy assessment, both quantitative and qualitative analyses were performed in this study.

#### 3.5.1 Quantitative analyses

Quantitative analyses were performed on a large number of data collected from the student participants who completed the current version of the energy assessment. Particularly, attention was focused on the majority participants—the participants who were from the four sections of the *M&I* mechanics course in lecture hall settings in the 2006 fall semester at NCSU. In an attempt to evaluate these majority participants, comparisons were conducted between the pre-test scores and post-test scores. The purpose of the comparisons is to detect whether or not the course instructions introduced a change (gain) in student performance on the energy assessment.

One noteworthy aspect of such comparisons is that only paired data are meaningful, as a simple direct comparison in test scores is flawed. The major reason lies in the very fact that the two sets of data collected before and after the course instructions are *not* independent. More specifically, most students who participated in the pre-test also took the post-test. Based on this consideration, only the data collected from those students who completed both the pre-test and post-test were retained for the pair-wise comparisons.

To compare the pre and post test scores, two levels of pair-wise comparisons were carried out. One was the comparison in total scores; the other was the comparison in item scores. As a result, the former comparison yielded the change (gain) in student overall performance on the assessment, whereas the latter revealed the change (gain) in student

performance on the individual items. Additionally, the pre vs. post comparisons were conducted individually as well as collectively for the four parallel sections. Consequently, it became possible to detect the difference, if any, in the gains across the four sections taught by three different instructors; in other words, to detect whether the difference characterized mainly by the different instructors would have any effect on the gains in test scores.

As for the data collected from the other participants—*M&I* (U)TAs at NCSU, senior physics major undergraduates at NCSU, and first-year physics major graduate students at Purdue—analyses were simply more descriptive in the sense that no inferential comparisons were particularly intended. Nevertheless, a few comparisons in test scores were performed between the (U)TAs and the Purdue graduate students, as they were fairly comparable in academic level and quality.

### 3.5.2 Qualitative analyses

To uncover student reasoning on the topics addressed in the current energy assessment, a qualitative method was used through a small number of student interviews. Two main reasons motivated this approach. First, students' final answers to the multiple-choice questions cannot reflect their thought processes; students' errors manifested on the written test often need further explanations. Therefore, it is useful to resort to students' explicit verbal accounts for supporting evidence. Second, student interviews can shed light on the further improvement of the test items in the current energy assessment, particularly issues of false positives and false negatives.

To conduct the interviews, nine student interviewees were recruited from those who attended one of the four sections of the *M&I* mechanics course in lecture hall settings in the

2006 fall semester at NCSU. Before participating in the interviews, all the nine recruited students took the post-test of the current energy assessment. These nine recruited students were selected from a volunteer pool based on the following considerations. First, efforts were made to ensure that the interviewees spread across the four parallel sections. Second, priority was given to those students whose post-test scores were at different levels. As such, the interviewees were purposefully chosen to be as diverse as possible.

Each of the nine students was scheduled to interview with the author individually and privately for an hour-long session. During the session, the student interviewees were asked to think aloud their reasoning while working on the questions that were selected from the current energy assessment. Since the time constraint did not allow students to “think aloud” every question in the current energy assessment, fifteen questions were used for the interviews. These questions were Q3a, Q5, Q6a, Q6b, Q7, Q8, Q9, Q12, Q14, Q15, Q16b, Q19, Q21, Q24 and Q27. Primarily, these questions were selected in such a way that they cover the main components of the test objectives. Particularly, priority was given to questions with a low percentage of correct responses in the post-test. In all cases, the student interviewees were able to complete the task within an hour. If there was more time left, then extra questions were given; these extra questions were Q1, Q11, Q18, Q22, Q23 and Q26. Students were given one question at a time. Each question was printed on a piece of paper with enough empty space left for students to write down anything that might help them answer the question.

Generally, students were instructed to talk aloud while working, verbalizing anything that came to mind. In case of vague or unintelligible utterance, students would be asked to repeat or explain more. Very occasionally, some follow-up questions were asked, for

example, “can you tell me the difference between a point-particle system and a real system?” If students were struggling between two or more choices, they would be asked how confident they were about each choice.

Felt-tipped pens were furnished for students to use during the interview session; no textbook, notebook or calculator was allowed. The interviews were videotaped with a camera placed directly above the paper to give a clear shot of the students’ handwritings on paper. Upon completion of the interview session, each student interviewee received \$15 as a monetary compensation.

The interview tapes were then transcribed. Due to some technology problems, one interview tape had extremely poor audio quality; thus making it impossible to transcribe. As a result, eight sessions were transcribed. Parts of the transcriptions are reported in Chapter 5, along with the quantitative data, to show students’ reasoning behind their choices.

Although the transcriptions contain a lot of rich and detailed information regarding student cognitive processes, it is not the purpose of this study to validate or construct any theories of these cognitive processes. Rather, the main purpose is to uncover students’ reasoning behind their written choices, and to provide possible directions for future work on the assessment. In other words, the analysis of the transcriptions is rather exploratory; the outcomes expected from the transcriptions are meant to be a display of some examples for curious eyes. To that end, no efforts were made to quantify the qualitative information contained in the transcriptions.

## **CHAPTER 4: Result I—Results on the Energy Assessment**

This study encompasses two major components. The first component is to design a valid and reliable energy assessment that is appropriate for the *Matter & Interactions* mechanics course (Chabay & Sherwood, 2006). The second component is to evaluate student understanding of the energy topics using the energy assessment. This chapter focuses on the first component, and reports the detailed results regarding the energy assessment. These results include the following four aspects:

- (1) Evaluating the reliability and discriminatory power of the two versions of the energy assessment
- (2) Categorizing the content and cognition level of the test items in the current energy assessment using the revised Bloom's taxonomy (Anderson & Krathwohl, 2001)
- (3) Determining the reasoning steps involved in the test items of the current energy assessment
- (4) Uncovering interesting issues on the test items, such as symbolic *vs.* numerical representations of physics quantities and diagrams *vs.* no diagrams

### **4.1 Reliability and discriminatory power of the energy assessment**

#### 4.1.1 Pilot version

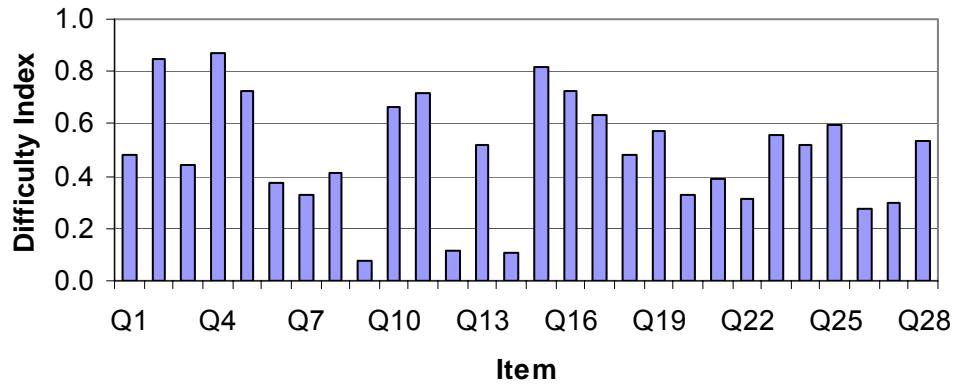
In the 2005 spring semester, the pilot version of the energy assessment was administered as an in-class pencil-and-paper test among 77 students, who were taking the *M&I* mechanics course in the SCALE-UP (Beichner *et al.* 2000) class at NCSU. Data collected from these 77 students were used to perform the item analysis and test analysis for

the pilot version of the energy assessment. For the item analysis results, Table 4.1 shows the difficulty index, discrimination index and point bi-serial coefficient values for the individual items in the pilot version of the energy assessment. Additionally, the graphical plots of these indices (coefficients) are presented in Figure 4.1 for readers' convenience.

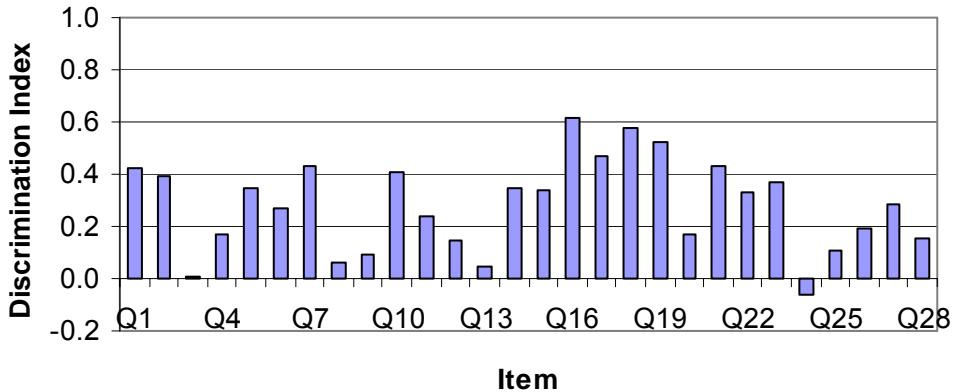
**Table 4. 1 Table 4.1 Item analysis of the pilot version of the energy assessment.**

	<b>Difficulty Index</b>	<b>Discrimination Index</b>	<b>Point Bi-Serial Coefficient</b>
<b>Q1</b>	0.48	0.42	0.31
<b>Q2</b>	0.84	0.39	0.39
<b>Q3</b>	0.44	0.01	0.07
<b>Q4</b>	0.87	0.17	0.26
<b>Q5</b>	0.73	0.34	0.25
<b>Q6</b>	0.38	0.27	0.25
<b>Q7</b>	0.32	0.43	0.31
<b>Q8</b>	0.42	0.06	0.15
<b>Q9</b>	0.08	0.09	0.07
<b>Q10</b>	0.66	0.41	0.26
<b>Q11</b>	0.71	0.24	0.29
<b>Q12</b>	0.12	0.14	0.11
<b>Q13</b>	0.52	0.04	0.05
<b>Q14</b>	0.10	0.35	0.44
<b>Q15</b>	0.82	0.34	0.35
<b>Q16</b>	0.73	0.62	0.47
<b>Q17</b>	0.64	0.47	0.37
<b>Q18</b>	0.48	0.58	0.43
<b>Q19</b>	0.57	0.52	0.40
<b>Q20</b>	0.32	0.17	0.20
<b>Q21</b>	0.39	0.43	0.37
<b>Q22</b>	0.31	0.33	0.29
<b>Q23</b>	0.56	0.37	0.34
<b>Q24</b>	0.52	-0.06	-0.05
<b>Q25</b>	0.60	0.11	0.07
<b>Q26</b>	0.27	0.19	0.16
<b>Q27</b>	0.30	0.28	0.24
<b>Q28</b>	0.53	0.16	0.12
<b>Average</b>	0.49	0.28	0.25

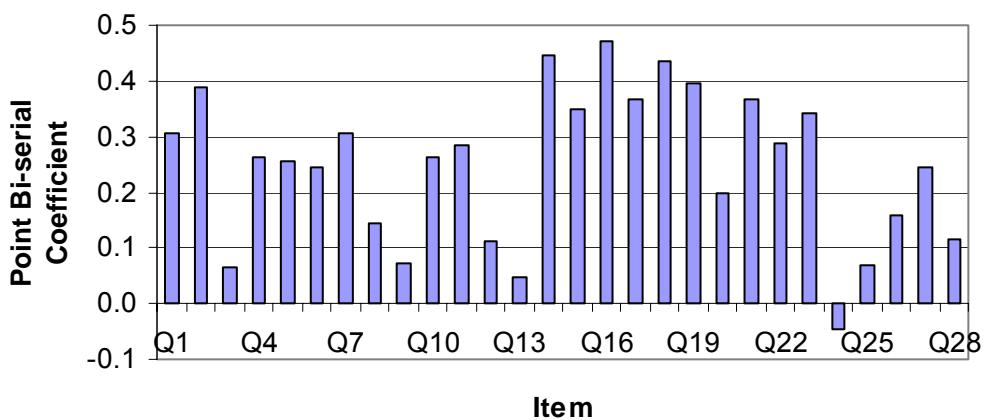
**(a) Item Difficulty Index**



**(b) Item Discrimination Index**



**(c) Item Point Bi-serial Coefficient**



**Figure 4.1 Item difficulty index, discrimination index and point bi-serial coefficient plots for the pilot version of the energy assessment.**

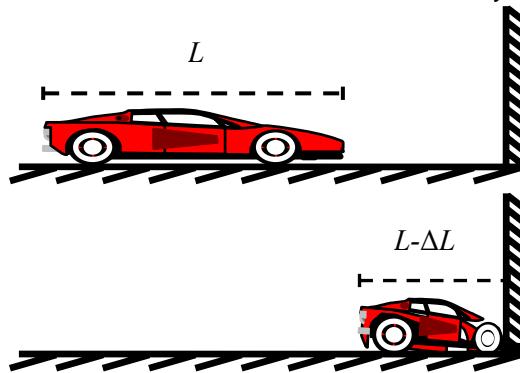
As can be seen from Figure 4.1(a), the majority of the test items in the pilot version have reasonable difficulty index ( $P$ ) values, which range from slightly below 0.3 to slightly above 0.8. Three items have noticeably low difficulty index values and thus are the most “difficult” questions. Difficult questions often deserve further attention, for it is possible that the questions may have ambiguous statements that have caused the low percentages of correct responses. Here, all these three questions are clearly stated as are shown in Figure 4.2, and it seems unlikely the question statements are confusing. Overall, the average difficulty level for the pilot version of the energy assessment is 0.49, which falls into the criterion range [0.3, 0.9] (Doran, 1980).

As for the item discrimination index ( $D$ ), Figure 4.1(b) shows that only half of the test items have the discrimination index values greater than 0.3. For the other half of the test items, the discrimination index values fluctuate around 0.15. Overall, the average item discrimination index is 0.28, which is slightly lower than the criterion  $D \geq 0.3$  (Doran, 1980). In particular, the discrimination index value for Q24 is negative ( $D = -0.06$ ), which is rather alarming. This item asks students to choose a correct symbolic expression for the energy principle in a case of a moving block compressing a low-mass spring. (See Figure 4.3.) During the subsequent expert review, this item was eliminated at the suggestion of the panel experts.

The item point bi-serial coefficient ( $r_{\text{pbi}}$ ) values are displayed in Figure 4.1(c). As one can see, for most items the point bi-serial coefficient values range between 0.2 and 0.5. Overall, the average value is 0.25, which satisfies the criterion  $r_{\text{pbi}} \geq 0.2$  (Kline, 1986). However, one noticeable feature of Figure 4.2(c) is worth mentioning. That is, the point bi-serial coefficient for Q24 is negative. This, again, indicates that item Q24 is problematic.

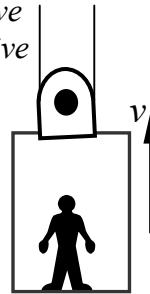
**Q9.** In a car crash test, a car of original length  $L$  hits a concrete wall and comes to a full stop without bouncing back. The wrecked car is now  $\Delta L$  shorter. During the collision, the average force applied to the car by the wall is  $F$ . What is the work done on the car by the wall?

- (a)  $F \cdot \Delta L$
- (b)  $F \cdot \Delta L / 2$
- (c)  $-F \cdot \Delta L$
- (d)  $-F \cdot \Delta L / 2$
- (e) Zero work



**Q12.** A person is standing in an elevator, which is moving upward at a constant speed. Consider the person as a system. As the elevator goes between the 3<sup>rd</sup> and 4<sup>th</sup> floor,

- (a) The total work done on the person by the Earth and the elevator is zero
- (b) The total work done on the person by the Earth and the elevator is positive
- (c) The total work done on the person by the Earth and the elevator is negative
- (d) Not enough information to determine



**Q14.** Consider a falling stone as a system. At the instant when the stone is several meters above the ground, the stone has:

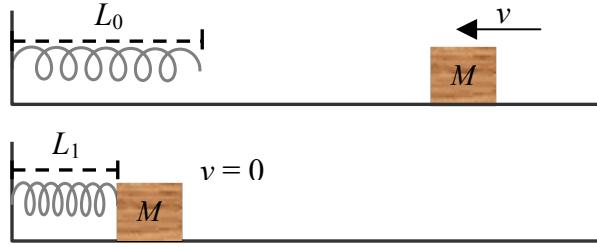
- (a) Kinetic energy only
- (b) Rest energy only
- (c) Kinetic energy and gravitational potential energy
- (d) Kinetic energy and rest energy
- (e) Kinetic energy, rest energy, and gravitational potential energy

Figure 4.2 Three items with the lowest difficulty index values in the pilot version of the energy assessment.

A light spring of stiffness  $k_s$  and relaxed length  $L_0$  has one end attached to a wall. A block of mass  $M$ , moving with speed  $v$  on a surface with negligible friction, approaches and compresses the spring, until it comes to a full stop. The spring is compressed to a length  $L_1$ . During the process there is no temperature change.

**Q24.** If we take the spring and block as the system, which of the following is correct?

- (a)  $\frac{1}{2}k_s(L_0 - L_1)^2 = \frac{1}{2}Mv^2$
- (b)  $\frac{1}{2}k_s(L_0 - L_1)^2 = -\frac{1}{2}Mv^2$
- (c)  $\frac{1}{2}k_sL_0^2 = \frac{1}{2}Mv^2 + \frac{1}{2}k_sL_1^2$
- (d)  $Mv = k_s(L_0 - L_1)$
- (e)  $Mv = k_sL_1$



**Figure 4.3** An item with a negative discrimination index and point bi-serial coefficient in the pilot version of the energy assessment.

In addition to the above item analysis, a test analysis for the pilot version of the energy assessment yields another two indices—KR-20 reliability ( $r_{\text{test}}$ ) and Ferguson's delta ( $\delta$ ). The reliability value for the pilot version of the energy assessment is 0.6, which is lower than the standard criterion for group measurement  $r_{\text{test}} \geq 0.7$  (Doran, 1980). As for the Ferguson's delta ( $\delta$ ), the value for the pilot version is 0.9, which meets the standard  $\delta \geq 0.9$ . (Kline, 1986)

In brief, Table 4.2 summarizes the item and test analysis results for the pilot version of the energy assessment.

**Table 4.2 Evaluation of the pilot version of the energy assessment.**

<b>Test statistics</b>	<b>Desired values</b>	<b>Values for the pilot version</b>
Difficulty Index	[0.3, 0.9]	Average 0.49
Discrimination Index	$\geq 0.3$	Average 0.28
Point bi-serial coefficient	$\geq 0.2$	Average 0.25
Reliability Index	$\geq 0.7$	0.6
Ferguson's Delta	$\geq 0.9$	0.9

#### 4.1.2 Current version

In total, 389 student participants completed the current version of the energy assessment in the 2006 fall semester. (See Table 4.3.) Data collected from these student participants are used to perform the item analysis and test analysis for the current version of the energy assessment. One noteworthy aspect about the analysis is that the pre-test data are not included. The reason lies in the fact that students, prior to the course instructions, often do not have a good mastery of the subject matter and are likely to guess on most questions, if not all questions; thus their pre-test performance may not be as reliable as their post-test performance. Since the focus of item and test analysis is on the assessment rather than on the students, it is reasonable only to use the post-test data, which are more consistent than the pre-test data, to evaluate the reliability and discriminatory power of the assessment.

**Table 4.3 Student participants whose data were used for the item analysis and test analysis of the current version of the energy assessment.**

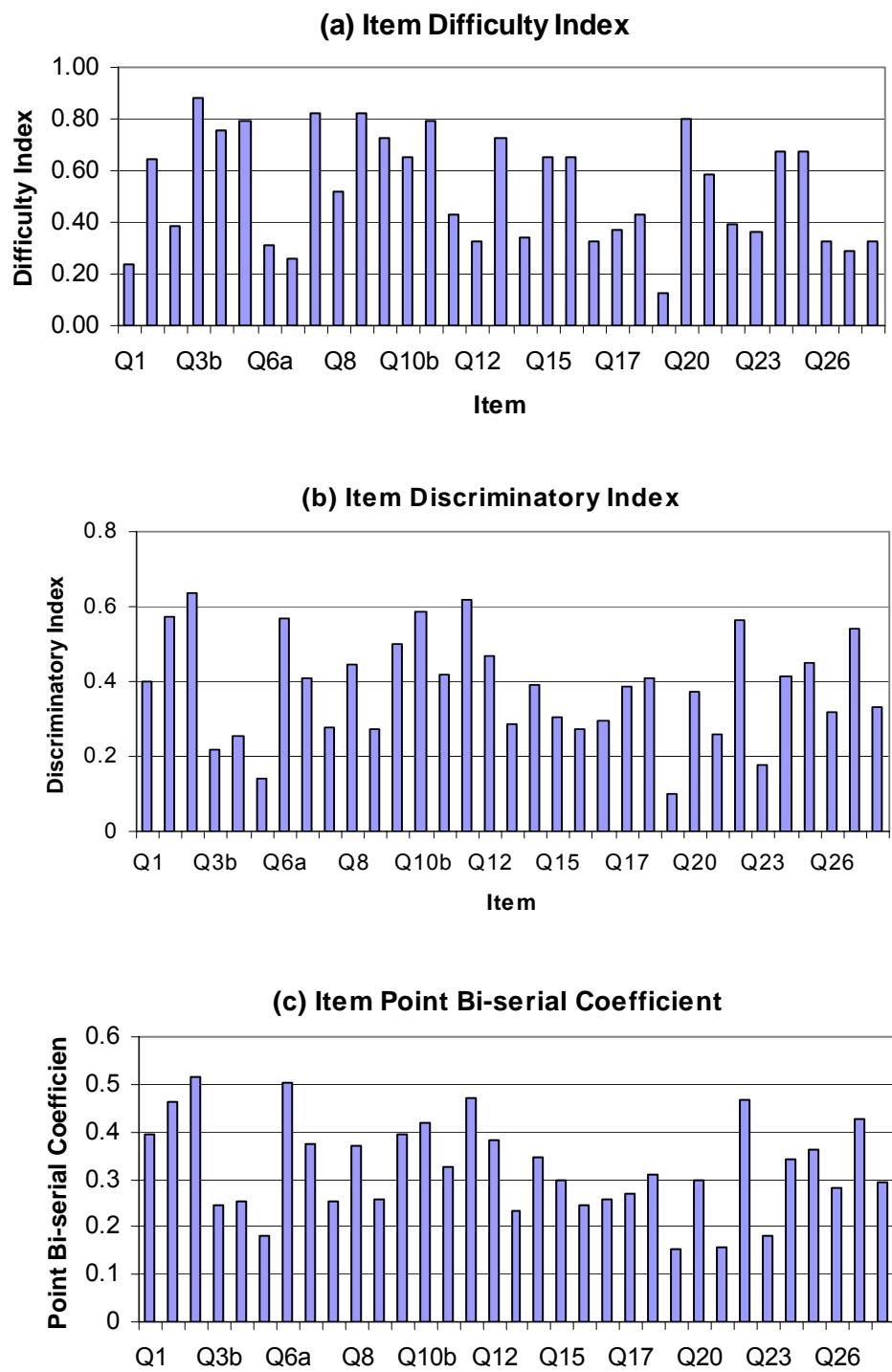
Participants	Number of Participants	Institution
Students from M&I lecture hall class (post-test only)	308	NCSU
M&I (U)TAs	16	NCSU
Students from PY401/501	36	NCSU
First-year physics graduate students	29	Purdue
<u>Total number of participants</u>	<u>389</u>	

For the item analysis results, Table 4.4 shows the difficulty index, discrimination index and point-bi serial coefficient values for the individual items in the current energy assessment. Additionally, Figure 4.4 displays the graphical plots of these indices (coefficients).

As shown in Figure 4.4 (a), the item difficulty index ( $P$ ) values vary from 0.13 to 0.86 with most items between 0.3 and 0.8. Among all the items, Q19 apparently is the most “difficult” item in the current energy assessment. This item is essentially the same as Q9 in the pilot version as shown in Figure 4.2, except for an explicit indication of a car being a real system. Overall, the average item difficulty index value is 0.53, which falls into the criterion range  $P \in [0.3, 0.9]$  (Doran, 1980). This result indicates that items in the current energy assessment have fairly reasonable difficulty levels.

**Table 4.4 Item analysis of the current version of the energy assessment.**

Item	Difficulty Index	Discrimination Index	Point Bi-serial Coefficient
<b>Q1</b>	0.24	0.40	0.39
<b>Q2</b>	0.65	0.57	0.46
<b>Q3a</b>	0.39	0.64	0.52
<b>Q3b</b>	0.88	0.22	0.24
<b>Q4</b>	0.76	0.25	0.25
<b>Q5</b>	0.79	0.14	0.18
<b>Q6a</b>	0.31	0.57	0.50
<b>Q6b</b>	0.26	0.41	0.37
<b>Q7</b>	0.82	0.28	0.26
<b>Q8</b>	0.52	0.45	0.37
<b>Q9</b>	0.83	0.27	0.26
<b>Q10a</b>	0.72	0.50	0.39
<b>Q10b</b>	0.65	0.59	0.42
<b>Q10c</b>	0.79	0.42	0.32
<b>Q11</b>	0.43	0.62	0.47
<b>Q12</b>	0.32	0.47	0.38
<b>Q13</b>	0.73	0.29	0.23
<b>Q14</b>	0.34	0.39	0.35
<b>Q15</b>	0.65	0.30	0.30
<b>Q16a</b>	0.66	0.27	0.25
<b>Q16b</b>	0.33	0.30	0.26
<b>Q17</b>	0.37	0.39	0.27
<b>Q18</b>	0.43	0.41	0.31
<b>Q19</b>	0.13	0.10	0.15
<b>Q20</b>	0.80	0.37	0.30
<b>Q21</b>	0.59	0.26	0.16
<b>Q22</b>	0.39	0.56	0.47
<b>Q23</b>	0.36	0.18	0.18
<b>Q24</b>	0.67	0.41	0.34
<b>Q25</b>	0.67	0.45	0.36
<b>Q26</b>	0.33	0.32	0.28
<b>Q27</b>	0.29	0.54	0.43
<b>Q28</b>	0.32	0.33	0.29
<b>Average</b>	0.53	0.38	0.33



**Figure 4.4 Item difficulty index, discrimination index and point bi-serial coefficient plots for the current energy assessment.**

For item discrimination index ( $D$ ), Figure 4.4 (b) shows the values for the individual items in the current energy assessment. As one can see, the discrimination index values vary from 0.1 to over 0.6 with most items being near or above the standard 0.3. Among all the items, Q5 and Q19 are rather conspicuous due to their lowest discrimination index values. It is, however, not surprising considering the rather high difficulty level for Q5 ( $P = 0.79$ ) and low difficulty level for Q19 ( $P = 0.13$ ). As to the average item discrimination index, the calculated value is 0.38 that is higher than the criterion  $D \geq 0.3$  (Doran, 1980). This average value is considered to be satisfactory, for it indicates that the overall item discrimination power of the current energy assessment is acceptable.

Figure 4.4(c) shows the point bi-serial coefficient plot for the items in the current energy assessment. In this plot, all the items display a satisfactory value being near or above the standard 0.2. As a result, the average item point bi-serial coefficient is 0.33, which is above the criterion  $r_{\text{pbi}} \geq 0.2$  (Kline, 1986). Clearly, this result suggests that the overall consistency of the items with the entire assessment is satisfactory.

As for the entire assessment, the test analysis of the current energy assessment yields the KR-20 reliability index 0.74 and the Ferguson's delta 0.97. These two values are both higher than their respective criteria. Therefore, the current version of the energy assessment is considered to be a reliable and discriminatory assessment.

In sum, Table 4.5 shows the item/test analysis results for the current version of the energy assessment.

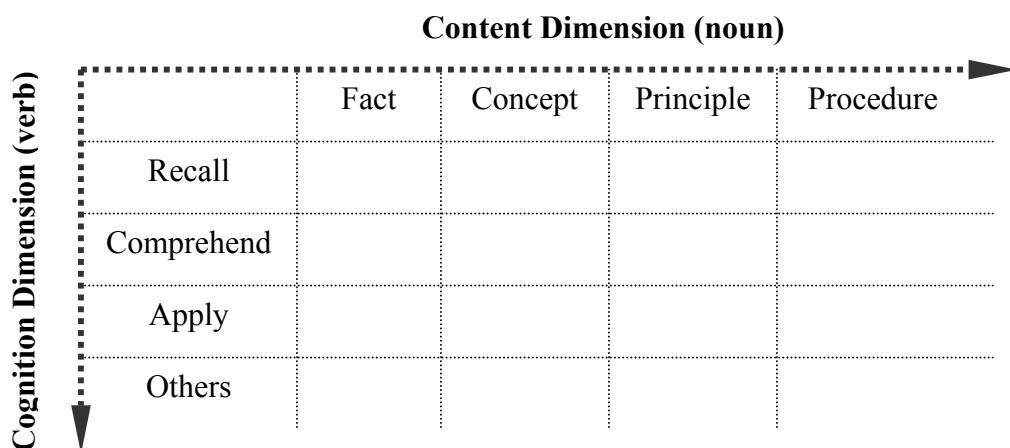
**Table 4.5 Evaluation of the current version of the energy assessment.**

<b>Test statistics</b>	<b>Desired values</b>	<b>Values for the current energy assessment</b>
Difficulty Index	[0.3, 0.9]	Average 0.49
Discrimination Index	$\geq 0.3$	Average 0.38
Point bi-serial coefficient	$\geq 0.2$	Average 0.33
Reliability Index	$\geq 0.7$	0.74
Ferguson's Delta	$\geq 0.9$	0.97

## **4.2 Content and cognition level of the test items**

The major motivation for categorizing the test items in terms of the content and cognition levels originates from such an intention that the energy assessment should not merely measure low-level thinking, i.e. recalling a formula. Rather, it is hoped that the items in the energy assessment measure higher-level thinking, i.e. applying a concept or a principle. In other words, the energy assessment is not designed to test whether or not students can recite a concept or a principle verbatim, but to test whether or not students can use the appropriate concepts or principles to explain or predict physical phenomena. To this end, it is useful to classify the individual items in terms of the content and cognition levels. In so doing, it becomes possible to detect any undesirable items as well as to determine the overall content and cognition level for the energy assessment.

In this study, the revised Bloom's taxonomy (see, for example, Anderson & Krathwohl, 2001; Krathwohl, 2002.) is employed to classify the test items in the current energy assessment. As opposed to the original Bloom's taxonomy (Bloom *et al.* 1956) where the classification of test items is of one dimension, the revised taxonomy is composed of two dimensions—content and cognition. (See Figure 4.5.) These two separate dimensions embody the noun aspect and the verb aspect of a test item, respectively. Specifically, the content dimension describes “what” (noun), whereas the cognition dimension addresses “how” (verb). For example, if a test item requires one to apply the energy principle, then the content level for this item is “principle” (noun), and the cognition level is “apply” (verb).



**Figure 4.5 The revised Bloom's taxonomy (See, for example, Haladyna, 1999; Anderson & Krathwohl, 2001; Krathwohl, 2002.)**

Within the content dimension, classification solely focuses on the type of knowledge involved in the test items. Primarily, there exist four types of knowledge in the content dimension, which, from the lowest level to the highest level, include: facts, concepts, principles and procedures. (See, for example, Haladyna, 1999; Anderson & Krathwohl, 2001.) In general, the first three levels are best suitable for multiple-choice tests, while the fourth level—procedure—is best measured through performance tests. In this particular study, no items in the current energy assessment are intentionally designed to test any procedural knowledge, and the first three types of knowledge are thus sufficient to categorize the item content levels. Table 4.6 shows the detailed descriptions for each type of knowledge. These descriptions are used to categorize the content levels for items in the current energy assessment.

Within the cognition dimension, on the other hand, classification is fully based on the cognitive behaviors that are required in the test items. These cognitive behaviors, from lower to higher level, include: recall, comprehend, apply, and even higher-level behaviors, such as synthesize and create. For the first three cognition levels—recall, comprehend and apply, most researchers agree upon both the names and order. (See, for example, Haladyna, 1999; Anderson & Krathwohl, 2001.) Nevertheless, for the other cognition levels, different researchers seem to find no clear agreement on what to name and how to order. (See, for example, Anderson & Sosniak, 1994; Haladyna, 1997; Linn & Gronlund, 2001; Aderson & Krathwohl, 2001.) In this study, the first three cognitive behaviors suffice to determine the item cognition levels in the current energy assessment, as no items are designed purposefully to require such behaviors as synthesis or creation. Therefore, only the first three levels are used, and their detailed descriptions are provided in Table 4.7.

**Table 4.6 The content dimension of the revised Bloom's taxonomy. (See, for example, Haladyna, 1999; Anderson & Krathwohl, 2001.)**

Content level	Description	Example
<b>Fact</b>	<p>Facts are discrete isolable content elements (smallest pieces and bits of information), including terminology, specific details and elements.</p>	<p>Scientific specifics, terms and standard symbols are in this category. For example, “the speed of light is <math>3 \times 10^8</math> m/s”, “protons have positive charge”, “<math>v</math> is speed while <math>\vec{v}</math> is velocity”</p>
<b>Concept</b>	<p>Concepts are discrete bits of information and the linking connections among them. This type of knowledge includes schemes, models, and criteria. It is a more complex and organized knowledge than facts.</p>	<p>Most physics quantities are concepts, in that they are often defined in terms of, and are connected to, other quantities. For example: momentum in relation with velocity, kinetic energy of a multi-particle system, graphical representation of potential energy</p>
<b>Principle</b>	<p>Principles are statements of relationship, usually between two or more concepts. A principle often takes the form “if... then...”. In science, most principles (theorems) are immutable laws.</p>	<p>For example, the momentum principle and the energy principle.</p>

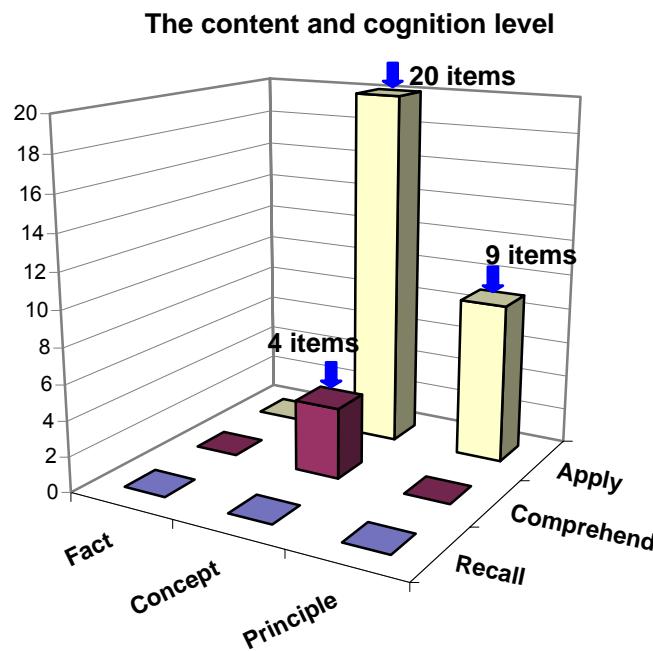
**Table 4.7 The cognition dimension of the revised Bloom's taxonomy. (See, for example, Haladyna, 1999; Anderson & Krathwohl, 2001.)**

Cognition level	Description	Example
<b>Recall</b>	This task requires an individual to retrieve relevant information from memories. The relevant information can be facts, concepts or principles. This task is a process of literal reiteration and does not require understanding at any level.	For example, write down the mass of an electron, or recite the relativistic expression of momentum
<b>Comprehend</b>	This task involves the interpretation, summarization, translation, or paraphrasing of information in a different manner from that in the original form. It is not equivalent to a fullest grasp of the information.	For example, interpret the graphical representations of physics quantities
<b>Apply</b>	This task involves the actual or described use of relevant information either to perform exercises or to solve problems in a <i>particular situation (scenario)</i> . It is a demonstration of comprehension.	For examples, use the work concept to figure out the work done on a given system, or relate work to a change in energy of a given system

Based on the descriptions in Table 4.6 and Table 4.7, test items in the current energy assessment were categorized independently by two researchers—the author and a physics professor at NCSU. Since three content levels and three cognition levels can yield nine different categories, both researchers decided only to focus on one dimension at a time to reduce possible confusion. For convenience, all the items were first categorized in terms of the content levels, and were then categorized in terms of the cognition levels. In an attempt to determine the agreement between the two researchers on the item content and cognition levels, the inter-rater reliability was calculated for both dimensions. For the content dimension, the inter-rater reliability is 94%; for the cognition dimension, the inter-rater reliability is 97%. Discussions between the two researchers on the disagreed items followed, and the disagreement was eventually resolved. As for the final results, Table 4.8 shows both the content and cognition levels for the individual items in the current energy assessment. Additionally, Figure 4.6 shows the 3-D bar chart for the overall results.

**Table 4.8 The content and cognition level for the individual items in the current energy assessment.**

<b>Item</b>	<b>Content</b>	<b>Cognition</b>	<b>Item</b>	<b>Content</b>	<b>Cognition</b>
<b>Q1</b>	Concept	Apply	<b>Q14</b>	Concept	Apply
<b>Q2</b>	Concept	Apply	<b>Q15</b>	Concept	Apply
<b>Q3a</b>	Concept	Apply	<b>Q16a</b>	Concept	Apply
<b>Q3b</b>	Concept	Apply	<b>Q16b</b>	Principle	Apply
<b>Q4</b>	Concept	Apply	<b>Q17</b>	Concept	Apply
<b>Q5</b>	Concept	Apply	<b>Q18</b>	Concept	Apply
<b>Q5a</b>	Concept	Apply	<b>Q19</b>	Concept	Apply
<b>Q5b</b>	Principle	Apply	<b>Q20</b>	Concept	Apply
<b>Q7</b>	Concept	Apply	<b>Q21</b>	Concept	Apply
<b>Q8</b>	Concept	Apply	<b>Q22</b>	Principle	Apply
<b>Q9</b>	Concept	Apply	<b>Q23</b>	Principle	Apply
<b>Q10a</b>	Concept	Comprehend	<b>Q24</b>	Principle	Apply
<b>Q10b</b>	Concept	Comprehend	<b>Q25</b>	Principle	Apply
<b>Q10c</b>	Concept	Comprehend	<b>Q26</b>	Principle	Apply
<b>Q11</b>	Concept	Comprehend	<b>Q27</b>	Principle	Apply
<b>Q12</b>	Concept	Apply	<b>Q28</b>	Principle	Apply
<b>Q13</b>	Concept	Apply			



**Figure 4.6 A 3-D bar chart of the content and cognition levels for the items in the current energy assessment.**

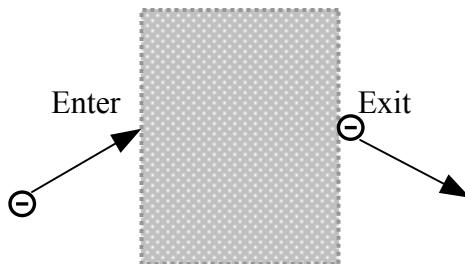
As one can see, all items in the current energy assessment target at either concepts or principles, and most items require the application of these concepts or principles. Specifically, the majority of the items, which account for 61% of the entire assessment, involve the application of appropriate concepts. For instance, Q2 tests rest energy in relation with mass and requires students to apply this concept in a comparison of rest energy between an electron and a proton. (See Figure 4.7.)

- Q2.** *A proton and an electron are moving at the same speed. Which one has greater rest energy?*
- Proton, because it has greater mass*
  - Electron, because it has less mass*
  - They have the same nonzero rest energy, because both are traveling at the same speed*
  - They both have zero rest energy, because neither is at rest*
  - None of the above*

**Figure 4.7 An item in the current energy assessment that requires the application of concept.**

Another nine items, which account for 27% of the entire assessment, require the application of correct principles, particularly the energy principle. Except for one item Q26 (see Figure 4.8), where students need to apply both the energy principle and the momentum principle, all items in this category require only the application of the energy principle.

**Q26.** An electron with speed  $1 \times 10^8$  m/s enters a box along a direction depicted in the diagram. Some time later the electron is observed leaving the box with the same speed  $1 \times 10^8$  m/s but different direction as shown. From this, we can conclude that in the box:



- (a) The net force on the electron was nonzero, and the net work done on the electron was nonzero
- (b) The net force on the electron was nonzero, but the net work done on the electron was zero
- (c) The net force on the electron was zero, but the net work done on the electron was nonzero
- (d) The net force on the electron was zero, and the net work done on the electron was zero
- (e) Not enough information to determine

**Figure 4.8** An item in the current energy assessment that requires the application of both the momentum principle and the energy principle.

In addition to the items mentioned above, a few items fall into the “comprehension of concepts” category. In fact, items in this category all require students to interpret the graphical representations of energy. For example, Q10(a) asks students to name the kind of energy that is represented by one curve in a graph, and Q10(c) asks students to identify the line that represents a bound state. (See Figure 4.9.)

Consider a system that consists of two asteroids in deep space. The following figure plots energy of several different states of the two-asteroid system versus distance  $r$  between them. Questions 10a to 10c refer to this figure.

**Q10a.** What does the curved line I represent?

- (a) Kinetic energy
- (b) Potential energy
- (c) Sum of kinetic energy and potential energy
- (d) Rest energy
- (e) None of the above

**Q10c.** Which of the horizontal lines represents a bound state?

- (a) Line II
- (b) Line III
- (c) Line IV
- (d) All of the above
- (e) None of the above

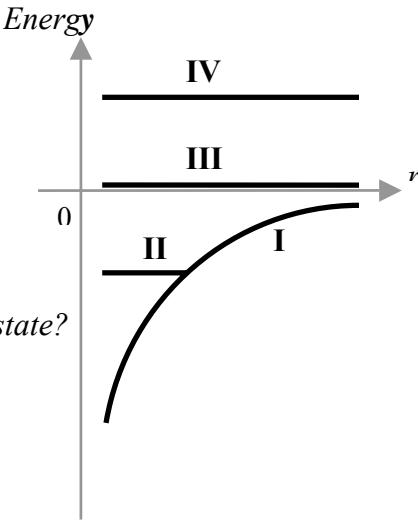


Figure 4.9 Items that require the comprehension of concepts.

In short, the current version of the energy assessment by and large tests the application of concepts and principles. Unlike items that ask students to recall specific formulas or concepts verbatim, items in the current energy assessment aim at higher-level thinking both in content and cognition.

### 4.3 Reasoning steps

Although it is now clear that the items in the current energy assessment require primarily the application of concepts and principles, it still remains unknown as to how many reasoning steps are involved in the individual items. In fact, the number of reasoning steps involved in test items is rather important for the multiple-choice test format, as it often determines how difficult or easy the interpretation of the test results will be. For multiple-choice items where long-chain reasoning is required, it is often difficult, if not completely

impossible, to pinpoint at which particular step students fail or succeed. If, on the other hand, a multiple-choice item requires only a short reasoning process, then the interpretation of student performance on this item becomes relatively more specific and convincing.

For items in the current energy assessment, it is hoped that the reasoning steps required are of a small number. In other words, long-chain reasoning was intentionally avoided in designing the energy assessment items, so that the subsequent interpretation of student performance is more meaningful. As a practical check, the following schemes are designed to determine the number of reasoning steps required for each item in the current energy assessment.

In general, if an item provides information on a specific quantity X and asks for information on another quantity Y, then the number of reasoning steps for this item is determined by the number of extra quantities needed to relate X and Y. To be more specific, consider the following situation. Suppose the given quantity X cannot directly generate information on the quantity Y, but can directly generate information on the quantity Z that is not given in the item. This quantity Z, however, can directly generate information on Y, which is the final answer. Then the number of reasoning steps from X to Y is one; equal to the number of extra quantities needed—only one in this case. Simply put, if  $X \leftrightarrow Z \leftrightarrow Y$  is the relation among X, Y, and Z, then the number of steps needed to obtain the answer Y from X is one.

If the given quantity X can directly generate information on the unknown quantity Y through a single relation, a formula for instance, without needing a third quantity as a bridge, then the number of reasoning steps is zero. In other words, if the given quantity X is in a

direct relation with the unknown quantity Y as  $X \leftrightarrow Y$ , then the number of steps needed to obtain an answer on Y from X is zero.

In particular, for qualitative items that require the following actions, the number of steps is regarded as zero. These actions include identifying the kinds of energy or the state of a given system, indicating the interactions between two bodies, naming the objects that do work on a system, choosing a system, and comparing qualitatively two entities.

In addition, interpreting or determining the relevancy of given information is not counted as a step. For example, the following actions do not pose a step: interpreting a system being far away from other objects as no external force, determining spring potential energy as independent of the original length of a spring.

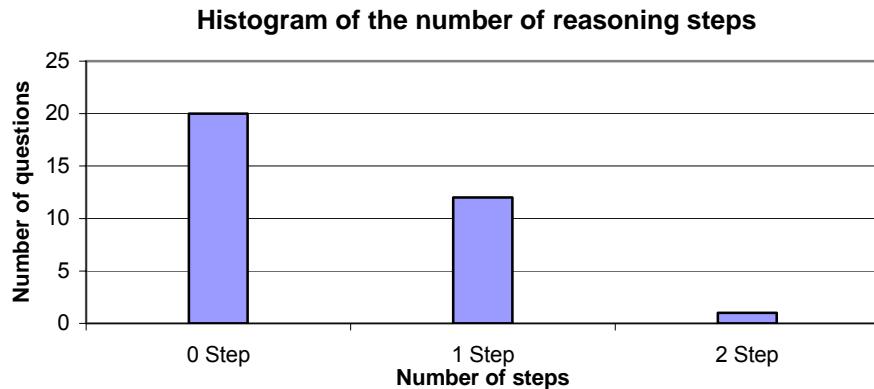
Moreover, a repetition of the same action is not counted as a reasoning step. For example, if an item requires one to calculate both the initial and final kinetic energy ( $K_i$  &  $K_f$ ) of an object given the initial and final speed ( $v_i$  &  $v_f$ ), the action of calculating  $K_f$  is regarded as a repetition of calculating  $K_i$  and is not counted as a new step. Furthermore, for any numerical calculation that involves only pure math without physics, it is not counted as a step. A final caveat is that if there exists more than one approach to answer an item, only the approach of the least number of reasoning steps is considered.

Based on these schemes, two researchers—the author and a physics graduate student at NCSU—individually determined the number of reasoning steps for each item in the current energy assessment. Results from both researchers were compared. It was found that both researchers agreed on 91% of the items in terms of the number of reasoning steps. For the other items, the disagreement on the number of steps differed only by 1. After discussions and clarifications of the schemes, the disagreement was resolved between the two

researchers. As for the results, Table 4.9 shows the number of steps for each item in the current energy assessment. Additionally, Figure 4.10 displays a histogram for the overall results.

**Table 4.9 The number of reasoning steps for each item in the current energy assessment.**

<b>Item</b>	<b>Number of steps</b>	<b>Item</b>	<b>Number of steps</b>
<b>Q1</b>	0	<b>Q14</b>	0
<b>Q2</b>	0	<b>Q15</b>	1
<b>Q3a</b>	1	<b>Q16a</b>	0
<b>Q3b</b>	1	<b>Q16b</b>	1
<b>Q4</b>	0	<b>Q17</b>	0
<b>Q5</b>	0	<b>Q18</b>	1
<b>Q6a</b>	0	<b>Q19</b>	0
<b>Q6b</b>	1	<b>Q20</b>	1
<b>Q7</b>	0	<b>Q21</b>	1
<b>Q8</b>	0	<b>Q22</b>	1
<b>Q9</b>	0	<b>Q23</b>	1
<b>Q10a</b>	0	<b>Q24</b>	0
<b>Q10b</b>	0	<b>Q25</b>	0
<b>Q10c</b>	0	<b>Q26</b>	2
<b>Q11</b>	0	<b>Q27</b>	1
<b>Q12</b>	0	<b>Q28</b>	1
<b>Q13</b>	0		



**Figure 4.10 A histogram of the number of reasoning steps for the items in the current energy assessment.**

Clearly, the majority of the energy assessment items require zero-step reasoning that can be formulated as “given  $X \leftrightarrow$  unknown  $Y$ ”. For clarification, it is worthwhile remembering that the schemes used in this study do not, and by no means intend to, capture every cognitive process that is necessary for one to be able to answer a question correctly. Some cognitive reasoning processes, such as interpretation of given information, are not considered as a “step” in the schemes. As a result, zero-step reasoning does not indicate no reasoning at all. Rather, it reflects a short reasoning process that requires no extra quantities or information to bridge the given with the unknown. As an example, the item Q2 in the current energy assessment is of zero-step reasoning. This item provides such information that “*a proton and an electron are moving at the same speed*” and asks for “*which one has greater rest energy*”. To answer this question, one needs to relate mass with rest energy as formulated above “ $m \leftrightarrow$  rest energy” and be aware that rest energy is independent of speed. Since the definition of rest energy relates mass directly to rest energy and no extra quantities are needed, this item, according to the schemes, is regarded of zero-step reasoning. As for determining the relevancy of the given information “moving at the same speed”, comparing

the mass of an electron with that of a proton, or comparing the rest energy of an electron with that of a proton, they are not counted as a step according to the aforementioned schemes.

Aside from the majority items that are of zero-step reasoning, twelve items require one-step reasoning that can be formulated as “given X  $\leftrightarrow$  extra Z  $\leftrightarrow$  unknown Y”. An example from the current energy assessment is item Q23. This item gives the following information: “*In outer space an asteroid was initially moving with a constant speed v in a certain direction. Some time later, the same asteroid was observed to be moving with the same constant speed v in the opposite direction.*” Then this item asks for the total external work done on the asteroid between the two times. Here, from the given information  $v_i$  and  $v_f$  to the final answer  $W_{\text{ext}}$ , one needs an extra piece  $\Delta K (=K_f - K_i)$  as a bridge to link the given with the unknown. Simply put, the relations among  $v_i$  and  $v_f$ ,  $W_{\text{ext}}$ , and  $\Delta K$  can be formulated as  $v_i$  and  $v_f \leftrightarrow \Delta K \leftrightarrow W_{\text{ext}}$ ; therefore it is a one-step reasoning process according to the schemes mentioned above. Of note here is the repetition action of calculating  $K_i$  and  $K_f$ , and the simple math for calculating the difference between  $K_i$  and  $K_f$ . So essentially only one step is required in this case.

Of all the items in the current energy assessment, one item (Q26) is particularly conspicuous, for it is the only item that involves a two-step reasoning process. Indeed, this item (see Figure 4.8) requires one to relate  $\vec{v}_i$  and  $\vec{v}_f$  to both  $\vec{F}_{\text{net}}$  and  $W_{\text{ext}}$  through the energy and momentum principles. To do so, one has to have both  $\Delta\vec{p}$  and  $\Delta K$  to be able to determine  $\vec{F}_{\text{net}}$  and  $W_{\text{ext}}$ . In other words, the relations among these quantities can be formulated as  $\vec{v}_i$  and  $\vec{v}_f \leftrightarrow \Delta\vec{p} \leftrightarrow \vec{F}_{\text{net}}$  and  $v_i$  and  $v_f \leftrightarrow \Delta K \leftrightarrow W_{\text{ext}}$ . So, according to the aforementioned schemes, the number of the reasoning steps for this item is two.

In brief, items in the current energy assessment generally require short reasoning processes. Specifically, all but one of the items require no more than one step reasoning. Only one item involves a two-step reasoning process, in which both the energy principle and the momentum principle are needed.

#### **4.4 Issues on symbolic/numerical representations and diagrams**

Some issues about the symbols, numbers and diagrams in the test items arose during the development of the energy assessment. These issues mainly pertained to whether or not different representations—symbols *vs.* numbers, diagrams *vs.* no diagrams—would affect student performance on the multiple-choice items in the energy assessment. Admittedly, no systematic studies other than exploratory experiments were carried out, nor were definite answers obtained to address these issues. As a result, these issues, to a large extent, still remain unresolved.

In this section, reported are two relevant exploratory experiments and some preliminary results. One of the experiments aims at whether students' responses to the symbolic items and to the numerical items are different. The other experiment explores how a diagram of one energy assessment item affects the students' responses.

##### 4.4.1 Symbolic representation *vs.* numerical representation

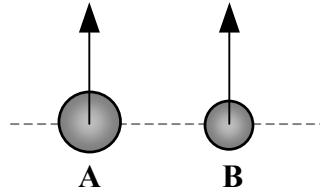
In physics, quantities are often expressed in either numbers or symbols. When designing physics questions, some instructors are more inclined to use symbols than numbers to represent physics quantities; some are in favor of numbers over symbols. Many others who have no particular preferences often adopt both representations to write physics questions.

Since whether using symbols or numbers to represent physics quantities is largely dependent on the writers' personal preferences, a related issue arises; that is, whether items written with different representations—symbolic or numerical—impose any difference on student performance. For convenience, two similar items that differ only in the representations of physics quantities are called “equivalent items”. Other than the fact that one item uses symbols to represent all the quantities while the other item uses numbers, these two “equivalent items” are otherwise identical.

In this experiment, two sets of equivalent items were designed. (See Figure 4.11 and Figure 4.12.) Within each set there are three multiple-choice items. These three items are either all symbolic questions or all numerical questions. Here, the set with all symbolic questions is called symbolic version; the other set with all numerical questions is called numerical version. Except for the representation difference, these two versions are otherwise identical. Figure 4.11 shows the symbolic version of these questions and Figure 4.12 shows the numerical version of these questions.

**Q1.** Two balls A and B are launched vertically upward from the same height at the same time. Both balls have the same initial kinetic energy ( $K_A = K_B$ ). Ball A is heavier than ball B ( $m_A > m_B$ ). Ignore air resistance. Compare the maximum heights of these two balls. Which ball can reach higher?

- (a) Ball A
- (b) Ball B
- (c) Both balls can reach the same height
- (d) Not enough information to determine



**Q2.** An object of mass  $m$  is initially at rest. A constant net force of magnitude  $F$  is applied to the object during a first time interval, and the object's speed increases from 0 to  $v$ . The constant net force continues to act on the object for a second time interval, and the object's speed increases further from  $v$  to  $2v$ . Consider the work done on the object. During which time interval is more work done on the object?

- (a) The first time interval
- (b) The second time interval
- (c) The work done during the two time intervals is the same
- (d) Not enough information to determine

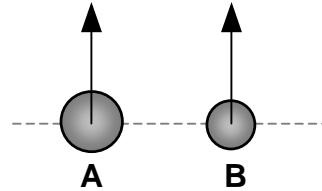
**Q3.** Two identical springs A and B have the same relaxed lengths of  $L_0$ . Spring A is stretched by  $|\Delta L|$ . Spring B is compressed by  $|\Delta L|$ . Compare the spring potential energy of spring A and spring B. Which one has greater spring potential energy?

- (a) Spring A
- (b) Spring B
- (c) Both springs have equal amount of spring potential energy
- (d) Not enough information to determine

Figure 4.11 The symbolic version of the quiz.

**Q1.** Two balls A and B are launched vertically upward from the same height at the same time. Both balls have the same initial kinetic energy 20 J. Ball A has mass 3 kg and ball B has mass 2 kg. Ignore air resistance. Compare the maximum heights of these two balls. Which ball can reach higher?

- (a) Ball A
- (b) Ball B
- (c) Both balls can reach the same height
- (d) Not enough information to determine



**Q2.** A 3 kg object is initially at rest. A constant net force of magnitude 10 N is applied to the object during a first time interval, and the object's speed increases from 0 m/s to 5 m/s. The constant net force continues to act on the object for a second time interval, and the object's speed increases further from 5 m/s to 10 m/s. Consider the work done on the object.

During which time interval is more work done on the object?

- (a) The first time interval
- (b) The second time interval
- (c) The work done during the two time intervals is the same
- (d) Not enough information to determine

**Q3.** Two identical springs A and B have relaxed lengths of 5 cm. Spring A is stretched by 2 cm. Spring B is compressed by 2 cm. Compare the spring potential energy of spring A and spring B. Which one has greater spring potential energy?

- (a) Spring A
- (b) Spring B
- (c) Both springs have equal amount of spring potential energy
- (d) Not enough information to determine

Figure 4.12 The numerical version of the quiz.

These two versions of questions were administered as a voluntary in-lab quiz among students who attended the *M&I* mechanics labs at NCSU in the 2005 fall semester. In that semester, over ten lab sections were offered. In each lab section, the instructor randomly distributed the two versions of quiz among the student volunteers. Each student volunteer would take either the symbolic version quiz or the numerical version quiz. In total, 236 engineering and non-physics science majors participated in completing the quiz. Among these 236 non-physics majors, 123 students took the symbolic version, and 113 participants took the numerical version. Additionally, 21 physics majors completed the quiz as well. Among them, 10 students took the symbolic version, and 11 students took the numerical version (see Table 4.10). Regardless of the major, all student volunteers were taking the *M&I* mechanics course in that semester and had studied the relevant materials on energy before taking the quiz.

**Table 4.10 Students who participated in taking either the symbolic or numerical version of the quiz.**

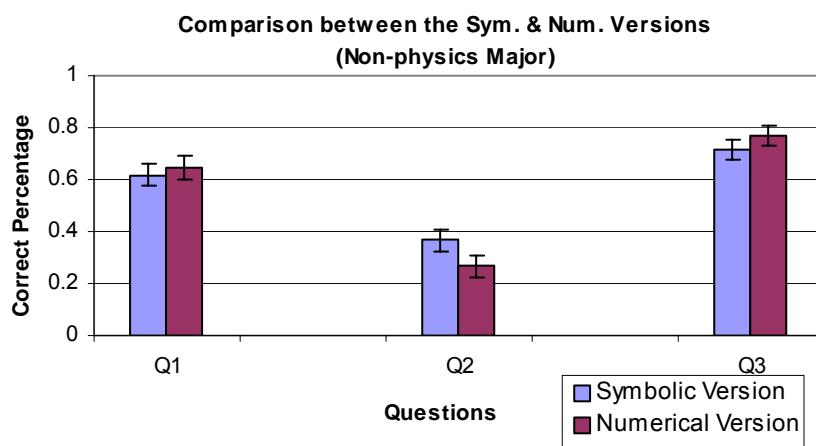
<b>Version</b>	<b>Number of Participants</b>	
	Non-physics major	Physics major
Symbolic version	123	10
Numerical version	113	11
<u>Total</u>	<u>236</u>	<u>21</u>

To compare the students' performances on the two versions of equivalent items, the percentages of correct responses are analyzed separately according to student major. For the 236 non-physics majors, Table 4.11 shows the results and Figure 4.13 displays a bar chart. As can be seen, student performance is rather similar on both versions for all the three

questions. A t-test in the percentage of the correct responses further confirms this result; that is, there is no significant difference in the percentage of correct responses between the symbolic version and the numerical version for all the three questions (Q1:  $df = 233$ ,  $t = 0.45$ ,  $p = 0.66$ ; Q2:  $df = 234$ ,  $t = 1.66$ ,  $p = 0.10$ ; Q3:  $df = 234$ ,  $t = 0.95$ ,  $p = 0.34$ ).

**Table 4.11 The percentage of correct responses among non-physics major students.**

	Percentage of correct answers		
	Q1	Q2	Q3
<b>Symbolic</b> (123 students)	62%	37%	72%
<b>Numerical</b> (113 students)	65%	27%	77%

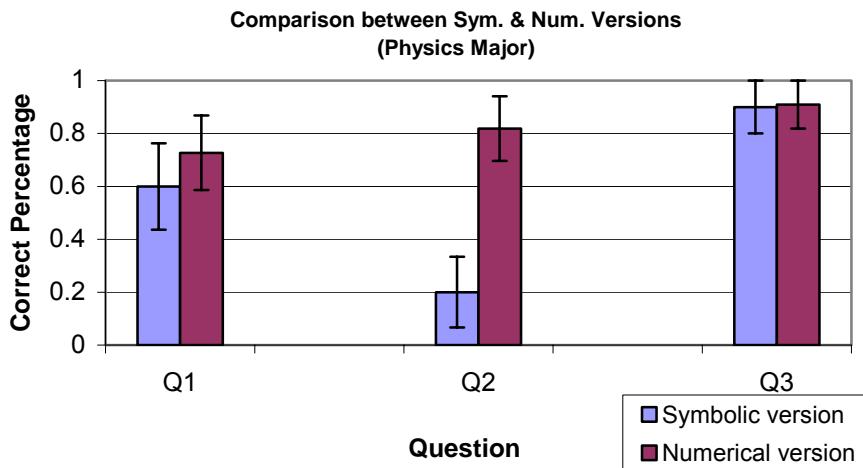


**Figure 4.13 A bar chart of the percentage of correct responses for the two versions among the non-physics major students.**

As for the 21 physics major students, Table 4.12 shows the percentage of correct responses for each question in both versions. In addition, Figure 4.14 displays a graphical bar chart for readers' convenience. Here, for Q1 and Q3, student performance is similar on both versions. But for Q2, student performance is apparently much better on the numerical version than on the symbolic version. A nonparametric Fisher's exact test is conducted for each question further to detect the significant difference in student performance on the two versions. The reason for using the Fisher's exact test is that the number of participating physics majors in each version is small; therefore the parametric t-test or chi-squared test is no longer appropriate. In this experiment, results from the Fisher's test show that for both Q1 and Q3 there is no significant difference in the percentage of correct responses between the two versions (Q1: two-tail  $p = 0.66$ , FET; Q3: two-tail  $p = 1.00$ , FET). However, the difference for Q2 is statistically significant. (Q2: two-tail  $p = 0.0089$ , FET)

**Table 4.12 The percentage of correct responses among physics major students.**

	Percentage of correct answers		
	Q1	Q2	Q3
<b>Symbolic</b> (11 students)	60%	20%	90%
<b>Numerical</b> (12 students)	73%	82%	91%



**Figure 4.14 A bar chart of percentage of correct responses for the two versions among the physics major students.**

It is now interesting to compare Figure 4.13 with Figure 4.14. As readers may have noticed, the fairly high percentage of correct responses to Q1 and Q3 is rather consistent across different versions as well as different majors. Additionally, the aforementioned statistical comparisons further support this observation. To this end, it *seems* that different representations—symbolic or numerical—may not impose a significant difference in student performance on “easy” questions like Q1 and Q3.

As for Q2, it in fact is a rather difficult question, as one can see from Figure 4.13 and Figure 4.14. However, this is certainly not the case for the physics majors who took the numerical version of Q2. (See Figure 4.14.) Indeed, the percentage of correct responses to the numerical version of Q2 is high among the physics majors, as if this question were just as easy as the other two. This interesting result is rather puzzling and it lacks convincing explanations or further supporting evidence. One possible explanation, however, is that the physics major students may have a stronger ability to utilize the given numbers to aid in answering difficult questions. On the other hand, the non-physics majors seem to lack such

ability. Consequently, the difference in student performance between the two versions of Q2 is found to be significant among the physics majors but insignificant among the non-physics majors. Another possible explanation is concerned with the small number of physics majors in this experiment. Because of the small number, the results may not necessarily be meaningful even using the nonparametric statistics as is done in this case.

Although efforts are made to explain the above results, it still needs further studies to fully explore the underlying reasons. Due to the lack of definite conclusions, the current energy assessment adopts the symbolic version of Q2 and the numerical version of Q3 for no particular reasons.

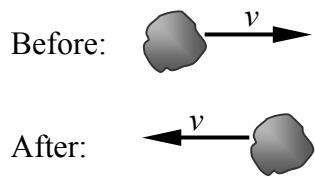
#### 4.4.2 Diagram vs. no diagram

In writing physics test items, physics instructors often consider providing diagrams to accompany the textual statements to facilitate the convey of the question scenario. Despite the instructors' good intentions, students often do not, or perhaps cannot, make good use of the diagrams. Consequently, the diagrams are likely to degrade into a mere decoration. In the worst case, students sometimes misinterpret the diagrams. As a result, the diagrams become a misleading feature rather than just a pure decoration.

A similar situation occurred during the development of the energy assessment. One item as shown in Figure 4.15 was initially created with a diagram to help convey the scenario. It was during the panel review that the diagram gradually entered the attention of the author and some of the panel experts, and became increasingly questionable. To ascertain what made the diagram questionable, one exploratory experiment was conducted as follows.

*In outer space an asteroid was initially moving with a constant speed  $v$  in a certain direction. Some time later, the same asteroid was observed to be moving with the same constant speed  $v$  in the opposite direction. Between the two times, the total external work done on the asteroid was:*

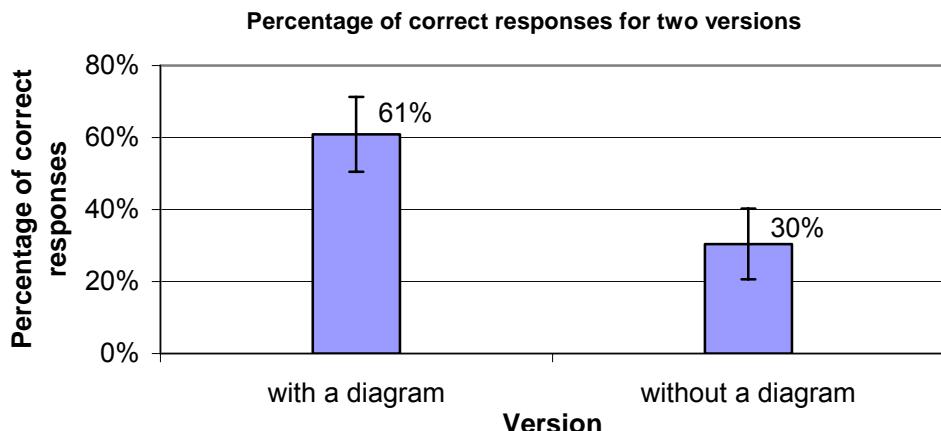
- (a) Positive
- (b) Negative
- (c) Zero



**Figure 4.15 An item was initially created with a diagram.**

Two versions of this question were prepared on paper. One version was provided with a diagram, and the other version was with no diagram but otherwise identical. These two versions were tested with 46 student volunteers as a written quiz in the SCALE-UP (Beichner *et al.* 2000) *M&I* mechanics class at NCSU in the 2006 spring semester. Prior to the quiz, all the student volunteers studied in class the relevant topics regarding energy. In taking the quiz, these 46 students were randomly assigned to one version, and were asked to provide their answers as well as written explanations to the question. Among these 46 students, 23 took the version with a diagram, and the other 23 took the version without a diagram. Student performance on both versions was then analyzed.

Figure 4.16 shows the results. As can be seen, for those students who took the version with a diagram, the percentage of correct responses is 61%, which is over twice as much as the percentage of correct responses (30%) for those who took the version without a diagram. As a follow up check, the chi-squared test is performed, and the result suggests that the difference in student performance on the two versions is statistically significant ( $df = 1$ ,  $\chi^2 = 4.293$ ,  $p = 0.038$ ).



**Figure 4.16 The percentage of correct responses for the versions of the item.**

Further, a close inspection of student written explanations reveals an interesting aspect about the result. Specifically, for those who took the version with a diagram, nearly 1/3 of them provided a correct choice based on the incorrect assumption that the initial position of the asteroid was the same as its final position. As a matter of fact, this information is neither given nor implied in the question statement. Where, then, did such an assumption come from? It is likely that the diagram may be the reason! In the diagram, two arrows of equal length but opposite directions are drawn to indicate the motion directions of the asteroid in the initial and final state, and one arrow is drawn right above the other. If students did not carefully read the problem, it is possible that they may have misinterpreted the diagram as it showing the same initial and final position for the asteroid. Indeed, 1/3 of the correct answers to the version with a diagram were based on such an explanation that: “because there is no change in displacement”, “no displacement = zero work”, “no displacement”.

Interestingly enough, for students who took the version without a diagram, no such explanation was found. More interestingly, for those students who took the version without a

diagram and correctly answered it, they all provided appropriate reasoning except for one student who did not write down any explanations. It seems that those students who took the version without a diagram were immune to this groundless assumption that the displacement of the asteroid was zero. On the other hand, for students who took the version with a diagram, it seems that the diagram may have very likely misled these students to assume that the displacement of the asteroid was zero.

Although it may be too early to draw any definite conclusions from this exploratory experiment, it is clear that students sometimes misinterpret diagrams and the physics instructors' good will is thus twisted. Based on this consideration, the current energy assessment adopts this item without a diagram.

#### **4.5 Summary**

The item and test analysis of the current energy assessment results in five indices (coefficients) as shown in Table 4.5. These five indices (coefficients) indicate that the current version of the energy assessment is a reliable test with adequate discriminatory power. As for the pilot version of the energy assessment, the item and test analysis shows that it lacks sufficient reliability and overall item discriminatory power.

In an attempt to determine the content and cognition levels of the items in the current energy assessment, the revised Bloom's taxonomy (see, for example, Anderson & Krathwohl, 2001) is employed. The revised Bloom's taxonomy consists of two dimensions—content and cognition. In the content dimension, items are categorized in terms of the types of knowledge involved—facts, concepts or principles. In the cognition dimension, items are categorized in terms of the cognitive behaviors required—recall, comprehend, and apply. It is

found that the majority of the items in the current energy assessment require the application of concepts or principles. Only a few items require the comprehension of concepts, where the interpretation of energy graphs is needed.

To determine the number of reasoning steps involved in the items, a set of schemes are designed and applied to the individual items in the current energy assessment. The main idea of the schemes is that the number of reasoning steps is equal to the number of extra quantities needed to bridge the given quantity with the unknown quantity. For most items in the current energy assessment, the number of reasoning steps is no more than one. Only one item in the current energy assessment requires a two-step reasoning process, where both the energy principle and the momentum principle are involved.

In addition, some interesting yet unresolved issues regarding the symbols, numbers and diagrams in the multiple-choice items are uncovered through two exploratory experiments. One experiment aims at whether different representations of physics quantities—symbols or numbers—cause any difference in student performance. The other experiment explores how a diagram affects student response to one item in the energy assessment. Although no definite conclusions are drawn from the exploratory experiments, it is clear that test writers ought to pay close attentions to these issues in writing the test items.

## **CHAPTER 5: Result II—Student Performance on the Energy Assessment**

One major component of this study is to evaluate student understanding of energy topics using the energy assessment. To fulfill this component, the current version of the energy assessment was administered, either on paper or on computer, to a large number of students. Data collected from these students are analyzed quantitatively to examine student performance in large scale. In addition, student verbal explanations are analyzed qualitatively through a small number of student interviews. The purpose of conducting student interviews is to uncover student reasoning for their answers. Based on the quantitative and qualitative analyses, this chapter reports the detailed results of student performance on the current energy assessment and the reasoning behind student responses.

For convenience, this chapter is organized into six sections. The first four sections—section 5.1 to section 5.4—report the quantitative results obtained from four samples of students who completed the current energy assessment in the 2006 fall semester. These four samples of students are listed in Table 5.1 as “sample 2” to “sample 5”. As for students of sample 1, they only completed the pilot version of the energy assessment and thus are not included for discussion. Based on the data collected from students who completed the current version of the energy assessment, section 5.5 describes the factor analysis of student responses as well as the attempted interpretations of the results. Finally, in the section 5.6, the qualitative discussion of student interviews is presented to reveal the reasoning that students employed in answering the questions. Students who participated in the interviews were recruited from “sample 2”, and all completed the written posttest of the current energy assessment prior to the interviews.

**Table 5.1 Five samples of students who completed either the pilot version or the current version of the energy assessment.**

	<b>Version</b>	<b>Participants</b>	<b>No. of Participants</b>	<b>Institution</b>	<b>Time</b>
<b>Sample 1</b>	Pilot	Students from <i>M&amp;I</i> SCALE-UP class	77	NCSU	2005S
<b>Sample 2</b>	Current	Students from <i>M&amp;I</i> lecture hall class	Pre-test: 319 Post-test: 308	NCSU	2006F
<b>Sample 3</b>	Current	<i>M&amp;I</i> (U)TAs	TA: 9* UTA: 11	NCSU	2006F
<b>Sample 4</b>	Current	Students from PY401	Undergrad. : 27 Grad. (Other): 5	NCSU	2006F
<b>Sample 5</b>	Current	First-year physics graduate students	29	Purdue	2006F

\* Four TAs took the energy assessment in a quantum physics class (PY401).

## 5.1 Results for student sample 2—*M&I* students

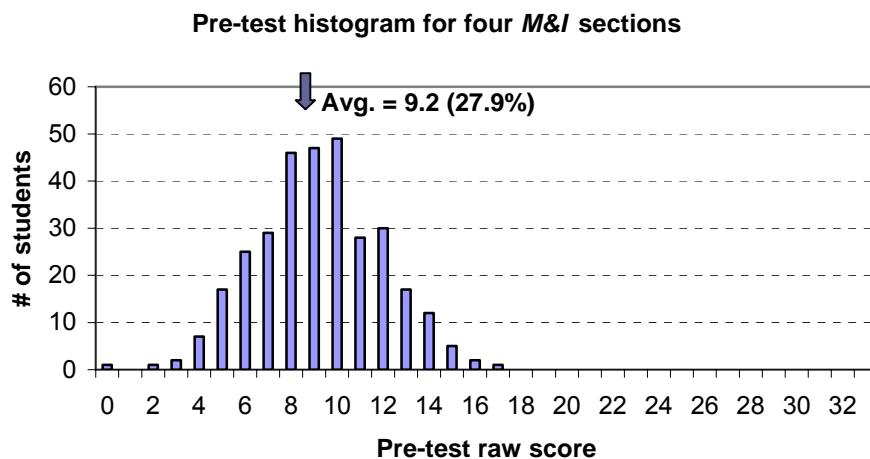
In the 2006 fall semester, the current energy assessment was administered as both the pretest and posttest among students who were then taking the *M&I* mechanics course in lecture hall settings at NCSU. For the pretest, 319 students completed the current energy assessment on WebAssign as a password-protected and time-limited homework assignment in the second week of that semester. For the posttest, 308 students took the current energy assessment on paper as a voluntary in-class test in the second to last week of the same semester. Among them, 262 students completed both the pretest and the posttest.

This section details the results on the pretest (section 5.1.1), posttest (section 5.1.2), and comparisons between the pre and the posttest (section 5.1.3).

### 5.1.1 Results on pretest

#### A. Pretest total score

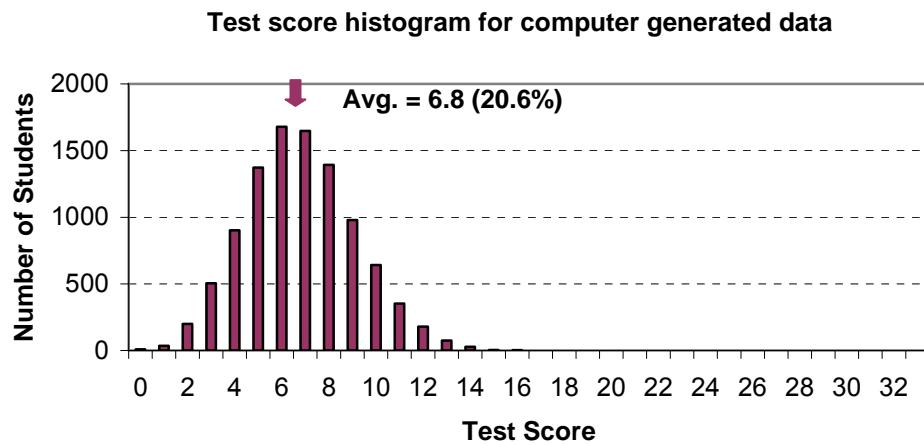
Altogether 319 students from four parallel sections of the *M&I* mechanics class participated in taking the pretest. The overall average total score is 9.2 out of 33 (27.9%) with a standard deviation 2.7 (8.3%). The minimum total score is 0 (0%), and the maximum total score is 17 (51.5%). Figure 5.1 displays the pretest histogram for all four sections. As shown, the score distribution is of a bell shape and fits the Gaussian distribution.



**Figure 5.1** Pretest histogram of the current energy assessment for four *M&I* sections collectively.

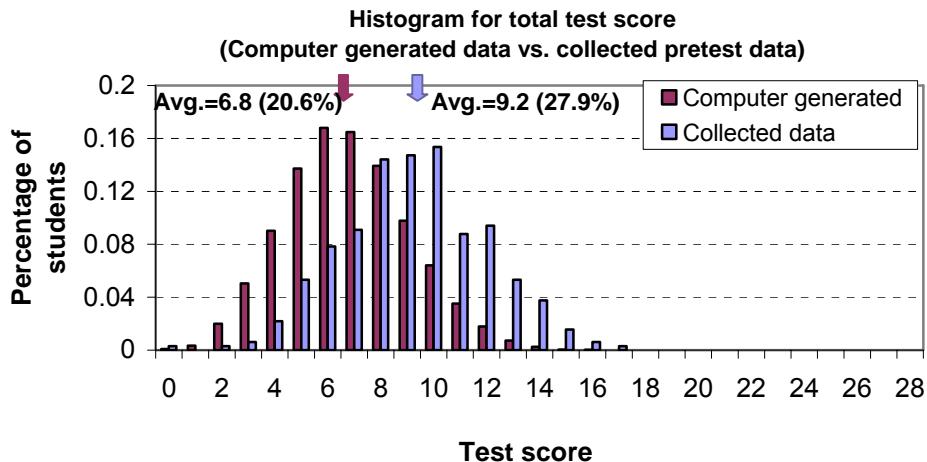
Since the pretest of the current energy assessment was conducted before the students learned in class any relevant topics on energy, it is interesting to see if student responses prior to the course instructions are a mere random guess. To do so, the Monte Carlo method is employed using a Python computer program to simulate student responses that would result solely from random guessing. Figure 5.2 shows the distribution of the total scores produced by pure random guessing. This simulated distribution is based on a computer-generated sample of 10,000 data. As indicated in Figure 5.2, the average total score of the

computer-generated sample is 6.8 out of 33 (20.6%) with a standard deviation 2.3 (7.0%). The minimum total score is 0 (0%), and the maximum total score is 16 (48.5%). To a large extent, the shape of the score distribution in Figure 5.2 resembles that of the score distribution in Figure 5.1. However, their average scores differ by 2.4 (7.3%).



**Figure 5.2 Histogram of the test scores produced by random guessing (computer generated data).**

For readers' convenience, Figure 5.3 displays both histograms in one graph. From this graph, it is clear that there exists a shift between the two distributions. Further, a t-test in the average total score is performed to compare the collected pretest data with the computer-generated data, and the result shows that the difference in the total score between the two sets of data is statistically significant. ( $df=318, t = 15.73, p < .0001$ ) In other words, it is unlikely that student responses in the pretest are based on pure random guessing; rather, students seem to have some prior knowledge about energy that may have helped them correctly answer the questions in the current energy assessment.

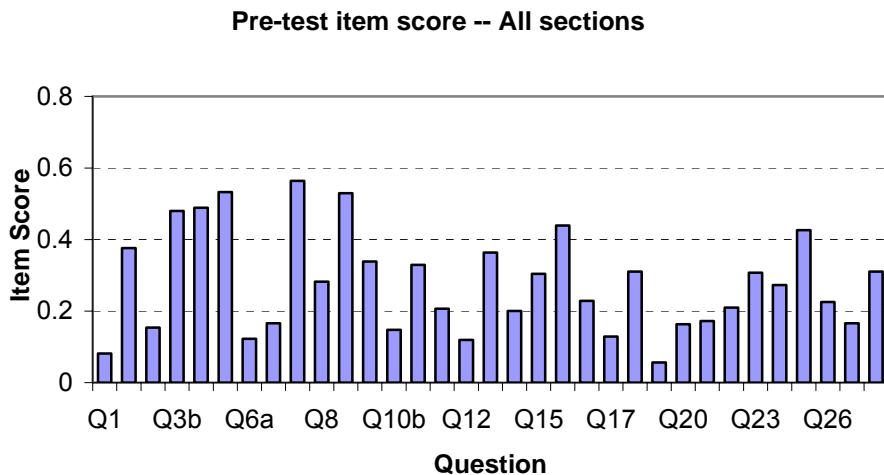


**Figure 5.3 Comparison between two sets of data: computer generated data and collected pretest data.**

### B. Pretest item score

In order to examine student performance on the individual questions in the pretest, the item score, which is the percentage of correct responses for each question, is calculated for the four sections collectively. Figure 5.4 displays the pretest item scores for the individual questions. As one can see, the item scores vary from 0.06 to 0.56 with most scores being above 0.15. In particular, there are several questions whose item scores are noticeably low, such as Q1, Q17, and Q19. It is worthwhile to mention that these questions in fact target some unique topics covered in the *M&I* mechanics course. For example, Q1 addresses the validity of approximation in relation to the relativistic form of kinetic energy; Q17 and Q19 both aim to test the work done on a macroscopic and multi-particle system—a person jumping off a rigid floor in Q17 and a car crashing into a concrete wall in Q19. Several other questions with low item scores (<0.15), such as Q6a, Q10b and Q12, also focus on the important topics emphasized in the course, including gravitational potential energy, graphical

representation of energy and bound/unbound state. As to these questions, it is interesting to see how students performed in the posttest.



**Figure 5.4 Pretest item scores.**

As for the choice distributions of the individual items, Table 5.2 shows the percentage of responses for each alternative choice. From this table, one can find out the major distracters for each question. Take Q1, which has a rather low item score, for an example. This item asks how the final kinetic energy of an electron compares to its initial kinetic energy given that its final speed is twice as fast as its initial speed  $1.4 \times 10^8$  m/s. As shown in Table 5.2, the most attractive distracter in the pretest, which accounts for 47% of the total responses, is choice (b)—“it is twice the initial kinetic energy, because it’s traveling twice as fast.” As opposed to the majority vote for choice (b), only 8% of the students chose the correct answer (d)—“none of the above, because its speed is close to  $3.0 \times 10^8$  m/s, the speed of light.”

**Table 5.2 Pretest alternative choice distributions (The correct answers are underlined.)**

	A (%)	B (%)	C (%)	D (%)	E (%)	F (%)	G (%)	No response (%)
<b>Q1</b>	14	47	29	<u>8</u>				1
<b>Q2</b>	<u>36</u>	12	17	31	3			2
<b>Q3a</b>	54	20	<u>15</u>	4	5			2
<b>Q3b</b>	2	16	<u>48</u>	4	29			1
<b>Q4</b>	31	<u>49</u>	5	5	8	2		1
<b>Q5</b>	3	7	6	20	<u>53</u>	5	5	1
<b>Q6a</b>	32	19	<u>12</u>	36				1
<b>Q6b</b>	<u>15</u>	11	15	25	8	12	12	1
<b>Q7</b>	10	9	<u>56</u>	7	12	4		2
<b>Q8</b>	22	<u>28</u>	45	4				1
<b>Q9</b>	8	34	<u>53</u>	4				1
<b>Q10a</b>	34	<u>34</u>	18	7	7			2
<b>Q10b</b>	23	31	<u>15</u>	23	7			2
<b>Q10c</b>	<u>32</u>	30	13	18	6			1
<b>Q11</b>	<u>19</u>	14	18	44	4			2
<b>Q12</b>	39	24	<u>12</u>	20	4			1
<b>Q13</b>	5	30	13	13	<u>36</u>			2
<b>Q14</b>	4	<u>20</u>	21	39	14			2
<b>Q15</b>	<u>29</u>	44	14	6	5			2
<b>Q16a</b>	15	16	20	<u>44</u>	4			2
<b>Q16b</b>	56	<u>23</u>	15	4				1
<b>Q17</b>	<u>11</u>	13	20	18	16	21		2
<b>Q18</b>	10	11	29	14	<u>31</u>	3		2
<b>Q19</b>	36	27	19	10	<u>6</u>			2
<b>Q20</b>	32	<u>16</u>	18	25	5			4
<b>Q21</b>	35	22	<u>17</u>	11	11			3
<b>Q22</b>	25	25	14	<u>21</u>	13			2
<b>Q23</b>	19	34	<u>31</u>	13				3
<b>Q24</b>	11	21	23	17	<u>27</u>			2
<b>Q25</b>	11	<u>43</u>	31	12				3
<b>Q26</b>	22	<u>23</u>	32	13	8			2
<b>Q27</b>	<u>14</u>	31	34	10	8			3
<b>Q28</b>	30	<u>31</u>	30	6				2

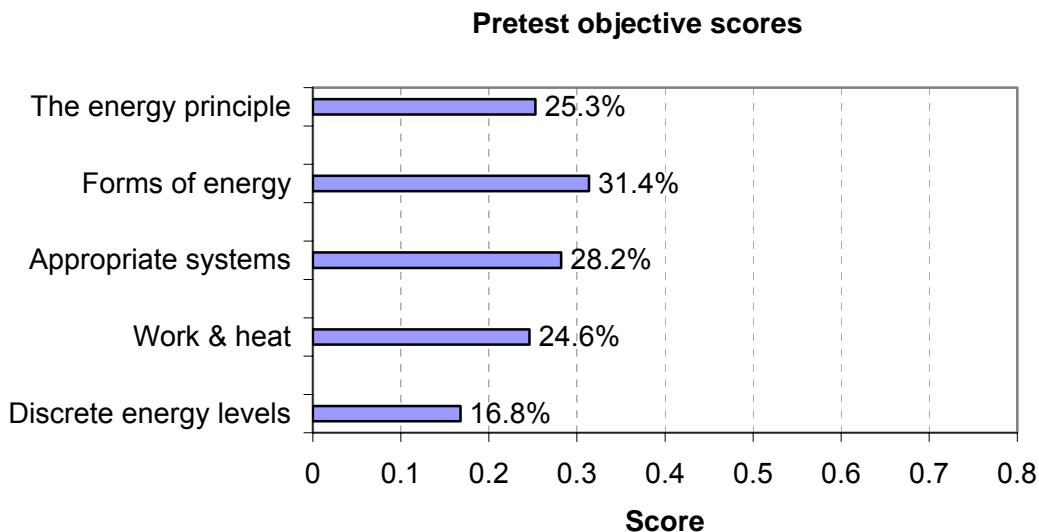
### C. Pretest objective score

Beyond the item scores, student performance on the individual test objectives is another interesting feature to note, namely the test objective scores. As one may recall, there are five major test objectives covered in the current energy assessment, and they are: application of the energy principle, identification and determination of different forms of energy, specification of appropriate systems, determination of and differentiation between  $W$  and  $Q$ , and calculation of absorption/emission spectrum. According to these five major objectives, questions in the current energy assessment can be classified into five groups; within each group questions share the same test objective. In order to detect how students perform on these five objectives individually, it is useful to calculate the average item score of questions that address the same test objectives.

Table 5.3 lists these five major objectives together with the corresponding questions that match these objectives. Also included in Table 5.3 are the individual pretest objective scores. (Also see Figure 5.5.) Not surprisingly, for the pretest, all the objective scores are fairly low, as students did not possess, prior to the course instructions, solid knowledge on the topics covered in the current energy assessment. Particularly, the objective score for “calculation of absorption/emission spectrum” is only 17%, which suggests that students had little prior knowledge on how to deal with the discrete energy levels in calculating atomic spectra. Another two objectives that also demonstrate low scores are related to the topics on work and heat as well as the energy principle. Again, it suggests that, prior to the course instructions, students didn’t know how to deal with work and heat or how to apply the energy principle in answering the questions.

**Table 5.3 Five major test objectives and pretest objective scores.**

Test Objective	Questions	Test Objective Score
Application of the energy principle	Q6b, Q16b, Q22, Q23, Q24, Q25, Q26, Q27, Q28	25.3%
Determination of energy & interpretation of energy graphs	Q1, Q2, Q3a, Q2b, Q4, Q5, Q6a, Q7, Q8, Q9, Q10a, Q10b, Q10c, Q11, Q12	31.4%
Specification of systems	Q13, Q14	28.2%
Determination of & differentiation between work and heat	Q15, Q16a, Q17, Q18, Q19, Q24	24.6%
Calculation of absorption/emission spectra	Q20, Q21	16.8%

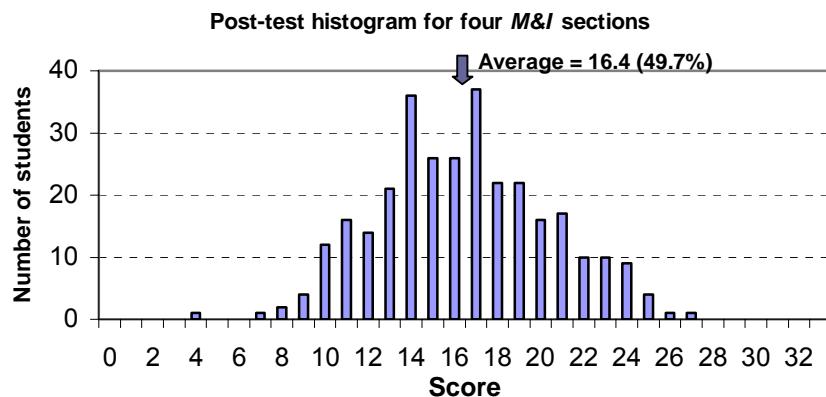


**Figure 5.5 A bar chart of the pretest objective scores. (The objectives are indicated by the corresponding topics names.)**

### 5.1.2 Results on posttest

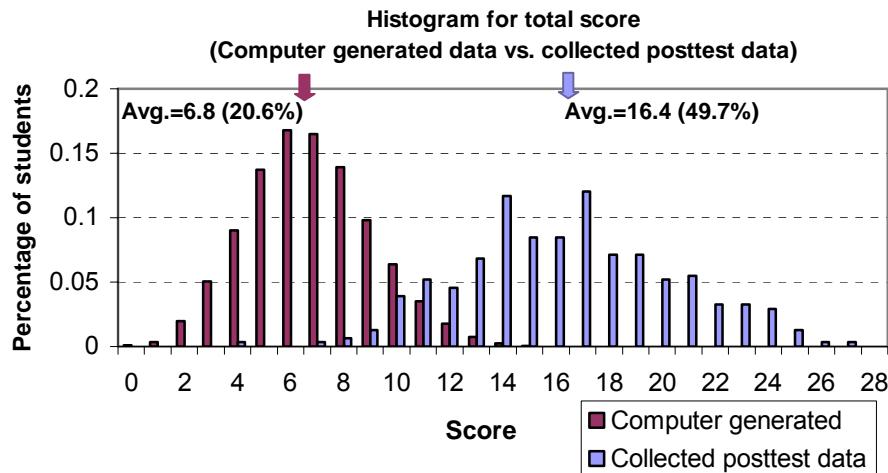
#### A. Posttest total score

In the second to last week of the 2006 fall semester at NCSU, 308 students in the four parallel sections of the *M&I* mechanic course completed the posttest of the current energy assessment. Overall, the average posttest score is 16.4 out of 33 (49.7%) with a standard deviation 4.0 (12.1%). The following graph—Figure 5.6—displays the histogram of the total scores for the four parallel sections collectively. As shown, the posttest total scores vary from 4 (12%) up to 27 (82%), and over forty-percent of the students scored between 14 (42%) and 17 (52%).



**Figure 5.6 Posttest histogram of the current energy assessment for four *M&I* sections collectively.**

Unarguably, the average posttest score is considerably higher than the average that random guessing would produce. As Figure 5.7 shows, the distribution of the computer-generated scores by random guessing differs noticeably from that of the student posttest scores. A t-test in comparison between the real posttest data and the computer-generated data provides further evidence that the student responses in the posttest are better than random guessing. ( $df=307, t=42.18, p < .0001$ )

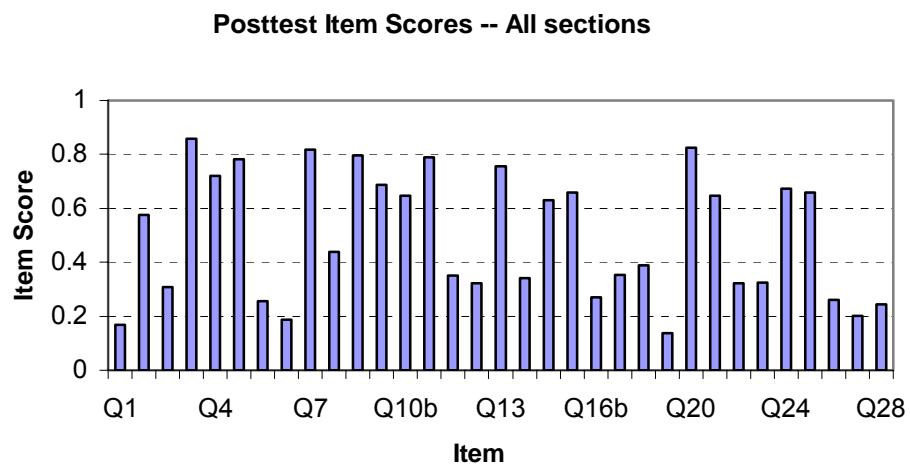


**Figure 5.7 Comparison between two sets of data: computer generated data and collected posttest data.**

### B. Posttest item score

To examine student performance on the individual questions in the posttest, the item scores are calculated, and the results are presented in Figure 5.8. In addition, the alternative choice distributions for the individual questions are tabulated in Table 5.4 as well. As one can see, the posttest item scores vary from 0.14 to 0.86 with most questions being above 0.2. Of all the questions, Q1 and Q19 are the most conspicuous ones due to their lowest item scores. In fact, the pretest item scores of these two questions are also noticeably low (See Figure 5.4). For Q1, however, the choice distribution is rather different in the posttest than in the pretest. Recall that, in the pretest, 47% of the students chose (b) for Q1; that is, 47% of the students answered that the kinetic energy of a fast moving electron would double if its speed is twice as fast. In the posttest, on the other hand, the most popular answer shifted from choice (b) to choice (c), and over two thirds of the students answered that kinetic energy would be quadrupled “because kinetic energy is proportional to speed squared”. Although the correct answer is choice (d)—none of the above, because the final speed is close to the speed

of light—instead of choice (c), it is clear that the majority of the students, after the course instructions, are aware of speed in relation with kinetic energy in a low speed approximation. The fact that these students chose (c) over the correct answer (d) may largely result from their carelessness about the final speed of the electron being near the speed of light. As to the detailed student reasoning that can provide further supporting evidence for that matter, section 5.5 discusses the relevant excerpts from the student interviews.



**Figure 5.8 Posttest item scores.**

Q19 is another question with a rather low item score in the posttest. The question provides a scenario of a car crash test, in which a car hits a concrete wall and stops. Given the magnitude of the average force applied on the car by the wall  $F$  and the length change of the car  $\Delta L$  during the collision, students are asked to figure out the work done on the car as a real system by the wall. The correct answer is (e) zero work, as the contact point between the car and the wall does not move. However, nearly 50% of the students chose (a)  $F \cdot \Delta L$  as the answer in the posttest, and another 23% of the students chose (c)  $-F \cdot \Delta L$ . In fact, regardless of

which distracter students may have chosen, both of these choices—choice (a) and (b)—are concerned with the length change of the car. To this end, it seems that, in answering this question, many students didn't focus on the contact point between the car and the wall; instead, they were overwhelmingly distracted by the configuration change of the car.

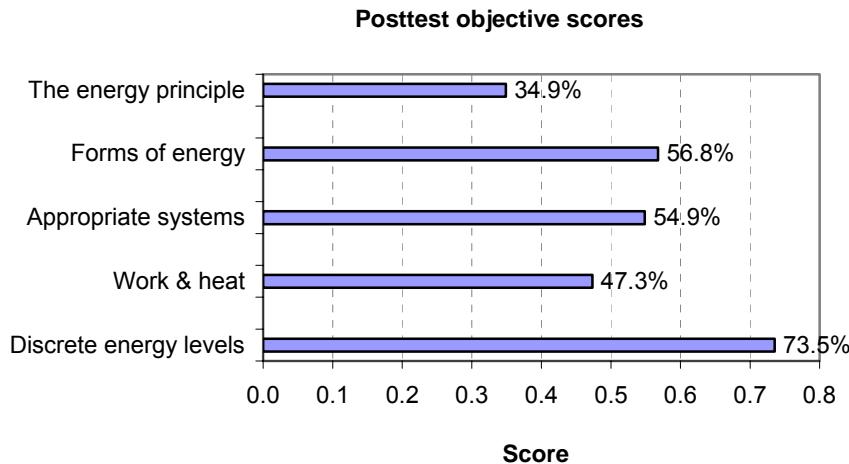
Before closing the discussion on the posttest item scores, it is worthwhile to mention questions Q6a, Q10b, Q12 and Q17, as section 5.1.1b calls for attention to their low pretest item scores ( $< 0.15$ ). Of note here is that, in the posttest, all of these questions have an item score higher than 0.25; thus suggesting that there is an improvement in student performance on the topics covered in these questions. In specific, these topics are: the sign of gravitational potential energy in Q6a, the graphical representation of  $K+U$  of a two-asteroid system in Q10b, the sign of  $K+U$  of a bound system in Q12 and the work done on a jumping person by a rigid floor in Q17. As for whether or not the improvement is significant, section 5.1.3 reports the results.

**Table 5.4 Posttest alternative choice distributions (The correct answers are underlined.)**

	A (%)	B (%)	C (%)	D (%)	E (%)	F (%)	G (%)	No response (%)
<b>Q1</b>	2	13	67	<u>17</u>	0	0	0	1
<b>Q2</b>	<u>57</u>	3	9	27	4	0	0	1
<b>Q3a</b>	62	7	<u>31</u>	0	0	0	0	0
<b>Q3b</b>	1	12	<u>86</u>	0	2	0	0	0
<b>Q4</b>	21	<u>72</u>	1	2	3	0	0	0
<b>Q5</b>	2	3	4	8	<u>78</u>	2	2	0
<b>Q6a</b>	48	6	<u>26</u>	20	0	0	0	1
<b>Q6b</b>	<u>19</u>	3	7	39	10	19	3	0
<b>Q7</b>	3	3	<u>82</u>	1	8	3	0	0
<b>Q8</b>	13	<u>44</u>	43	1	0	0	0	0
<b>Q9</b>	2	13	<u>80</u>	6	0	0	0	0
<b>Q10a</b>	10	<u>69</u>	17	2	2	0	0	0
<b>Q10b</b>	18	5	<u>65</u>	7	5	0	0	0
<b>Q10c</b>	<u>79</u>	13	3	2	2	0	0	0
<b>Q11</b>	<u>35</u>	9	10	44	2	0	0	0
<b>Q12</b>	19	21	<u>32</u>	14	14	0	0	0
<b>Q13</b>	4	16	2	2	<u>76</u>	0	0	0
<b>Q14</b>	2	<u>34</u>	5	44	15	0	0	0
<b>Q15</b>	<u>63</u>	21	1	1	13	0	0	0
<b>Q16a</b>	17	8	9	<u>66</u>	1	0	0	0
<b>Q16b</b>	61	<u>27</u>	6	5	0	0	0	0
<b>Q17</b>	<u>35</u>	12	8	13	11	21	0	0
<b>Q18</b>	3	1	50	3	<u>39</u>	3	0	0
<b>Q19</b>	49	8	23	6	<u>14</u>	0	0	0
<b>Q20</b>	4	<u>82</u>	4	3	6	0	0	0
<b>Q21</b>	8	18	<u>65</u>	6	3	0	0	0
<b>Q22</b>	24	23	7	<u>32</u>	13	0	0	0
<b>Q23</b>	17	28	<u>32</u>	21	1	0	0	0
<b>Q24</b>	4	13	10	6	<u>67</u>	0	0	1
<b>Q25</b>	6	<u>66</u>	16	11	0	0	0	1
<b>Q26</b>	37	<u>26</u>	26	6	5	0	0	1
<b>Q27</b>	<u>20</u>	21	46	5	6	0	0	2
<b>Q28</b>	17	<u>24</u>	45	11	0	0	0	2

### C. Posttest objective score

Since the objective scores can provide insights into student performance on the individual test objectives covered in the energy assessment, the posttest objective scores are calculated, as are done for the pretest, and the results are presented in a bar chart in Figure 5.9.



**Figure 5.9 A bar chart of posttest objective scores. (The objectives are indicated by the corresponding topics names.)**

As one can see, all of the five objective scores are near or above 0.35. Particularly, the objective “calculation of emission/absorption spectrum” has the highest score and is over 0.7. It seems that the majority of the students, after the course instructions, are able to deal with discrete energy levels in calculating the emission and absorption spectrum. On the other hand, the objective “application of the energy principle” has the lowest score 0.35. In fact, this relatively low objective score, as opposed to the other scores in the posttest, is not unreasonable or alarming, for application of the energy principle inevitably involves other test objectives and therefore is the most comprehensive and challenging one. This is, indeed, why “application of the energy principle” takes the top position in the hierarchical structure of the test objectives as shown in Figure 3.1.

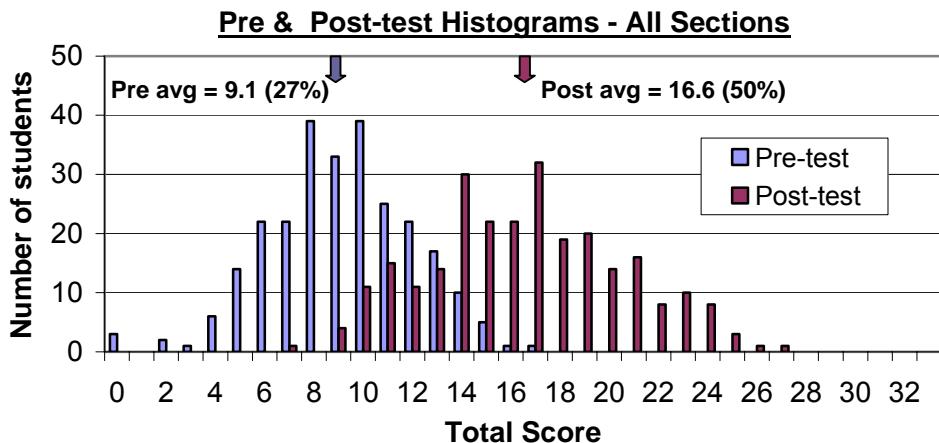
### 5.1.3 Comparisons between the pretest and the posttest

#### A. Comparisons in total score

To detect the effect of the course instructions on student performance, it is useful to compare the posttest scores with the pretest scores. One noteworthy aspect of such comparisons is that only paired sample data are meaningful, as the simple direct comparison in overall scores is flawed. The main reason lies in the very fact that the two sets of data that were collected before and after the instructions are not independent. Specifically, the majority of the students who participated in the pretest also took the posttest. Therefore, it is appropriate only to retain the data collected from those students who completed both the pretest and posttest for pair-wise comparisons.

After rather tedious and time-consuming processes of “find-and-match”, the paired data in the four sections of the *M&I* mechanics course were constructed. In total, 262 students completed both the pretest and the posttest. Among them, 74 students were from Sec001, 66 students from Sec002, 65 students from Sec003, and another 57 students from Sec004.

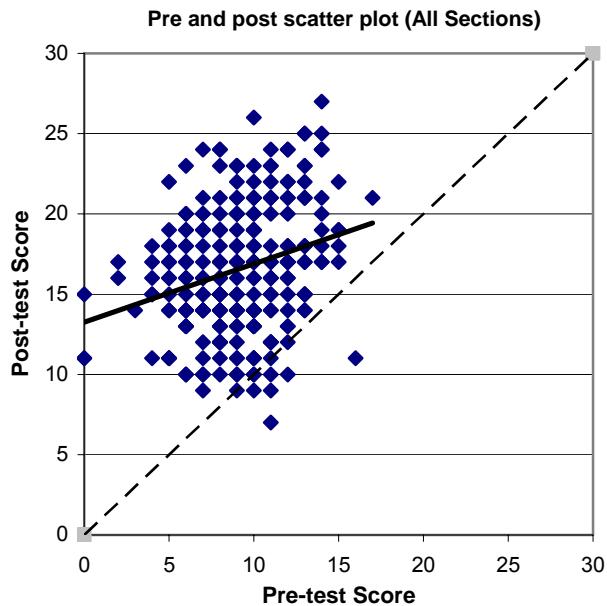
The overall pre- and post-test histograms, based on the 262 paired data from all four *M&I* mechanics sections, are plotted in Figure 5.10. For these 262 students, the average pretest score is 9.1 (27%), and the average posttest score is 16.6 (50%). As one may notice, the increase in the test score from 27% to 50% has marked the posttest average almost twice of the pretest average.



**Figure 5.10 A pair-wise comparison between pre- and post-test scores.**

In comparing the pair-wise pre and post total scores, a scatter plot, as shown in Figure 5.11, is rather illuminating. In Figure 5.11, each dot represents a student, and the dashed line corresponds to the  $Y = X$  function. Dots that are above the dashed line indicate a positive gain in the test score ( $\text{gain} = \text{post score} - \text{pre score}$ ); dots on the dashed line indicate a zero gain, and dots below the dashed line indicate a negative gain. From this plot<sup>10</sup>, it is clear that most dots are above the dashed line; thus indicating a positive gain in total score for the majority of the students. Further, a paired t-test in the total scores is conducted to confirm this finding. The result shows that the pair-wise difference between the pre- and post-test scores is statistically significant ( $df = 261$ ,  $t = 28.41$ ,  $p < .0001$ ). Simply put, there is a significant increase in student overall performance on the energy assessment after the course instruction.

<sup>10</sup> The solid line in Figure 5.11 is the trendline of post-test vs. pre-test scores; its large intercept and relatively small slope suggest that student pre-instructional knowledge may not impose a major effect on student post-test scores. Students at all levels seem to benefit from the course instruction in a fairly similar manner, although students who scored higher in the pre-test tend to have slightly higher post-test scores.



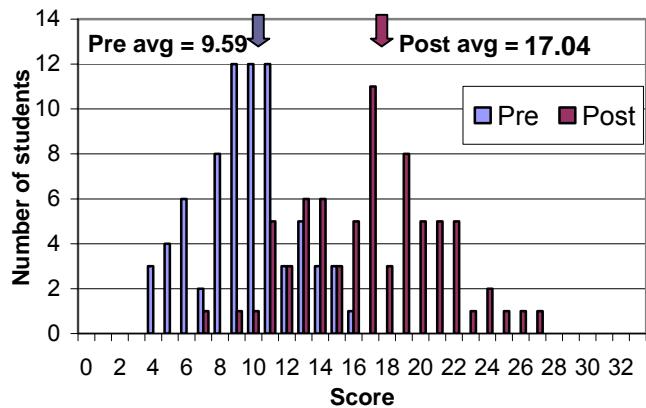
**Figure 5.11** A scatter plot for the pair-wise pre- and post-test scores.

The above analyses denote that the gain in total score is significant for the four sections collectively. It is then of great interest to investigate whether the gains for the individual sections are significant. Moreover, it is meaningful to detect the difference, if any, in the gains across the four sections taught by three different instructors. The question, in other words, is whether the difference characterized mainly by the different instructors has any effect on the gains in total score. To answer these questions, it is rather useful to examine the gains of these four sections individually.

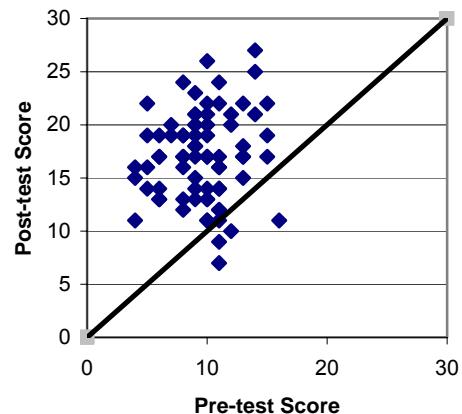
First, let's look at the plots in Figure 5.12. These plots display the histograms of the pre and post total scores, as well as the scatter plots, for the individual sections. As one can see, for each section there is a positive gain in the average score, varying from 6.6 (20%) to 7.8 (24%).

Figure 5.12 Pre- and post-test histograms and scatter plots for four *M&I* sections individually.

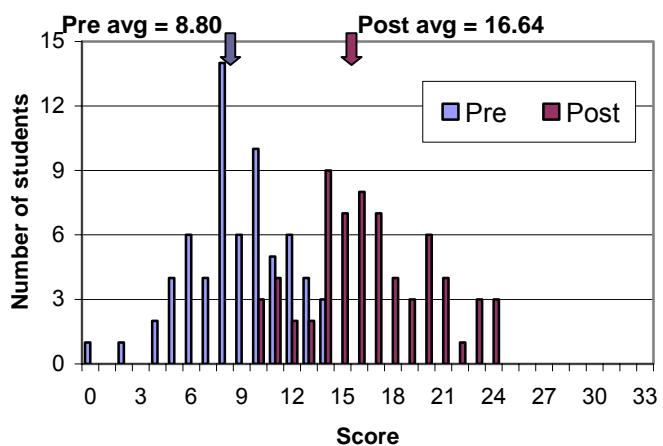
Pre & Post-test Histograms (Sec001)



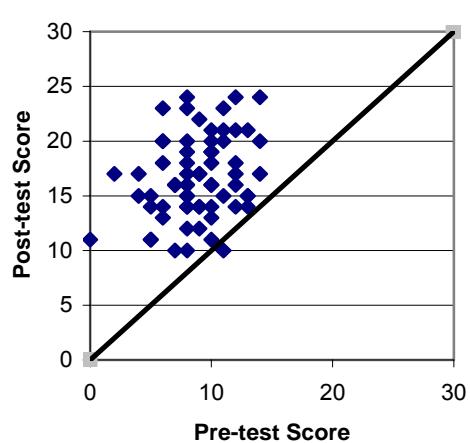
Pre and post scatter plot (Sec001)



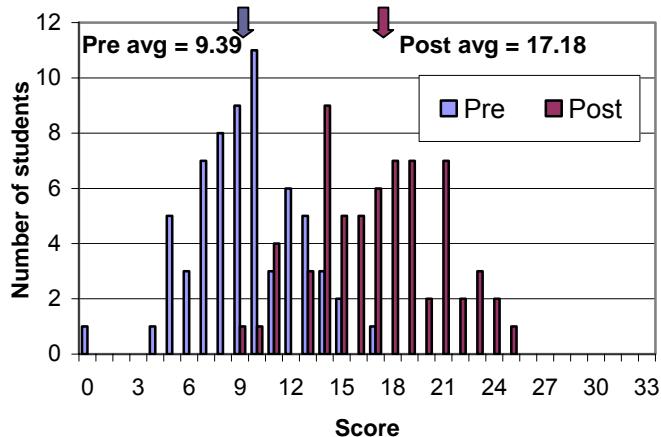
Pre & Post-test Histograms (Sec002)



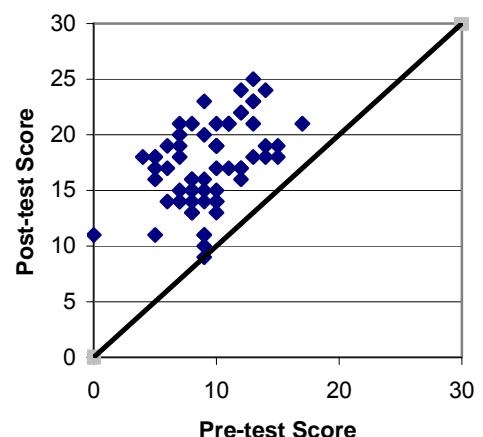
Pre and post scatter plot (Sec002)



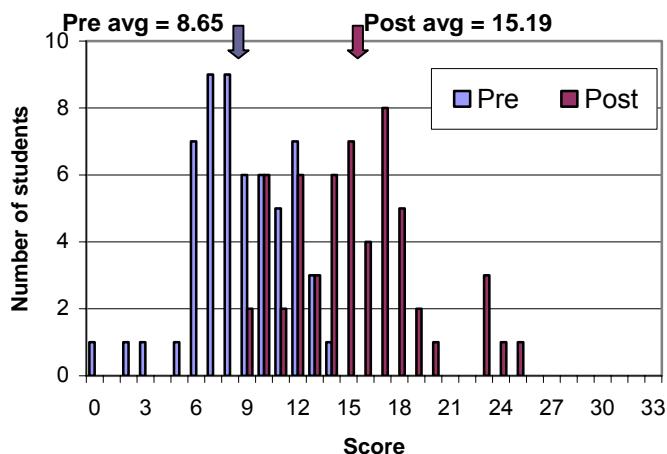
**Pre & Post-test Histograms (Sec003)**



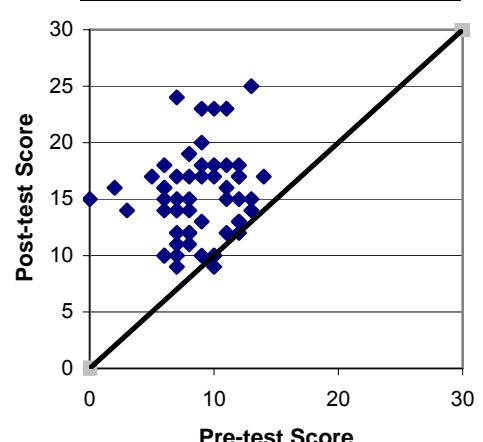
**Pre and post scatter plot (Sec003)**



**Pre & Post-test Histograms (Sec004)**



**Pre and post scatter plot (Sec004)**



A paired t-test is conducted for each section to examine if each individual gain is significant. The results show that the individual gains for these four sections are all statistically significant; therefore indicating a significant improvement in student performance on the energy assessment after instructions for each section (Sec001:  $t$  (73) = 13.46,  $p < .0001$ ; Sec002:  $t$  (65) = 16.90,  $p < .0001$ ; Sec003:  $t$  (64) = 17.40,  $p < .0001$ ; Sec004:  $t$  (56) = 11.21,  $p < .0001$ ).

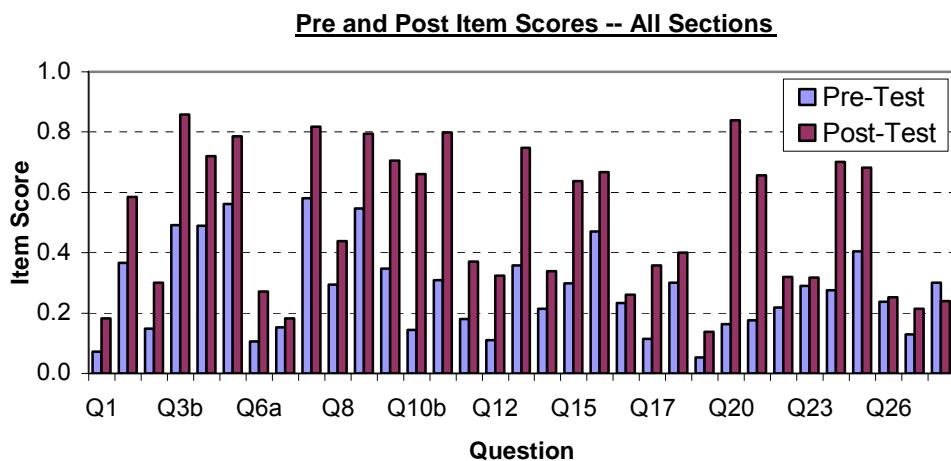
Moreover, a one-way ANOVA is employed to test if the gains are statistically different across these four parallel sections. The purpose of this analysis is to detect if the factor “section” plays a significant role in the total score gains. Results from the ANOVA show that there is no significant difference in the gains among these four sections ( $F$  (3, 258) = 1.20,  $p = .3107$ ). In other words, the gains in total scores are statistically uniform across the four sections and thus are independent of the “section” factor.

## B. Comparisons in item scores

Up to now, the comparisons have been focused on the (gains in) total score. In order to inspect how students performed on any particular items before and after the course instructions, it is necessary to have a closer look at the individual items, more specifically, the gains in the item scores. In so doing, it becomes clear whether or not the course instructions have helped students perform better on the individual topics covered in the energy assessment.

Based on the same argument that valid pre and post comparisons ought to be pairwise, only data collected from those who took both the pre- and post-test are retained for analysis. For the 262 students who took both the pre- and post-test, Figure 5.13 displays the

item scores for the individual questions. As can be seen, there is an increase in item scores for each question except for Q28. In fact, Q28 is not an easy question; students are asked to compare the work done on an object by a constant force during two equal time intervals, given a speed increase from 0 to  $v$  in the first time interval and from  $v$  to  $2v$  in the second time interval. In the posttest, 45% of the students answered that that equal work was done on the object during the two time intervals. It is likely that these students who chose this answer may have focused on the fact that the change in speed is the same for the two time intervals and mistakenly invoked the momentum principle in answering this question.



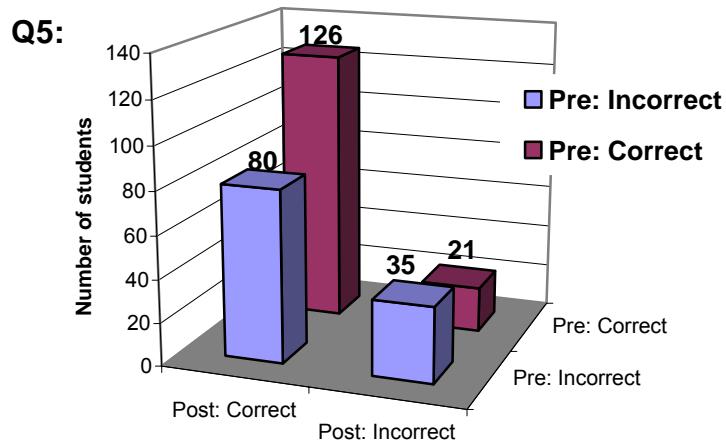
**Figure 5.13 A pair-wise comparison between the pre- and post-test item scores.**

In order to compare the pair-wise item scores between the pre- and post-test, it is useful to construct a  $2\times 2$  contingency table (matrix) for each question. Take Q5 for an example. As Table 5.4 shows, 126 students of the total 262 students provided a correct answer in both the pre- and post-test. Another 80 students who answered incorrectly in the pretest provided a correct answer in the posttest. Yet another 21 students who answered correctly in the pretest provided a wrong answer in the posttest. The last 35 students

answered consistently wrong in both the pre- and post-test. In Figure 5.14, a 3-D bar chart is also constructed to illustrate this case.

**Table 5.5 A contingency table for Q5.**

Q5		<u>Post-test</u>	
		Correct (score 1)	Incorrect (score 0)
Pre-Test	Correct (score 1)	126	21
	Incorrect (score 0)	80	35



**Figure 5.14 A 3-D bar chart for Q5.**

Based on the  $2 \times 2$  contingency tables, the McNemar's test<sup>11</sup> is conducted for each individual question to detect if the gain in the item score is significant. In other words, the McNemar's test is to examine whether or not students performed significantly better on each question in the posttest than in the pretest. Results show that there is a significant increase in the posttest item scores for all the individual questions except for Q6b ( $S = 0.346, p = 0.889$ ),

<sup>11</sup> McNemar's test is used for comparing proportions of two dependent samples. (See, for example, Agresti & Finlay, 1997)

Q16b ( $S = 0.505, p = 0.477$ ), Q23 ( $S = 0.527, p = 0.468$ ), Q26 ( $S = 0.346, p = 0.683$ ) and Q28 ( $S = 0.251, p = 0.112$ ). This suggests that the course instructions have enhanced student performance on most of the topics covered in the energy assessment.

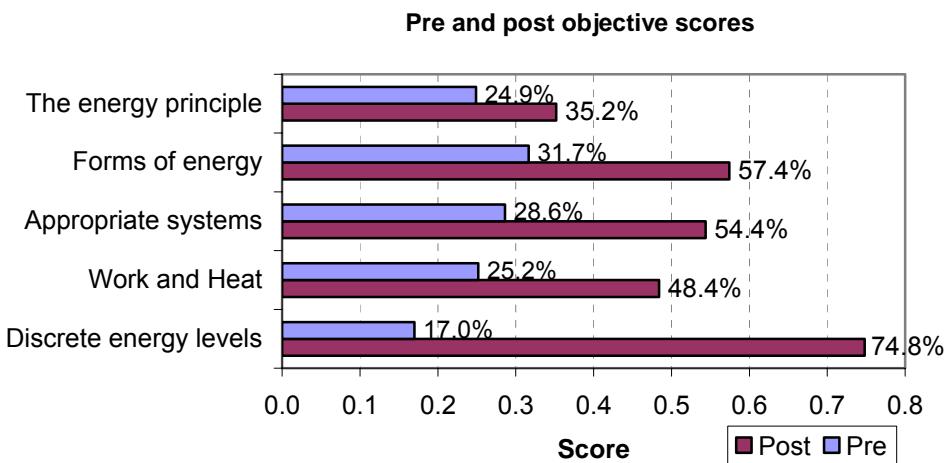
It is interesting to find that all of these questions, for which there is no significant gain in the item scores, require an application of the energy principle. One noteworthy aspect, however, is that, for the other four questions that also require an application of the energy principle, the gains are significant. These four questions are: Q22 ( $S = 6.57, p = 0.01$ ), Q24 ( $S = 81.45, p < .0001$ ), Q25 ( $S = 41.96, p < .0001$ ) and Q27 ( $S = 8.067, p = 0.005$ ).

As for the reasoning behind student answers to some of the aforementioned questions, section 5.5 provides detailed discussions on the relevant excerpts from student interviews.

### C. Comparisons in objective scores

Since the gains in item scores are insignificant for some of the questions that involve the energy principle, it is then of great interest to see whether or not the gain is significant for all the questions collectively that require an application of the energy principle; in other words, to examine the gain in the objective score.

In order to do so, it is again necessary to use only the pair-wise data collected from those students who completed both the pre- and post-test. For the 262 pair-wise data, Figure 5.15 displays the individual pre- and post-test objective scores. As one can see, there is a positive gain for all the five objectives, including “application of the energy principle”.



**Figure 5.15 A pair-wise comparison between pre- and post-test objective scores.**  
(The objectives are indicated by the corresponding topics names.)

As for whether or not the gains are significant, a paired t-test is conducted to answer this question. Results from the t-test show that all the gains are significant (See Table 5.6); thus suggesting that student better performance on the individual objectives in the posttest is statistically meaningful and is not by chance. Based on this result, it is reasonable to infer that the course instructions have promoted student performance on all the objectives covered in the energy assessment, including the objective “application of the energy principle”, although some individual questions have insignificant gains in the item scores.

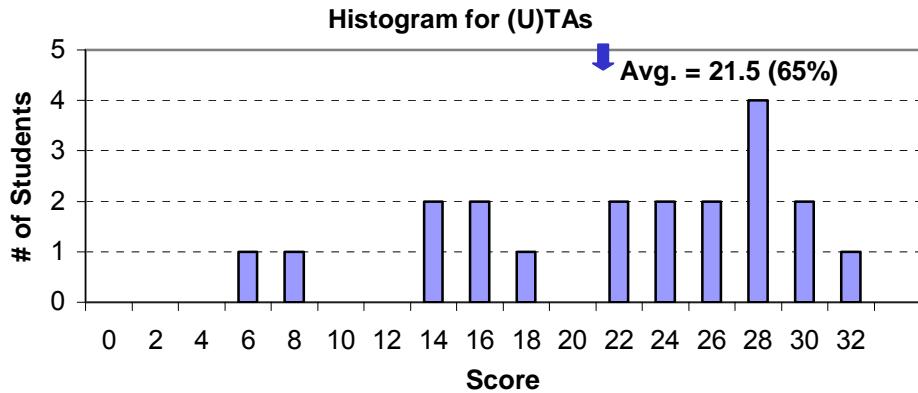
**Table 5.6 Results of the paired t test in objective scores.**

Test Objectives	<u>Paired t test</u>		
	df	t	p value
The energy principle	261	8.07	<.0001
Forms of energy	261	24.24	<.0001
Appropriate systems	261	9.22	<.0001
Work and heat	261	14.92	<.0001
Discrete energy levels	261	20.84	<.0001

## **5.2 Results for student sample 3—*M&I* (U)TAs**

In the 2006 fall semester, the current energy assessment was also administered to the graduate and undergraduate *M&I* teaching assistants (TAs and UTAs) at NCSU. Presumably, these students have a better mastery of physics knowledge than the typical undergraduates in science and engineering majors; thus their performance on the energy assessment provides useful information on the “ceiling effect”—the highest imperfect score that the test audience can reasonably be expected to achieve. (The audience of the current energy assessment includes students who have been exposed to the *M&I* materials, or students who have learned from other advanced physics classes the energy topics comparable to those covered in the *M&I* mechanics course.)

Altogether nine TAs and eleven UTAs completed the current energy assessment. For these twenty (U)TAs, the overall average is 21.5 (65%) with a standard deviation 7.5 (23%). As shown in Figure 5.16, the maximum score is 32 (97%), and the minimum is 6 (18%). Specifically for the nine TAs, the total scores vary from 6 (18%) up to 32 (97%) with an average 23.0 (70%). As for the eleven UTAs, the total scores range from 7 (21%) to 30 (91%), and the average is 20.3 (61%). According to these results, it seems possible for well-prepared students to obtain a total score as high as over 90% on the energy assessment.

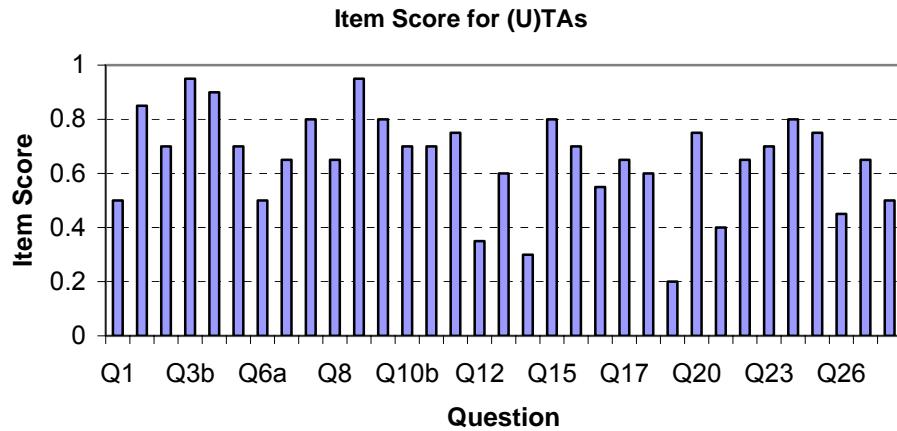


**Figure 5.16 Score histogram of the current energy assessment for the *M&I* (U)TAs.**

Interestingly enough, these TAs and UTAs, albeit from different academic levels, appear to have rather similar scores on the energy assessment. In order to provide further evidence, a nonparametric Wilcoxon-Mann-Whitney (WMW) test is conducted due to the small numbers of the participating (U)TAs. The result shows that there is no significant difference ( $S = 141$ ,  $p = 0.2510$ ) in the total scores between the TAs and UTAs. In this respect, these TAs and UTAs have a comparable mastery of the *M&I* course materials regarding the energy topics.

In addition to the total score, the item scores are calculated as well for all the twenty (U)TAs, and the results are displayed in Figure 5.17. As one can see, the most distinctive feature of Figure 5.17 is the lowest percentage of correct responses for Q19. This question asks for the work done on a car as a real system by a concrete wall as the car crashes into the wall. A close inspection further reveals that the most popular choices for this question are choice (a)  $F\Delta L$  and choice (c)  $-F\Delta L$ , which account for 40% and 30% of the total responses respectively. This result closely resembles that reported in the previous section, where 50% of the *M&I* students—those who were taking the *M&I* mechanics course in the 2006 fall

semester at NCSU—chose (a)  $F\cdot\Delta L$  as the answer, and another 23% chose (c)  $-F\cdot\Delta L$ . It seems that this item poses a considerably difficult question for the *M&I* teaching assistants as much as it does for the *M&I* students.

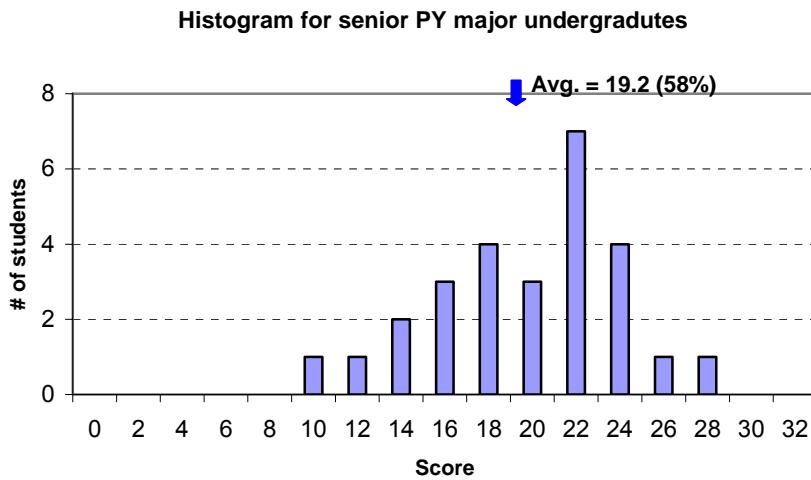


**Figure 5.17 Item scores for the *M&I* (U)TAs.**

### 5.3 Results for student sample 4—senior physics major undergraduates

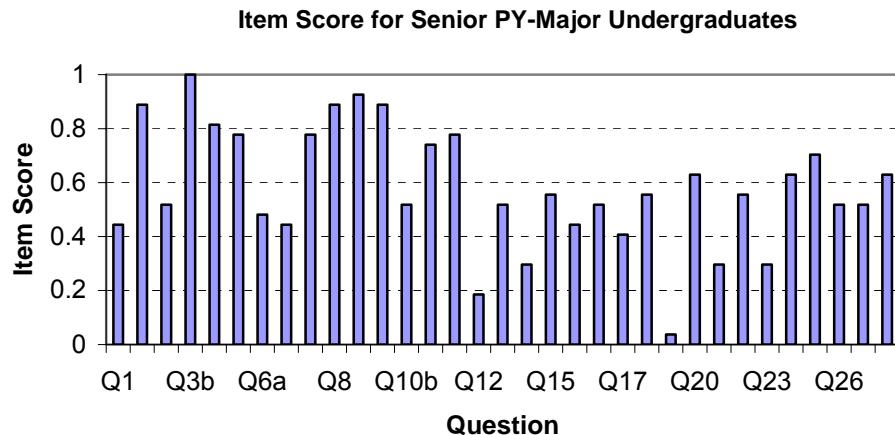
In the 2006 fall semester, twenty-seven senior physics-major undergraduates and five first-year graduate students completed the current energy assessment as a written test in a quantum physics class at NCSU.

For the twenty-seven senior physics major undergraduates, the average score is 19.2 (58%), ranging from 10 (30%) to 27 (82%) with a standard deviation 4.2 (13%). Figure 5.18 provides the overall score distribution for these twenty-seven undergraduates. As it shows, the score distribution is of a slightly skewed Gaussian shape.



**Figure 5.18 Score histogram of the current energy assessment for the senior physics major undergraduates.**

In addition to the total score, the item scores of the individual questions are calculated as well, and the results are presented in Figure 5.19. As one can see, Q19 once again shows the lowest percentage of correct responses. (This question asks for the work done on a car as a real system by a wall as the car crashes into the wall.) Similar to what has been reported for the *M&I* students and (U)TAs in the previous sections, the most popular answers to Q19, among the senior physics major undergraduates, are choice (a)  $F \cdot \Delta L$  and choice (c)  $-F \cdot \Delta L$ , which account for 42% and 33% of the total responses respectively. Apparently, like the *M&I* students and (U)TAs, these students were largely distracted by the configuration change of the car, and failed to focus on the contact point between the car and the wall in answering the question.



**Figure 5.19** Item scores for the senior physics major undergraduates.

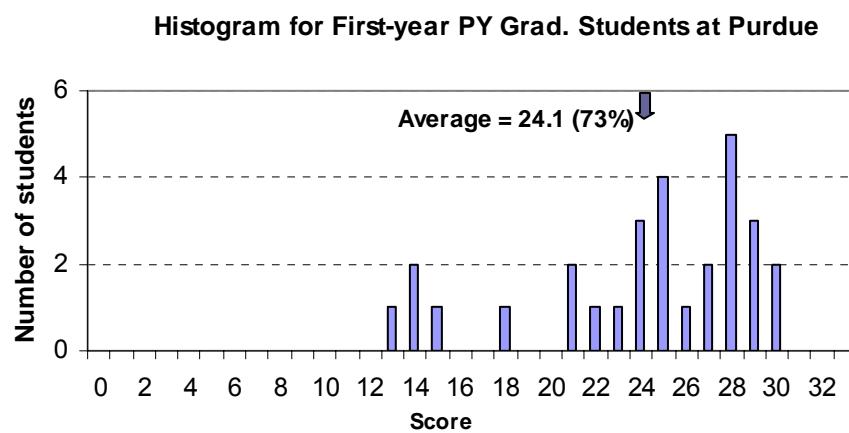
As for the five graduate students who also completed the energy assessment, the average score is 14.4 (44%) with a standard deviation 3.8 (11%). Among them, the maximum score is 19 (58%), and the minimum is 11 (33%). Since the number of these graduate students is too small to yield any other meaningful information, no further analysis is attempted.

#### 5.4 Results for student sample 5—first year physics major graduate students at Purdue

In the 8<sup>th</sup> week of the 2006 fall semester, Mark Haugan—a professor at Purdue—administered the current energy assessment among twenty-nine first-year physics major graduate students at Purdue University. Since these students had been reportedly reading the *M&I* mechanics textbook before taking the energy assessment, their performance provides useful baseline data on students at a higher academic level.

For these twenty-nine graduate students, the average total score is 24.1 (73%) with a standard deviation 5.0 (15%). As one can see from the histogram in Figure 5.20, the scores vary from 13 (39%) to 30 (91%) with the majority being above 20 (61%). In a comparison to

the average score of the nine *M&I* TAs at NCSU (raw score: 23, percentage: 70%), the average for the Purdue first-year graduate students is not statistically different (WMW:  $S = 167$ ,  $p = 0.780$ ). To this end, the TAs at NCSU and the first-year physics major graduates at Purdue have a rather similar mastery of physics knowledge regarding the energy topics covered in the energy assessment.

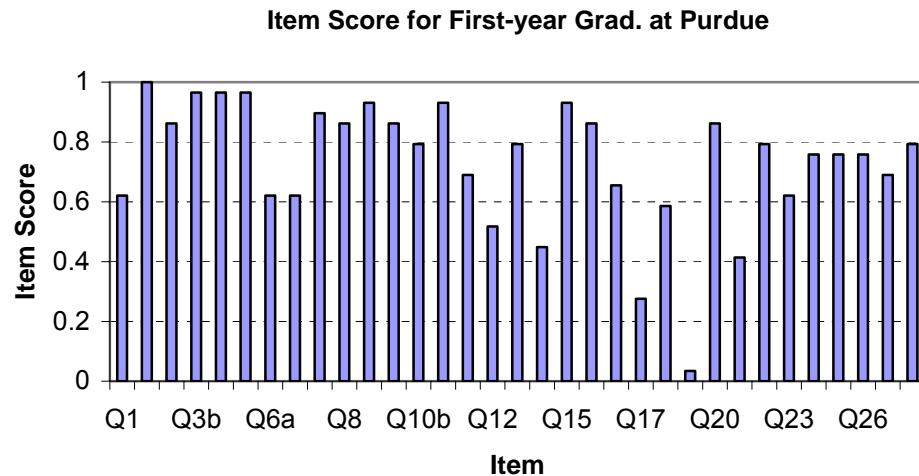


**Figure 5.20 Score histogram of the current energy assessment for the first-year physics major graduate students at Purdue.**

For the item scores, Figure 5.21 shows the percentages of correct responses for the individual questions. The striking feature of Figure 5.17 is, once again, the distinctively low item score of Q19. (This question asks for the work done on a real system—a car—by a wall as the car crashes into the wall.) Further, the most popular choices are, again, choice (a)  $F \cdot \Delta L$  and choice (c)  $-F \cdot \Delta L$ , which together account for over two thirds of the total responses.

So far, the percentage of correct responses for Q19 has been consistently low among different students, including the *M&I* students and (U)TAs at NCSU, the senior physics major undergraduates at NCSU, and the first-year physics major graduates at Purdue. More interestingly, the majority of these students are in favor of either choice (a)  $F \cdot \Delta L$  or choice

(c)  $-F \cdot \Delta L$ , which suggests that these students have been sidetracked by the shape change of the car in answering this question.



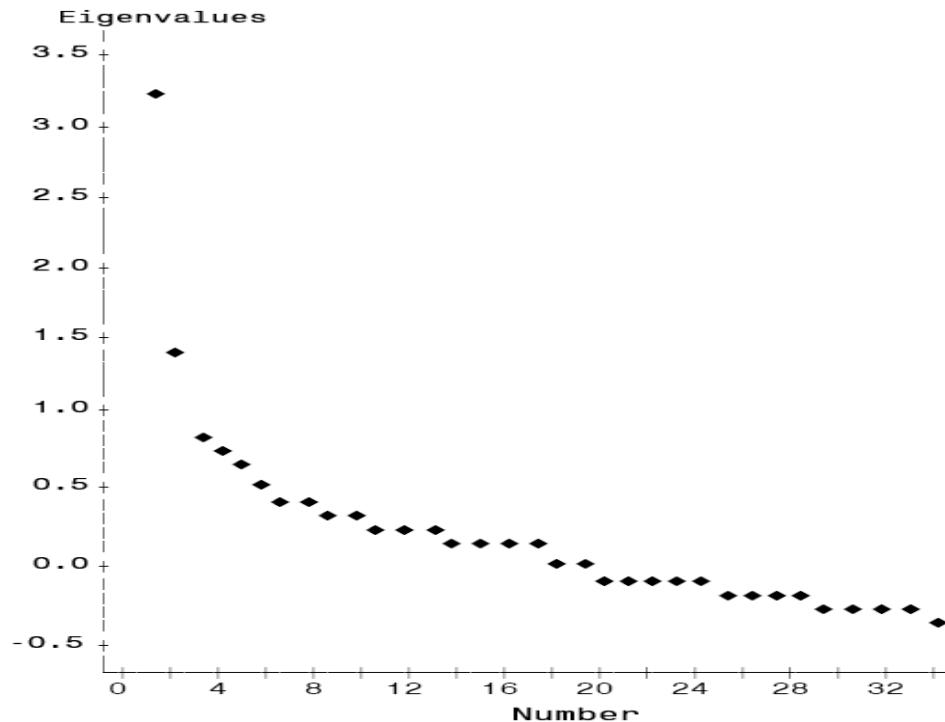
**Figure 5.21** Item scores for the first year graduate students at Purdue.

## 5.5 Factor analysis of student responses to the energy assessment

At an advanced level and receiving much attention in standardized assessment tools is the factor analysis. Factor analysis examines the correlations among individual items, and groups together the items that essentially measure the same underlying variables (factors). Based on the data collected from these aforementioned four student samples, a factor analysis is performed to investigate how the items in the current energy assessment are clustered together.

Figure 5.22 is a “scree plot”, and shows the eigenvalues for all the possible “factors” that can be possibly extracted from the raw data. A standard criterion for selecting the factors is to retain only those whose eigenvalues are greater than 1. (See, for example, Kim & Mueller, 1978.) As can be seen, there are two eigenvalues greater than one; they are 3.175

and 1.376. Their corresponding factors account for, respectively, 47% and 20% of the total variation. It is rather rare to obtain factors that can explain such a significant amount of variation considering the raw data are binary data.



**Figure 5.22** A scree plot showing the eigenvalues of the extracted factors from the factor analysis.

In Table 5.7 and Table 5.8, test items that group together on these two factors are presented. As it shows, items Q6b, Q16b, Q22, Q23, Q26, Q27 and Q28 are all grouped on the first factor. It is interesting to notice that all of these items are either related to the energy principle or energy flow within a closed system—a special case with  $W_{\text{ext}} + Q = 0$ . Nevertheless, Q24 and Q25, which also require an application of the energy principle, are not included. (Q24 asks for the increase in the thermal energy of some water given external work and thermal transfer of energy; Q25 asks for a brief explanation for two moving clay balls smashing together.)

**Table 5.7 The first factor from the factor analysis.**

<b>Factor 1 (47% of variance)</b>			
<b>Item Number</b>	<b>Factor Loading</b>	<b>Concepts</b>	<b>Question Situation</b>
Q6b	0.55	Energy principle	Two asteroids passing each other
Q16b	0.25	Energy principle	Person in elevator moving up
Q22	0.45	Energy principle	Person pushing to stop box
Q23	0.23	Energy principle	Asteroid changing direction
Q26	0.41	Energy principle	Electron passing through box
Q27	0.44	Energy principle	Two pucks being pushed by springs
Q28	0.39	Energy principle	Particle moving in two time intervals
Q1	0.45	Relativistic $K$	Fast electron moving
Q2	0.30	Rest energy	Moving proton and electron
Q3a	0.41	Non-relativistic $K$	Two runners running opposite
Q6a	0.42	Gravitational $U_{\text{grav}}$	Two asteroids in deep space
Q8	0.36	Spring $U_{\text{spring}}$	Two springs being compressed

Also grouped together on the first factor are items Q1, Q2, Q3a, Q6a and Q8. Although it seems puzzling that how these items can be related, a closer inspection reveals that there is something in common about these items, that is, they all measure the concepts of the most basic forms of energy. In fact, the concepts that are measured in these items are no more complex than the definitions of rest energy (Q2), relativistic kinetic energy (Q1), non-relativistic kinetic energy (Q3a), gravitational potential energy (Q6a), and spring potential energy (Q8).

Based on the above discussions, it seems reasonable to describe the first factor as such that it measures the application of energy principle and the definitions of the basic forms of energy. Logically, it also seems sensible, since one has to grasp the definitions of the basic forms of energy first before being able to apply the energy principle.

As for the second factor, there are ten items clustered together: Q2, Q3a, Q6a, Q10a, Q10b, Q10c, Q11, Q12, Q21, and Q25. It is unclear as to what all these items have in common other than the fact that they all involve energy. However, if one includes only the six items that are most highly correlated to the factor, then what remains are Q6a, Q10a, Q10b, Q10c, Q11, and Q12. It now seems a bit clearer as to what these items have in common. As one can see, all of these items pertain to (the graphical representation of) gravitational/electric potential energy and their relations with kinetic energy in a closed system. Admittedly, there are other items in the energy assessment that can possibly be classified into this category as well but yet not included.

**Table 5.8 The second factor from the factor analysis.**

<b>Factor 2 (20% of variance)</b>			
<b>Item Number</b>	<b>Factor Loading</b>	<b>Concepts</b>	<b>Question Situation</b>
Q6a	0.37	Gravitational $U_{\text{grav}}$	Two asteroids passing each other
Q10a	0.58	$U_{\text{grav}}$ Graph	Closed system of two asteroids
Q10b	0.59	$K+U$ Graph	Closed system of two asteroids
Q10c	0.37	Bound/unbound state	Closed system of two asteroids
Q11	0.30	$K & U$ Graph	Two-particle system
Q12	0.38	Bound/unbound state	Binary-star system
Q2	0.24	Rest energy	Moving proton and electron
Q3a	0.27	Non-relativistic $K$	Two runners running opposite
Q21	0.24	Absorption spectrum	Objects with discrete energy levels
Q25	0.27	Energy principle	Clay balls smashing together

After all, the factor analysis extracts underlying factors based on student responses, which, under many circumstances, may not be as consistent as what the test designers hoped for. It is, therefore, not surprising to obtain factors that are either insignificant or significant but difficult to interpret. In this particular case, two significant factors are yielded, which together account for 67% of the total variance. Although it is almost impossible to fully understand why the test items are grouped as such, it at least seems sensible to interpret the two factors as discussed above.

## 5.6 Qualitative discussion of student interviews

The purpose of conducting student interviews is to probe students' reasoning behind their written answers. In the 2006 fall semester, nine student volunteers from four *M&I*

mechanics sections at NCSU participated in hour-long one-on-one interviews, where the students verbalized their explanations while working through the questions selected from the energy assessment. These selected questions include Q3a, Q5, Q6a, Q6b, Q7, Q8, Q9, Q12, Q14, Q15, Q16b, Q18, Q19, Q21, Q24 and Q27. Moreover, if time allowed for a student to complete extra questions, he/she then would be asked to answer the following: Q1, Q11, Q18, Q22, Q23 and Q26. Student interviews were videotaped and transcribed. Of the nine interview tapes, one has extremely low audio quality and thus is impossible to transcribe.

In this section, the qualitative discussion of student reasoning is presented through relevant excerpts from the interviews. For convenience, the discussion is organized according to the groups of questions that address the same test objectives.

#### 5.6.1 Application of the energy principle

Q16b, Q22, Q23, Q26, Q27 and Q6b involve the energy principle or the energy flow within a closed system. In answering these questions, students were, in fact, able to provide correct answers with satisfactory explanations as long as they chose to start from the energy principle. Nonetheless, if a student didn't invoke the energy principle, it is more than likely that this student would fail to provide a correct final answer.

Take Q16 for an example. This question describes a person moving up at a constant speed in an elevator and asks if the total work done on the person as a system by the earth and the elevator is positive, negative, or zero. As one can conclude by using the energy principle, the change in kinetic energy of the person is zero; thus the total work done on the person must be zero. However, many student interviewees didn't start from the energy principle; instead, they focused on the motion direction of the person. Consequently, these

students answered that the total work done on the person was positive and reasoned, in almost every case, as follows.

"The work done by the elevator and the earth is positive. Well, if they were zero, then the elevator wouldn't be moving... If they were negative, that would be the elevator going down."

"It's positive because the person is being pulled upward. If this person was coming down, you would say it's negative... If the person was coming down, you could say total work would be in negative direction, it would be negative."

"Since it's moving upward, then the work done by the elevator, uh, the total work has to be positive. The work done by the elevator is gonna be greater than the work done by the earth..."

One student, however, did mention that she ought to consider the fact that the speed of the person is constant. But she didn't proceed to invoke the energy principle and thus couldn't relate the constant speed with the change in kinetic energy and the total work done on the person.

"I feel like I should take into account of the fact that it's a constant speed. But I am not sure how. ...The elevator is moving the person up; the force is up and the displacement is up, so I feel like it should be a positive work. Then I am not sure whether it's the same or the work of the earth is greater... I can't think of how the fact it's at a constant speed makes a difference."

Another question Q22 also involves the energy principle. This question describes that a person leans against a wall with arms stretched and applies a pushing force to stop a box of mass  $m$  moving on a low-friction surface. Students are asked to figure out the work done on the box by the person given the initial speed of the box  $v$  and the final speed of the box zero. Clearly, the correct approach to this question is to use the energy principle, as some of the

student interviewees did. In particular, one student provided rather detailed and comprehensive explanations:

“So when I think of this question, I think of energy principle. So I’ll think of change in energy in a system equals work external plus  $Q$ , no thermal energy so that equals zero. Um, it has rest energy, but no mass is changing, so I won’t write that down... There’s nothing else: negligible friction, no change in  $E$  internal... So, I’ll just say change in kinetic energy equals work external (*writing down “ $\Delta K = W_{ext}$ ”*). Um, I guess I’ll write everything out:  $\frac{1}{2}mv$  final squared minus  $\frac{1}{2}mv$  initial squared equals work external (*writing down “ $\frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2 = W_{ext}$ ”*)... Finally the personal brings the box to a full stop. So, to me, that will mean the final velocity is zero. So that would cancel that term out (*crossing out “ $\frac{1}{2}mv_f^2$ ”*). So you’re left with initial kinetic energy (*writing down “ $-\frac{1}{2}mv_i^2 = W_{ext}$ ”*), which is why I put (d).”

Here, choice (d), which states that the amount of work done on the box by the person was  $-\frac{1}{2}mv^2$ , is the correct answer. From the above reasoning, it is clear that this student was able to invoke the energy principle and work through it consistently in answering the question.

Another student also used the energy principle to obtain a correct answer, although she doubted the relevance of the term “ $\frac{1}{2}mv^2$ ” in the choices at the beginning. Once she started considering the energy principle, her subsequent explanations flowed rather smoothly.

“I am not sure how to take the kinetic energy into account, like these answers have  $\frac{1}{2}mv$  squared. I am not sure what that means, since... Oh! I guess... delta  $E$  is the work external, which is delta  $K$  in this case (*writing down “ $\Delta E = W_{ext} = \Delta K$ ”*), because there’s no temperature change, and there’s no change in rest energy. So, since the final is, the final kinetic energy is zero, I guess that will be negative  $\frac{1}{2}mv$  squared is delta  $K$  (*writing down “ $\Delta K = -\frac{1}{2}mv^2$ ”*). Yeah, I think it’s (d).”

One student started from the energy principle but mistakenly confused  $K$  with  $\Delta K$ . As a result, he chose choice “(d) the amount of work done on the box by the person was  $+\frac{1}{2}mv^2$ ”

as the answer. In fact, this is the most popular wrong answer among the *M&I* students in the posttest, which accounts for 24% of the total responses.

"I use the energy principle, delta  $E$ , work external, since no temperature change, so... Delta  $E$  is work external. Energy is, energy associated with this is kinetic energy... There is only kinetic energy. That's gonna equal the work external, which is in this case, is the work done by the person. So, and the work done by the person is gonna be kinetic energy, which is  $\frac{1}{2}mv^2$  squared."

Later, this student realized that he confused  $K$  with  $\Delta K$  and quickly recovered from this mistake.

"Uh, wait. This is delta  $E$ , which is delta  $K$  is work external (*writing down* " $\Delta K = W_{ext}$ "). Final is zero, then there's gonna be a negative (*writing down* " $0 - \frac{1}{2}mv^2 = W_{ext}$ ").

Contrary to the above cases, some students didn't invoke the energy principle in answering the question. Rather, they only focused on the definition of work, which involves both force and displacement. However, this question doesn't provide any information about force or displacement. Consequently, these students could not determine the value of the work done on the box and ended up choosing either (b) positive work with indeterminate value or (e) negative work with indeterminate value—the other two popular wrong choices among the *M&I* students in the posttest, which together account for 36% of the total responses. As an example, the following excerpt shows one student's explanation.

"...You need more information. You'll need to know how, over what distance the person stops the box. You need to know from the initial contact to when it stops. And you could figure out force the person had to apply over that distance to stop it in that zone..."

Q23 is yet another question that requires the application of the energy principle. This question provides a situation where an asteroid was initially moving at a constant speed

toward one direction, and later the same asteroid was observed to be moving toward the opposite direction at the same constant speed. Given this information, students are asked to determine if the external work done on the asteroid during the two times was positive, negative, zero or not enough information to determine. Again, in order to answer this question correctly, one needs to invoke the energy principle and relate the change in kinetic energy of the asteroid to the external work done on it. As was observed in the interviews, students who started from the energy principle managed to obtain the final answer correctly.

The following excerpt is an example.

“Well, I think work external will be delta  $K$  in this case (*writing down “ $W_{ext} = \Delta K$ ”*). And speed is the same both times, so it has the same kinetic energy both times, so there’s no change in kinetic energy. So work external must be zero.”

Another student struggled a bit before she came to think of the energy principle. But immediately after she started from the energy principle, she quickly figured out the correct answer.

“Okay, so, I am thinking... I am trying to visualize this situation in my head... Um... It could be positive or negative. I am not sure... Because again I was thinking of the same equation: change in kinetic equals work external (*writing down “ $\Delta E = W_{ext}$ ”*). So... (*Writing down “ $\frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2 = W_{ext}$ ”*) I think it is the same thing really... My original thought was, I was thinking if it was moving towards the right, this would be positive (*pointing at “ $v_f$ ”*). But it’s squared, so the direction doesn’t matter. So, zero, zero... Yeah, zero, it’s my final answer.”

As this student admitted, her original focus was not on the change in kinetic energy, but rather on the direction of the velocity. Indeed, other student interviewees had the same issue. The difference, however, is that the above student resorted to the energy principle for help, while the others didn’t. Instead, they dwelled on the motion directions of the initial and final state of the asteroid. Since the question doesn’t specify the initial or final direction, they

either came up with an initial direction themselves and worked from it or simply answered “not enough information to determine”. For example, the following student defined the initial direction as positive.

“Initially it’s going at some direction. I might just say, uh, it’s direction... I would say it’s positive... Later going the opposite direction, which will be negative of the original direction. So, it’s moving exactly the opposite direction with the same speed. Obviously it was going the positive direction, now it’s going the negative direction. So, that means that had to be a force on this side, pushing the asteroid back into the negative direction. So that makes me think that the total work done on the asteroid was negative.”

Another student also focused on the motion direction and answered “not enough information to determine”; he explained:

“You’ll need to know what direction it was initially traveling in, and then you would know that if it was traveling in the positive direction to begin with, then the opposite direction later will be the negative. So you would say that negative force acting upon it. The total work that was done to it was in the negative direction. But because you don’t know what direction it was heading in to begin with, I would say not enough information.”

As for Q26, it involves both the energy principle and the momentum principle. This question provides the following situation: an electron enters a box along a certain direction with a speed  $1 \times 10^8$  m/s; later the electron leaves the box with the same speed  $1 \times 10^8$  m/s but different direction. Given this information, students are asked to conclude if the net force and net work done on the electron inside the box were zero, nonzero or not enough information to determine. As readers may have realized, if one applies the momentum principle and considers the momentum change of the electron, it immediately becomes clear that the net force was nonzero. Similarly, one can also conclude, by using the energy principle, that the

net work done on the electron was zero, as the kinetic energy of the electron remains the same.

For most student interviewees, this question is, in fact, quite challenging, for they need to distinguish between momentum and kinetic energy as well as to relate these quantities to net force and net work through the fundamental principles. Nevertheless, for students who started from the fundamental principles, they were often able to draw a correct final conclusion. Here is an example.

"I know the kinetic energy hasn't changed, because it has the same speed. Then I think the net work should be zero. So it's one of these two (*marking choice (b) and choice (d)*). But the net force wasn't zero, because it somehow changed direction. So, I think it should be (b)."

Another student invoked the momentum principle in considering the net force but didn't use the energy principle to find the net work.

"It has to be some force on it, and some force in the box. Um, it'll be able to change direction, otherwise like... if you have this principle, momentum principle (*writing down* " $\Delta\vec{p} = \vec{F}_{\text{net}} \Delta t$ "), uh...  $F_{\text{net}}$  has to be nonzero if you are gonna change the momentum. Momentum, uh, contains the directional; uh, it has direction. So... um, so this can't be zero (*pointing at* " $\vec{F}_{\text{net}}$ "). And then if the force isn't zero, and it's moving some distance in the block, then there's gonna be nonzero work on it, on the electron also."

Later, this student was asked to pay attention to the kinetic energy of the electron. At that moment, he realized that the initial and final kinetic energy of the electron is the same; then he started to think of the energy principle and came up with the correct answer regarding the work done.

"Okay (*writing down* " $\Delta E = W_{\text{ext}}$ ")... If there's change  $E$  (*writing down* " $\frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2 = W_{\text{ext}}$ ")... Oh! They are equal; that means that external has to be zero. Ok, so, they are equal..."

One student mistakenly related net force to the change in speed instead of the change in momentum (direction). Consequently, this student chose (d)—both the net force and the net work were zero—as the answer.

“Okay... so if it’s entering the box with the same speed as it’s leaving the box. I guess I put (d) because the speed isn’t changing... (Unintelligible)... Because the speed isn’t changing so it appears there wasn’t a force applied to it. There is nothing to make it change... That’s why I put (d)”

Q27 involves the energy flow within a system. This question provides a situation where two pucks of the same shape and size are launched along a low-friction surface using two identical springs with the same amount of compression. Given the information that puck 2 is twice as massive as puck 1, students are asked to compare the kinetic energy of these two pucks far away from the springs. To answer this question, students need to choose both the spring and the puck as a system. Since no external work is done on the system, the amount of the initial spring potential energy is equal to the amount of the final kinetic energy of the puck. Considering that both springs have the same amount of initial spring potential energy, it is clear that the amount of the final kinetic energy of the two pucks ought to be the same.

The most popular answer to this question, among the student interviewees, is that the kinetic energy of puck 1 is half the kinetic energy of puck 2. This finding is similar to that reported from the posttest, where choice (c)—it is half the kinetic energy of puck 2—is the most popular answer among the *M&I* students and accounts for 46% of the total responses. Student interviewees who supported this answer all started from the definition of kinetic energy  $K = \frac{1}{2}mv^2$ . Focusing on the mass difference between the two pucks, these students either overlooked the difference in speed between the two pucks, or simply assumed that the

speed for the two pucks is the same. As a result, they concluded that the more massive puck ought to have more kinetic energy as it leaves the spring. The following two excerpts illustrate such reasoning.

“So the kinetic energy of the first one will be  $\frac{1}{2}mv$  squared; the kinetic energy of the second one will be  $\frac{1}{2}2mv$  squared, which is just  $mv$  squared.”

“Kinetic energy formula is one half times mass times velocity squared... So strictly going by that, if you would plug the mass of puck 2 into this formula, it will give you a greater number than plug in the mass of puck 1, simply because puck 2 has the greater mass... Just look right off the bat, it doesn’t seem the speed of them is gonna be any different... They are both being pushed by a spring with the same force.”

Another popular answer among the student interviewees is that the kinetic energy of puck 1 is twice the kinetic energy of puck 2. This result is also consistent with that obtained from the written posttest, where around 21% of the *M&I* students selected this answer. Student interviewees who chose this answer realized that the speed for the two pucks was different. They claimed that the speed of the more massive puck—puck2—would be half the speed of the less massive puck—puck1. Therefore, using the definition of kinetic energy  $K = \frac{1}{2}mv^2$ , they found that puck 1 had twice the kinetic energy of puck 2. The following excerpt from one student is such an example.

“Because they are pushing off with same force, and since this one (*pointing at puck 2*) is, I don’t want to say heavy, but the mass is twice as much as mass 1, so I am just thinking that the velocity of this one (*pointing at puck 2*) should be half the velocity of that one (*pointing at puck 1*), since it’s pushing off the same force...”

Following this reasoning, the student then drew on the definition of kinetic energy and performed math calculations to find out that the kinetic energy of puck 1 was twice the kinetic energy of puck 2.

Although students like the above are sophisticated enough to notice a difference in speed between the two pucks, their conclusion on this matter is oversimplified. In this case, it is true that the puck of greater mass moves more slowly than the other, but the speed of a puck is not simply inversely proportional to the mass of the puck. In fact, it is completely unnecessary to pursue this information. As mentioned above, the key point of this question is the constant energy of the puck-spring system. Realizing the initial spring potential energy eventually turns into the final kinetic energy of the puck, one can easily reach the correct conclusion that the two pucks ought to have the same amount of kinetic energy.

Q6b is another question that involves the energy flow within a system. This question describes an isolated system of two asteroids passing each other. Initially these two asteroids are at a certain distance apart. Later, as the two asteroids are twice as far apart, this question asks students to determine how the gravitational potential energy and the kinetic energy of the system change.

Most student interviewees realized that the total  $K+U$  of the closed system is a constant. Therefore, they answered either (a)  $U$  increased,  $K$  decreased or (d)  $U$  decreased,  $K$  increased. Students who chose the correct answer (a) tended to relate this question to the energy graphs they learned in class. The following excerpts are some examples.

“Well, we learned on this as graphs that whenever  $U$  increases then (*drawing the K and U gravitational graphs*)... Whenever distance increases, that will mean that...  $U$  increases and  $K$  decreases.”

“I was thinking of that graph; that graph popped in my head (*drawing the U gravitational and K graphs*). So... as  $r$  increases, kinetic energy is gonna decrease and potential energy is gonna increase.”

For students who answered (d), it is possible that they mistakenly confused the magnitude of gravitational force ( $|\vec{F}_{\text{grav}}|$ ) with gravitational potential energy ( $U_{\text{grav}}$ ). As the following student explained, he thought the gravitational potential energy of the system would become smaller simply because the “gravitational pull would be weakened” as the asteroids were farther apart.

“Ok, the two are farther apart and this means that gravitational potential would have weakened, because you get further and further away, the gravitational potential would, the gravitational pull would be weakened. Therefore, I would assume that  $U$  would decrease. And since  $U$  is decreasing, you have to have a conservation of energy, so therefore the kinetic energy must increase to make up or the difference.”

Another two students explicitly mentioned that the expression of gravitational potential energy was “ $m_1 m_2$  over  $r$  squared”. Therefore, they concluded that the gravitational potential energy of the system would decrease as  $r$  increased. Moreover, these two students answered that the kinetic energy of the system would remain constant. As the following excerpts reveal, these two students couldn’t immediately see how a change in distance could affect the “velocity” of the asteroids; thus, they assumed that the “velocity” of the asteroids would remain constant.

“ $K$  should remain constant, because the mass doesn’t change and the velocity doesn’t change...  $K$  doesn’t depend on the radius between them.”

“ $K$  should stay the same, because the mass and the velocity, it doesn’t say anything about these, so I am assuming they remain the same still as before. So  $\frac{1}{2}mv$  squared should be the same...”

One noteworthy aspect regarding the question—Q6b—itself arises during the student interviews. In answering this question, one student correctly drew the  $U_{\text{grav}}$  graph and

explicitly mentioned that the curve becomes “less negative” as  $r$  increases. However, he mistakenly thought “less negative” meant “decrease”; thus he answered  $U_{\text{grav}}$  would decrease as  $r$  increased. Although this is the only case found in the student interviews, such an error may well indicate that the question needs further improvement in order to avoid the false negative (wrong answer based on correct reasoning). Possible suggestion for future revision includes using phrases “less positive” and “more negative” to replace the word “decrease”, as well as using phrases “less negative” and “more positive” to replace the word “increase”.

### 5.6.2 Determination of energy and interpretation of energy graphs

Q1, Q3a, Q5, Q6a, Q7, Q8, Q9, Q11 and Q12 involve the determination of different forms of energy and the interpretation of energy graphs. Student performance on these questions is rather satisfactory; they could, in most cases, provide not only correct answers but appropriate explanations as well. One aspect highlighted by the interviews is the fact that students were able to identify the important features of different energy forms qualitatively without having a formula sheet. Specifically, some students couldn’t remember a particular formula but still managed to answer the questions correctly through sensible reasoning. Nevertheless, it is also found during the interviews that students would make unnecessary mistakes due to their carelessness in answering some rather simple questions, for example, Q1.

This question describes an electron moving with an initial speed  $1.4 \times 10^8$  m/s and later with a final speed twice as fast. With this condition, the question asks how the final kinetic energy of the electron compares to its initial kinetic energy. Although students knew that the correct kinetic energy expression at high speeds is the relativistic form, many of them still

chose the wrong answer by considering the non-relativistic kinetic energy due to their carelessness. To be specific, most students first responded that the final kinetic energy of the electron would quadruple by thinking of the non-relativistic form of kinetic energy  $K = \frac{1}{2}mv^2$ . When further asked if this expression is always correct, students suddenly realized their mistake and quickly changed to the correct answer. The following excerpt is an example.

Student: "See, that's twice the velocity, that'll be four instead, uh, the kinetic energy will quadruple, because that's, because the square, so that's four times the initial kinetic energy. It'll be (c)."

Interviewer: "Is this always correct?"

Student: "Oh! Well, actually whenever it's traveling close to the speed of light, it involves gamma. So, now, it's actually not quite right. Or it might be greater than, probably it in fact is greater than four times kinetic energy when it's getting closer and closer to the speed of light. Uh, I would say (d), because it's getting closer and closer to the speed of light. It's actually not gonna be increment of the quadruple, because you have this extra function gamma inside."

One student, in fact, did consider the relativistic form and also attended to the initial speed  $1.4 \times 10^8$  m/s. Unfortunately, he only compared the initial speed of the electron with the speed of light and overlooked the final speed that is twice as fast; thus he concluded that the speed of the electron was not fast enough to use the relativistic form of kinetic energy.

"Umm, I think it's just moving faster, so it's not gonna be the same (*pointing at choice "(a) the same as the initial kinetic energy, because it's the same electron"*). Um, it's speed squared, so it's not just gonna double (*pointing at choice "(b) twice the initial kinetic energy, because it's traveling twice as fast"*). If the mass is double, then that will be true. But, and I don't think that's close enough to the speed of light to use the other formula. So that's why. Yeah, that's something like half of the speed of light, so I wouldn't...."

As this student mentioned, he was aware of “the other formula” and noticed that the initial speed of the electron was only “something like half of the speed of light”. But he didn’t consider the final speed, which is twice as fast and is indeed close to the speed of light.

Q3a describes a situation where two runners of the same mass  $m$  are running toward the opposite directions at the same speed  $v$ . Based on this situation, students are asked to determine the total kinetic energy of the two-runner system. Most students realized that the motion direction of the runners would not affect their kinetic energy; therefore they correctly concluded the total kinetic energy of the system ought to be the sum of the kinetic energy of each runner. The following excerpts are some examples.

“I am thinking the kinetic energy is  $\frac{1}{2}mv$  squared. And then I think it shouldn’t depend on the direction; it should just depend on the mass and the velocity. So it will be... the total kinetic energy should just be the sum of the kinetic energy of the first runner and the kinetic energy of the second runner, which will be  $\frac{1}{2}mv$  squared plus  $\frac{1}{2}mv$  squared, which should be  $mv$  squared, which is choice (c). I think it’s right.”

“My initial thought was that since they are running opposite directions, you just subtract them. But since this is velocity squared, the negative isn’t gonna make any difference.”

Nevertheless, the following two students mistakenly assigned a direction to the kinetic energy of each runner and answered that the total kinetic energy of the system was zero.

“I notice that each of them have the exact speed, which means both are running at the same actual speed, but running the opposite directions. So, they have exactly opposite velocities. And therefore I know they are gonna cancel each other out, because everything else stays the same as well.”

“Because they both have the same mass and velocity [speed] and they are heading

the opposite directions, then the final is gonna be zero... His will be positive one number; his will be negative the exact same number, so for the system that will be zero total.”

Q5 tests the pair-wise nature of potential energy by asking for the total electric potential energy of a four-proton system in a tetrahedron configuration. Interestingly enough, all of the student interviewees answered this question correctly with appropriate reasoning. In every case, the student realized that potential energy should be associated with the pair-wise interactions, and that counting the number of different pairs would yield the correct answer. The following excerpts are some examples.

“It should be the sum of all the potential energies. It should be one potential energy for every pair of protons.”

“I think of potential energy means you have to have two objects. For every two objects, there’s one potential energy. So, since they’ve already drawn it out here... I just counted the number of  $d$ ’s.”

“Um, counting up all the different interactions, within the tetrahedron there are 6.”

Q6a asks for the sign of gravitational potential energy in a context of two asteroids passing each other. Among the eight student interviewees, only two students correctly answered that gravitational potential energy ought to be negative. Rather interestingly, these two students didn’t try to recall the formula of gravitational potential energy; instead they both referred to the energy graphs that they learned in class. In particular, one of the two students explicitly commented that she couldn’t remember the exact formula. However, she drew the gravitational potential energy graph, and, based on the graph, she correctly concluded that gravitational potential energy should be negative.

"I remember it being, it's related to the distance between them and their masses. I can't remember the exact formula, but... The curve always looks like,  $U$  versus  $r$  always looks like something like that (*drawing the  $U$  gravitational curve below the  $x$  axis in an  $x$ - $y$  coordinate*). I remember something like that. And this is zero (*pointing at the  $x$  axis*). That's why."

Ironically, students who attempted to recall the formula of gravitational potential energy eventually provided a wrong answer. These students mistakenly remembered either " $G \frac{Mm}{r^2}$ " or " $G \frac{Mm}{r}$ " as the expression for gravitational potential energy. Therefore, these students answered that gravitational potential energy was greater than zero. The following excerpt illustrates an example.

"I said (a) because this is the system of two asteroids. So the gravitational potential energy between them would be  $m_1$  times  $m_2$  over distance squared, which will give you a positive... I don't really see how it could be less than zero; and it's not zero."

Another two students neither drew energy graphs nor tried to recall the formula. Rather, they simply stated that gravitational potential energy *should* be greater than zero.

"Noticing that the only gravity between the two asteroids will be between asteroid 1 and 2, you would notice that there has to be a gravitational potential... Therefore it's gotta be greater than zero."

"There's gonna be a potential energy, so it has to be greater than zero... I'll definitely not think this is gonna be less than zero, just because, it seems like it should always be some potential energy. There should never be like negative potential energy."

Q7 involves Rutherford scattering in which an alpha particle is scattered by a gold nucleus and moves along a trajectory. Given several marked points on the trajectory, students need to decide at which point the system of the alpha particle and the gold nucleus has the

greatest electric potential energy. Almost all the student interviewees provided a correct answer with appropriate reasoning. Interestingly enough, students often commented that they couldn't remember the exact formula and thus were not confident about their answers. Nevertheless, they did, in fact, know qualitatively how electric potential energy is in relation with charge and distance, which is indeed sufficient for them to answer the question correctly. The following excerpts from two student interviews illustrate this situation.

"Umm, I guess the way I try to think about the problem is, I was trying to think about the formula and try to see what that tells me, because when I think about things logically in physics, I never quite, I never get very far. So, I... The electric potential energy, as the radius increases, the potential energy, hmm, I guess, I guess. If I am remembering correctly, electric potential energy the radius is in the denominator. So that means as the radius increases, the energy will decrease... That logic will mean that (c) is the correct when they are closest together."

"Um, they are both positively charged, so they should be repelling each other. So, I think the potential will be greatest when they are closer together... So I think it should be (c). That's also something, I always forget these equations, but I think this is something over, oh, the charge of the first one times the charge of the second one over the radius."

One student answered that, at a point farther away from the gold nucleus before the alpha particle approaching the gold nucleus, the system would have the greatest electric potential energy. From his explanation as quoted below, it seems that this student may have mistakenly thought the interaction between the alpha particle and the gold nucleus was attraction.

"Now... thinking about it, potential energy is before, before it really starts moving, once you start moving, the kinetic energy picks up, and you'll lose some of your potential. And so, thinking about it, I would pick (a)..."

Both Q8 and Q9 aim to test spring potential energy in relation with compression. Q8 provides a context in which two springs of the same stiffness but different relaxed lengths are compressed the same amount. Given this information, students are asked to compare the spring potential energy of the two springs qualitatively. In a different context, Q9 describes two identical springs being either compressed or stretched through the same amount and, as well, asks students to compare the spring potential energy qualitatively. Student performance on these two questions is rather satisfactory; not only could they select the correct answer but also articulate the reasoning properly. In almost every case, the student was aware that spring potential energy ought to be independent of its relaxed length. The following excerpts, for example, are from two students in answering Q8.

“They’re both compressed by 2 cm, so they have the same change. And so if the change is the same, that’s all what it really matters. And the spring constant is the same, so they are gonna be equal to each other. It doesn’t matter how long the relaxed length is.”

“Just because you have different lengths to start with, doesn’t really matter, because they are both being compressed the exact same distance, and both have the same stiffness. So the change in length is the same even though they have different lengths. Therefore, I will pick (b), delta  $U_1$  is equal to delta  $U_2$ .”

Moreover, students knew that compression and stretch of a spring by the same amount would result in the same amount of spring potential energy. Consequently, all the students correctly answered Q9. In particular, one student candidly admitted being unable to remember the exact formula; nonetheless, this student still managed to answer this question with correct reasoning.

“I think that’s the same. I can’t remember what the formula is. But I don’t think it has something to do with the direction. I think it has something to do with the absolute value of the stretch or something. So, I think it will be the same.”

Q11 involves the energy graphs of a two-particle system. In the graph, the kinetic energy curve increases whereas the potential energy curve decreases as the distance between the two particles becomes greater. When the two particles are at  $r_1$  apart, the system has the same amount of kinetic energy and potential energy; thus the two curves intersect. Given this graph, students are asked to determine the kind of interaction between the two particles.

Among all the student interviewees, only two students answered correctly. In particular, one student invoked the expression of electric potential energy to aid in answering the question.

“It’s gonna be either repulsion or attraction. Um, see, as the distance increases, um, kinetic energy increases. So,  $U$  is positive, uh, it’s decreasing. Um... Oh! Ok, I guess it will be easier to write this out (*writing down* “ $\frac{1}{4\pi\varepsilon} \frac{q_1 q_2}{r}$ ”). And if the charges are the same, then this curve is gonna be positive, so if the charges are gonna be the same, they are gonna be repelled by each other. So, I guess that makes sense.”

Another student, who also answered the question correctly, attended to the kinetic energy curve. Based on the trend of the curve, he concluded that the interaction ought to be repulsion.

“I think as  $r$  increases, they are farther apart, the potential energy is smaller and kinetic energy is greater. So, I don’t think it’s either of these (*pointing at choice (c) and (d)*), because the curve behaves the same way before and after  $r_1$ . So I don’t think it’s (c) or (d). I think it’s repulsion. Because if it were attraction, then I don’t think kinetic energy would be increasing as farther away. So it makes sense that they are pushing each other away and their speeds are increasing.”

As opposed to the above student who was clear on the trend of the curves and not confused by the distance  $r_1$ , quite a few students thought the distance  $r_1$  should contain more meaning than it actually does. As such, these students selected either choice (c) or choice (d)

as an answer. Particularly, one student explicitly expressed her view on the distance  $r_1$  as follows:

“Ok, the turning point is  $r_1$ , because before  $r_1$  the potential energy is greater than the kinetic energy. And after  $r_1$  the kinetic energy is greater than the potential energy, which probably indicates something. So, I guess that’s why I was drawn to (c) and (d). Um, I guess (d) (*laughing*)...”

Q12 describes a system of two stars orbiting around each other in deep space, and asks students to determine if the sum of the kinetic energy and gravitational potential energy  $K+U$  of the system is greater, equal or less than zero. To answer this question, it is crucial to notice that this system is in a bound state, for the two stars do not escape from each other. In fact, almost all the student interviewees realized this fact except for one student who was completely off the track and considered  $K$  and  $U$  both as positive; thereby he concluded that  $K+U$  ought to be positive as well. Other than this individual, all other students either chose “(c)  $K+U$  is negative” or “(e)  $K+U$  is either negative or zero” as the answer. The reason for some students to choose (e) is that they mistakenly thought  $K+U=0$  also indicated a bound state. The following excerpt from one student illustrates this error.

“I think it has to be either negative or zero, because if they are orbiting around each other, then it’s a bound system. And if  $K+U$  were positive, it would be unbounded. I think it can be zero; zero will be; I think it still can be zero; it will be bounded. I know it’ll have to negative or zero, not positive. I think zero is included (*laughing*).”

Students who supported answer (c)—the correct answer—based their reasoning on the same fact that the system is in a bound state, but they were aware that  $K+U$  could not be zero for a bound state. The following excerpt is an example.

“Um, since neither is breaking away,  $K+U$  can’t be positive, because that would imply that one star would break away from the other star. Um,  $K+U$  would not be zero either, because there’s still... I think they’ll move.”

### 5.6.3 Determination of work and differentiation between work and heat

Q15, Q18, Q19 and Q24 require the determination of and differentiation between work and heat. Specifically, Q15 asks for the work done on a system that can be simplified as a point-particle system, whereas Q18 and Q19 address the work done on real systems. As for Q24, it involves both work and heat; thereby testing whether students can identify work and heat as two independent energy-transfer processes. Student performance on these questions is rather reasonable; some students demonstrated a fairly good mastery of basic concepts regarding work and heat, while others had trouble in dealing with these topics.

Q15 describes a process in which a truck pulls a trailer due east with a force of magnitude  $F$  over distance  $d_1$ , makes a sharp turn and continues to pull the trailer with a force of the same magnitude due north over distance  $d_2$ . With this information, the question asks for the work done on the trailer by the truck in the entire process. Most student interviewees answered this question correctly with satisfactory reasoning. These students all realized that work is a scalar not a vector, and that the work done on the trailer by the car in this case is simply the sum of the work done over the two distances. For example, the following student based his answer on such reasoning. What’s more, he provided comments on other choices as well, including choice (b)  $F\sqrt{d_1^2 + d_2^2}$ , choice (d)  $F\sqrt{d_1 d_2}$  and choice (e)  $\langle Fd_1, Fd_2, 0 \rangle$ .

“Umm, work is, work is force times distance. Umm, it’s additive. So it’s gonna be the... Let’s see. Okay, it’s the same force for both of them. That’s gonna be the, you just have to add these (*writing down “ $Fd_1 + Fd_2$ ”*). And (a), so, yeah, that answer is right. That makes sense... Umm well, (e) definitely not, because work is

scalar, and that's a vector form. Umm, I don't know where you would get the square root from; umm, you wouldn't have to do that. You just take each of them separately and add them together.

One student was debating between two choices—choice (a) and choice (b)—and was considering if he needed to use the Pythagorean theorem to determine the work done. Although he finally chose the correct answer (a), his struggle in this case may reflect the trouble that other students have, for 21% of the *M&I* students selected choice (b) as the answer in the written posttest. The following excerpt shows this student's reasoning.

"I would say still the force times the distance here plus the force times the distance here... I was pondering if you can make a Pythagorean of a triangle out of it. But I wasn't sure if I would mess around in the test. But in this case you would just say that work here squared plus work here squared (*writing down  $W_1^2$  and  $W_2^2$* ) equal to work here squared (*writing down  $W^2$  along the hypotenuse of the triangle*). And that was just, and that would give you the total work done in this whole, this diagonal that would be the total work done..."

Two students selected (e)  $\langle Fd_1, Fd_2, 0 \rangle$  as the answer, and they mistakenly thought that work was a vector. The following excerpt from one student illustrates this error.

"Okay, work equals force times displacement, and force is constant through the entire maneuver. Work  $F$  over distance 1 and distance 2... Okay, there's work done on the east direction which is  $x$ , and there's work done on the north direction which is  $y$ . So I believe you need to make these vectors. And (e) will be the answer... The work done in the  $x$  direction will be  $Fd_1$  and the work done in the  $y$  direction will be  $Fd_2$ ."

Q24 involves a work process, in which a beater stirs some water doing 200J of work on it, as well as a heat process, in which there is a 300J of heat from a stove fire into the same water. Given the above information, students are asked to determine the change in the thermal energy of the water after the water stops swirling and the stove is turned off.

Almost all the student interviewees correctly answered this question with appropriate reasoning. These students recognized that this question involves two independent processes that have changed the thermal energy of the water; therefore they obtained the correct answer by adding 200J and 300J together. Some students were rather careful in answering this question; they did not only clarify the work and heat processes but also detailed how to relate work and heat to the change in thermal energy. As an example, the following excerpt from one student evidently reflects the clear thoughts this student had in answering this question.

“Ok, well, since it tells you the thermal transfer of energy into the water is from the fire 300J. I assume that the 300 joules energy’s been transferred. And then you also have this 200 joules of work being done from the, uh, the beater to the water. And let’s see... I would write the general equation: change energy is equal to work exterior. And then I would notice here that thermal transfer of energy to the water is 300J, so I would write change of energy is equal to the work done, which is 200J by the beater, plus the Q which is the thermal transfer of energy into the water, which is 300J. And I would say the change in energy would be 500J, because that’s just simple addition of Q plus the work. And I would then say the thermal energy of the water has increased by 500J.”

One student answered the change in the thermal energy of the waster was zero, and he argued:

“If everything is turned off, the water is not even moving, it doesn’t seem like the thermal energy should increase at all. So I will go with (a) 0.”

In this particular case, the student appeared to equate thermal energy with work and heat and didn’t realize that the effect from the work and heat processes, which causes the increase in the thermal energy of the water, would remain in the system.

Q18 introduces a spring as a real system being stretched on both ends over the same small distance by two equal but opposite forces. Given this information, students are asked to

determine the work done on this real system. The majority of the student interviewees realized that the work done on the spring is nonzero although the net force is zero. Some students even spontaneously verbalized their comments on the difference between a real system and a point-particle system. As the following excerpt indicates, the student analyzed this situation using both real system and point-particle system; thereby providing rather convincing explanations for the answer.

“Well, have a spring at one length (*drawing a spring*), you pull it this way and this way at the same time, and the distance is the same... Oh, I think it’s  $2Fd$ , maybe, because it will be zero with the point particle system, because the center of mass doesn’t move. But with the real system you’ve done  $Fd$  work on this end and  $Fd$  on that end. So I think it should be  $2Fd$ . Yeah.”

Further, when asked what to attend to in considering the work done on a real system as opposed to the work done on a point particle system, the same student also provided satisfactory replies:

“Um, on the real system, you’ll always have to look at where the forces are actually being applied, whether it’s like the distance it moves is the actual distance this end moves, whereas in a point particle system the distance moves will be the distance the center of mass moves.”

Two students answered that the total work done on the spring was zero. As one student explained, the reason he chose this answer is that the net force applied on the spring is zero.

“Okay, you pull the same spring at each end over the same distance. Consider the spring, the total work is gonna be zero. Because if you... the pull is equal and opposite, you get the fore is zero on it. So the total work will equal to zero.”

Another student who answered zero work mistakenly thought that the work done on two ends would “cancel each other out”; therefore he concluded the total work done on the spring was zero.

Q19 addresses the similar topic regarding work done on a real system, but provides a different scenario—a car hitting a concrete wall. It is known that the average force applied on the car by the wall during the collision is of magnitude  $F$  and the car finally becomes  $\Delta L$  shorter. Then this question asks students to determine the work done on the car as a real system by the wall during the collision. In answering this question, students need to pay close attention to the contact point between the car and wall. Considering the fact that the contact point does not move during the collision, it is clear that the work done on the car by the wall is zero.

Two students realized that they should focus on the contact point rather than the configuration change of the car. In particular, one student related this question to a similar case discussed in class regarding an ice-skater pushing off a wall.

“This is the same concept. I was thinking of it in terms of the ice-skater and the human and wall. The wall isn’t doing any work on the car in this instance... But I was mainly thinking of the example she told us in class. With the ice skater and her arms change, but I mean, I guess... Okay, I guess the difference is if you are looking at a point particle system or a real system. If you are looking at the real system, the force is still being applied here (*pointing at the contact point between the car and the wall*) and it’s not moving. But if you are looking at the point particle system, the force will be on the center of mass and so it moves then.”

Most students didn’t attend to the contact point; rather they were distracted by the configuration change of the car and concluded that the work done on the car by the wall was either  $F\Delta L$  or  $-F\Delta L$ . One student, however, did notice that the “wall doesn’t move” during the collision and was struggling between choice (a)  $F\Delta L$  and choice (e) zero work, alas he couldn’t resolve the conflicts.

“Right now, I am deciding between saying that is zero and saying it’s  $F$  times delta  $L$ , because the wall doesn’t move. Yeah, the wall doesn’t move, so... I am unsure as whether you consider the force... I know work is force times distance, which is

applied. And you can say that because the wall doesn't move, it will be force times zero distance, which means the wall does zero work. But you could say that the wall does constant force  $F$  times delta  $L$ , because that's how far the car is compressed. I think it will be the distance will... I think it's (a). You will say that the wall did its work over distance delta  $L$ ... I am really torn between (a) and (e)."

Interestingly enough, many students did appear to know, in general, the difference between a real system and a point particle system. But they were unclear about what to attend to in determining the work done on a real system. In the interviews, students tended to think that dealing with the work done on a real system means taking into consideration the "actual length", particularly the change in the "actual length", of the real system. The following student, for example, explicitly mentioned what he would pay attention to in calculating the work done on the car as a real system in this question.

"When you are dealing with work done to a real system as opposed to a point particle system, you pay attention to the actual length of the entire car and any actual forces that might be in this situation."

#### 5.6.4 Specification of systems

Q14 involves the specification of appropriate systems. In particular, this question asks students to identify the kinds of energy involved in a pre-specified system—a falling stone. When answering this question, all the student interviewees were aware that they should not take gravitational potential energy into account, for there was no stone-Earth interaction in the system. This finding is consistent with that from Q5, in which all the student interviewees correctly determined the electric potential energy of a four-proton system in a tetrahedron configuration. To this end, it is reasonable to believe that the pair-wise aspect of potential energy has rather deeply ingrained into student minds. What's more, in answering question

Q14 some students spontaneously, and also correctly, provided comments on how to cope with the interaction between the stone and the Earth. Therefore, it becomes evident that students were clear about how different system specification would affect energy accountings. For example, the following excerpt from one student interview illustrates this case.

“Alright, it should have rest energy, because rest energy is  $mc^2$ . It should have kinetic energy, because it’s moving, so that’ll be  $\frac{1}{2}mv^2$ . And since only the stone is the system, there’s no gravitational potential energy, because that would be the stone and the Earth in the system, I believe. So it should just... So that, that gravitational potential energy would be work external (*writing down “ $\Delta E = W_{ext}$ ”*). So on the left side of it delta  $E$  it should just be (b) rest energy and kinetic energy.”

One error identified from this question is that some students mistakenly thought an object in motion would not have rest energy. As an example, the following excerpt from one student illustrates this error.

“... Since it is falling, probably it has some rest energy; it may not be very large amount of rest energy. Um, in fact it might not even have rest energy, since it’s totally falling at this point, and it’s actually not in a stationary state. So, I am gonna actually rule out it’s got rest energy at all. I’m gonna assume it got nothing but kinetic energy, and go with (e) none of the above.”

### 5.6.5 Calculation of atomic spectrum

Q21 aims to test whether students are able to deal with the discrete energy levels in calculating the absorption spectrum of some quantum objects in the ground state. Student performance on this question is rather satisfactory; most students knew that the absorption spectrum should correspond to the energy transitions from the ground state to the excited states. Moreover, these students were also aware that there would be no absorption lines due

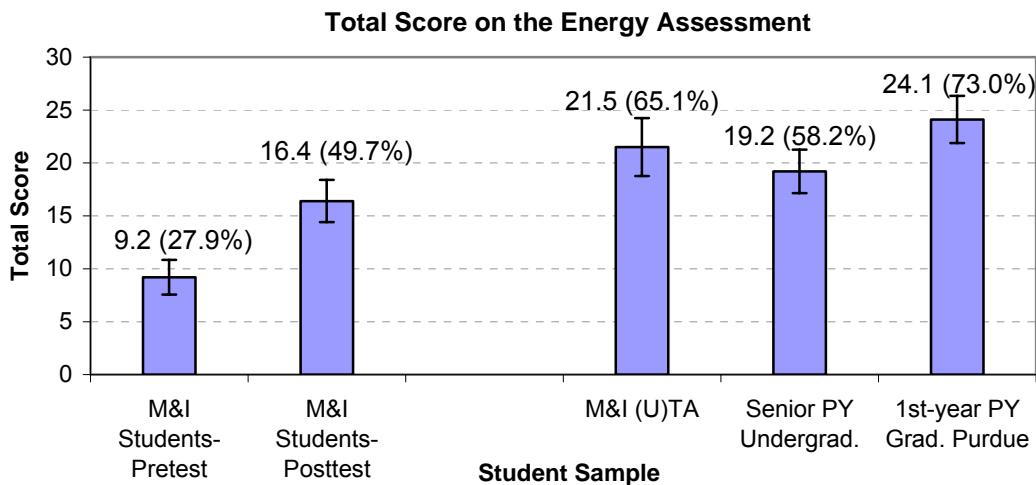
to the energy transitions among the excited states. One student even explained why he thought there would not be dark lines other than those corresponding to the jumps from the ground state.

“So, so you have some quantum, some, uh, yeah, some quantum objects in the ground state. (*Drawing multiple dots on the ground energy level*) Then the absorption is gonna be, uh, from going up to these energy levels (*pointing at the excited levels*) that are higher than this level (*pointing at the ground level*). And this happens through a very small amount of time; there’s no time for, uh, to go from here to here (*pointing from the first excited energy level to the second excited level*). So there’s gonna be only two absorption or dark lines. So this is the change 6ev, and this is change of 3ev. Yeah, so (c).”

Only one student provided an incorrect answer to this question, and he thought that the energy values of the dark lines were just the absolute values of the discrete energy levels in the energy diagram. In verbalizing his reasoning, this student revealed a deficiency in basic knowledge regarding the discrete energy levels. In particular, this student could not identify the ground state, which indeed is atypical among the *M&I* students.

## 5.7 Summary

This chapter reports the detailed results of student performance on the current version of the energy assessment. Specifically, four samples of students from different academic backgrounds completed the current energy assessment either on paper or on computer in the 2006 fall semester. For readers’ convenience, Figure 5.19 summarizes the average scores for the individual student samples.



**Figure 5.23 A summary of average total scores for four samples of students.**

In particular, the *M&I* students, who were then taking the *M&I* mechanics classes at NCSU, took both the pre and post-test of the energy assessment. It is found that student pretest performance is statistically better than what random guessing would produce; thus suggesting that students may have some prior knowledge regarding energy before instructions. Nevertheless, the item scores, which are the percentages of correct responses for the individual questions, are generally low. Further inspections on the objective scores—average item scores for groups of questions that address the same test objective—reveal that prior to the course instructions students appear to lack basic knowledge on the discrete energy levels, work, heat and the energy principle.

In the posttest, student performance is noticeably improved. Overall, the posttest average is close to 50% of the perfect score; individually, most item scores are above 0.25 with quite a few near or above 0.6. As for the objective scores, all values are near or above 0.35.

In order to detect the effect of course instructions on student performance, the pairwise comparisons are conducted to examine the gains ( $\text{gain} = \text{post score} - \text{pre score}$ ) in total score, item score and objective score respectively. Results show that there is a statistically significant gain in total score for the four *M&I* sections collectively as well as individually. Simply put, the course instructions have indeed promoted student overall performance on the energy assessment. Moreover, it is also found that the gain in total score is not statistically different across the four sections; thus suggesting that student better performance in the posttest is independent of the “section” factor.

For the individual questions, the gains in item scores are demonstrated to be significant for most items; therefore, it is reasonable to believe that the course instructions have enhanced student performance on most of the individual topics covered in the energy assessment. As for a few questions whose gains in item score are shown to be insignificant, it is interesting to find that these questions all aim to test the application of the energy principle.

Although some questions have insignificant gains in the item scores, the gain in the objective score for “application of the energy principle” is shown to be significant. In other words, students may have failed to provide correct answers to some particular questions in the posttest, but their overall performance on the objective “application of the energy principle” has improved, and such improvement is not by accident. Similarly, the gains in other objective scores are also found to be significant; therefore, it is valid to believe that the course instructions have promoted student performance on all the objectives covered in the energy assessment.

Results obtained from *M&I* (U)TAs, senior physics major undergraduates and first-year physics major graduate students at Purdue provide useful baseline data on students at higher academic levels. As the results show, it seems possible for a well-prepared student to achieve greater than 90% of the perfect score on the energy assessment. Interestingly enough, the *M&I* TAs and UTAs demonstrate rather similar performance on the assessment and thus seem to have a comparable mastery of physics knowledge regarding the energy topics. Moreover, the performance of the *M&I* TAs is also found to be similar to that of the first-year physics major graduate students at Purdue.

Based on the data collected from the students who completed the current energy assessment, the factor analysis of student responses is performed, and it yields two meaningful factors that together account for 67% of the variance. The first factor seems to be related to the energy principle and the most basic energy forms; the second factor appears to address the (graphical representation of) potential energy and kinetic energy in a closed system.

In order to probe student reasoning behind their answers, a small number of student interviews were conducted. One aspect highlighted by the interviews is that students are, in fact, able to relate the change in energy of a system to work and heat, so long as they choose to start from the energy principle. Another aspect reflected in the interviews is that students are able to answer the questions correctly without using the exact formulas but rather through sensible reasoning on the qualitative relations among different physical quantities.

In general, student interviewees appear to have a fairly good mastery of the topics regarding different forms of energy, appropriate systems and discrete energy levels.

However, some students show weakness in dealing with work, particularly the work done on real systems.

## **CHAPTER 6: Conclusion**

This study aims to design an energy assessment that is appropriate for the *M&I* mechanics course to evaluate student understanding of the energy topics. During the study, great efforts have been devoted to the development of the energy assessment as well as to the investigation of student performance on this assessment. To conclude the findings from the study, this chapter first summarizes the major results that are detailed in the previous chapters, then describes the implications for course instructions on the energy topics, and finally discusses possible directions for future work on this subject.

### **6.1 Summary of major results**

#### 6.1.1 Results regarding the energy assessment

Primarily, this study yields two major outcomes regarding, respectively, the current energy assessment and student performance on the energy topics. As for the former, this study does not only provide evidence on the validity and reliability of the assessment, but also examines the content/cognition levels and the reasoning steps of the individual items. As such, the current energy assessment is demonstrated to be an appropriate standardized tool to evaluate student performance on the energy topics.

*The current energy assessment is a valid and reliable instrument.* The validity of the assessment was established through careful design of the test objectives and rigorous review of the test items. The test objectives addressed in the current energy assessment match the goals of the *M&I* mechanics course and mainly include the following five aspects: application of the energy principle, determination of various forms of energy, specification of

appropriate systems, determination of and differentiation between work and heat, and calculation of absorption/emission spectrum. Test items, which are accordingly designed, were subjected to two rounds of review—expert review and faculty review—to ensure the appropriate subject matter and adequate coverage. Thereby, both the face validity and the content validity of the energy assessment are established.

As for the reliability of the assessment, item/test analysis is performed to evaluate the discriminatory power and the consistency of the energy assessment through five statistical measurements: item difficulty index, item discriminatory index, point bi-serial coefficient, KR-20 reliability index and Ferguson's delta. These five measurements, which are summarized in Table 4.2, indicate that the current energy assessment is a reliable and discriminatory test.

*The current energy assessment is such a test that requires higher level of thinking in both content and cognition.* Indeed, the current energy assessment by and large aims at the application of fundamental principles or basic concepts. Among the thirty-three items in the current energy assessment, nine items, which account for nearly one third of the entire assessment, require the application of the fundamental principles, particularly the energy principle. Another twenty items, which account for 60% of the entire assessment, target at the application of basic concepts. Only four items involve the comprehension of concepts, and all these four items pertain to the interpretation of energy graphs. Evidently, the energy assessment is not a test that focuses on the lower level of thinking, i.e. recall a formula. Rather, this assessment requires higher-level thinking both in content and cognition.

*The current energy assessment is a test in which only a short reasoning process is involved in the individual items.* The advantage of short reasoning for standardized multiple-

choice items is that the interpretation of student responses is facilitated. If students fail to answer a particular multiple-choice question correctly, it is often more challenging to pinpoint student difficulties if the question involves multiple steps of reasoning than if it involves only one step of reasoning. To this end, the small number of reasoning steps is often preferred for multiple-choice questions. For the current energy assessment, the number of reasoning steps involved in each item is determined based on the number of physical quantities needed to link the given information and the unknown information. According to this scheme, it is found that twenty items, which account for 60% of the entire assessment, involve zero-step reasoning, and another twelve items, which account for 36% of the entire assessment, require one-step reasoning. Only one item—Q26—requires a two-step reasoning process, in which the application of both the energy principle and the momentum principle is necessary. Interestingly enough, the relatively large number of reasoning steps for Q26 does not seem to have a negative impact on the item statistics (See Table 4.4).

### 6.1.2 Results regarding student performance on the energy assessment

Aside from the above results regarding the energy assessment, another major outcome yielded from this study pertains to student performance on the energy assessment. For the *M&I* students in particular, reported in this study are detailed quantitative results of student pre and post performance as well as the qualitative discussions of student reasoning. As such, this study provides useful insights into student learning outcomes as well as the course effectiveness.

*Prior to the course instructions, the M&I students appeared to lack adequate knowledge regarding energy, although their pretest performance was better than random*

*guessing.* As the results show, student pretest performance was statistically better than what random guessing would produce. However, student pretest performance on the individual items and test objectives was generally poor. Specifically, students seemed to lack basic knowledge on discrete energy levels, work, heat and the energy principle.

*After the course instructions, the M&I students achieved a significant improvement in performance on the entire energy assessment as well as on the individual items and test objectives.* In fact, improved student performance on the entire assessment was consistent across all four sections that were taught by three different instructors. In this respect, it is reasonable to consider that these four *M&I* sections are equally effective in promoting student overall performance on the energy topics. Additionally, student performance on the individual questions was improved as well. For most questions in the posttest, the increase in the percentage of correct responses is proven to be significant; thus suggesting that the course instructions have enhanced student performance on most of the topics covered in the assessment. As for a few questions on which student posttest performance didn't show a significant improvement, they all require the application of the energy principle. This result notwithstanding, the *M&I* students still achieved, in the posttest, a significant improvement in the individual test objectives, including "application of the energy principle". It is, therefore, sensible to conclude that the course instructions have promoted student performance on all the test objectives, including "application of the energy principle".

Moreover, a small number of student interviews that were conducted among the *M&I* student volunteers provide useful information on student reasoning. Three aspects highlighted by the interviews are worthwhile mentioning, and they are summarized as follows.

*M&I students were, in fact, able to apply the energy principle correctly in answering the questions if they chose to consider this fundamental principle.* To a large extent, it is often not an issue whether or not students can use the energy principle in an appropriate manner. Rather, what seems to be an issue is whether or not students consider the energy principle, when needed, as a starting point to tackle the relevant problems. As observed from the interviews, rarely did those students who started from the energy principle fail to provide satisfactory responses to the relevant questions, nor did those who didn't start from the energy principle succeed in presenting correct answers.

*M&I students were able to answer questions correctly without using the exact formulas but rather through qualitative reasoning on the relationships among the physical quantities.* As observed in the student interviews, those students who could not recall exact formulas were still able to provide correct answers with reasonable explanations in answering the qualitative questions. Ironically, those students who tried to recall certain formulas failed on some rather simple questions. One noteworthy aspect reflected in the interviews is that students seemed to lack confidence in their own answers, albeit correct, without having at hands the exact formulas.

*M&I students, in general, showed a reasonable mastery of all the topics covered in the energy assessment after the course instructions. However, students had noticeable difficulties in dealing with the work done on real systems.* It is found that in determining the work done on a real system students often focused on the configuration change of the real system instead of on the contact point where force was applied.

*Furthermore, results from students at higher academic levels show that it is possible for a well-prepared student to achieve a score higher than 90% of the perfect score on the*

*current energy assessment.* These students who were from higher academic levels included the *M&I* undergraduate teaching assistants, *M&I* graduate teaching assistants, senior physics major undergraduate students<sup>12</sup> at NCSU and first-year physics major graduates at Purdue. Their participation in the energy assessment, like the participation of the *M&I* students, provides valuable baseline data on the current energy assessment.

Finally, *a factor analysis of student responses to the energy assessment yields two meaningful factors, which together account for 67% of the variance.* The first factor seems to be concerned with the energy principle and the basic forms of energy, while the second factor appears to relate to (the graphical representation of) the gravitational/kinetic energy of a closed system. A caveat of the factor analysis is that the results are highly dependent on student responses, which, under most circumstances, generate either insignificant factors or significant but hard-to-interpret factors. (See, for example, Heller & Huffman, 1995; Hestenes & Halloun, 1995)

## 6.2 Implications for instructions

As the above results suggest, the *M&I* mechanics course is rather successful in promoting student performance on the energy topics. To maintain and advance such a success, findings from this study, particularly from the student interviews, provide some practical implications for the instructions.

In order to encourage students to start from the fundamental principles, instructors may first need to promote student confidence in these fundamental principles throughout the

---

<sup>12</sup> One noteworthy aspect regarding the physics major undergraduates is that the highest score of this sample was only 82%. Also note that their average score 58% is not considerably higher than the average of the *M&I* students.

entire course sequence. In conveying the essence of the energy principle, instructors could purposefully lead students' attention to the causal relationship between the energy transfer processes ( $W$  and  $Q$ ) and the change in energy of a system ( $\Delta E_{\text{sys}}$ ). Continual practice on the energy principle in various contexts may be helpful in familiarizing students with its application. In particular, practice questions that require students to infer the energy transfer processes given the change in energy are just as helpful as questions that require students to determine the change in energy given the energy transfer processes. As for the format of the practice questions, qualitative and conceptual questions could be rather beneficial on both the students' side and the instructors' side.

In fostering student conceptual understanding of the energy topics, it is useful that students learn to analyze physical situations qualitatively without frequent reference to the formula sheet. Instructors may need intentionally to emphasize that blind use of special case formulas is not a healthy way to practice physics, and that qualitative analysis is a feasible and often an effective way to tackle certain physics problems. As for the fact that students feel less comfortable without exact formulas at hands when answering questions, instructors need to help students build necessary confidence in this matter. What's more, instructors could encourage students not to use formula sheets in doing qualitative questions in homework.

As for the topic regarding work done on a particular system, it is often quite challenging for many students, for it requires different treatments depending on different systems—point particle system or real system. For real systems in particular, it is important for students to realize that when calculating the work done on a real system they should attend to the contact point where the external force is applied. Instructors may need to clarify

explicitly that to determine the work done on a real system does not necessarily mean to take into consideration the actual length change, shape change or other kinds of configuration change of the system. In many cases, the configuration of a real system can change dramatically, but the work done on the system is still zero. For this type of physical situations, instructors can design various questions for students to practice either in class or after class.

### **6.3 Future work**

There are a variety of research directions worth pursuing to extend the current study. One immediate example is to concentrate on the further development of the current energy assessment. Aside from the possible minor revisions, it is interesting to consider a computer-based version with animations. As readers may notice, nearly one-third of the questions in the current energy assessment require the application of the energy principle, and almost all of these questions involve situations where objects are in motion. In the current energy assessment, all these situations are described in words and are sometimes accompanied by static pictures. Indeed, computer animations can be another means to present these situations and may even facilitate student responses. One study on the animations for the Force Concept Inventory has reported the potential advantage of using animations for questions regarding Newton's laws (Dancy, 2000). As a further extension on this topic, it is worthwhile to study the effect of animations on student responses to the energy related questions, i.e. questions in the current energy assessment.

Another topic worth pursuing is the effect of symbolic and numerical representations in physics questions on student performance. Although an exploratory experiment in this

study has provided some preliminary results on this matter, more rigorous research is needed to answer if and how the different representations affect student performance. One suggestion for future research on this topic is to take into consideration the difficulty level of questions, as it may confound the results.

Also worth pursuing is an assessment on energy that is appropriate for both the *M&I* mechanics course and the traditional mechanics course. In so doing, it is feasible to evaluate students from different classes and thus to compare the effectiveness of different curricula on the energy topics. For studies on this topic, the current energy assessment in fact has laid a practical foundation, since quite a few questions can be regarded as appropriate for the traditional mechanics classes. These questions include Q3a, Q3b, Q6a, Q6b, Q4, Q8, Q9, Q15, Q16a, Q16b, Q22, Q23, Q26, Q27 and Q28.

Finally, it is important to note that the *ultimate* goal of all these aforementioned studies is to promote student understanding of the energy topics. Also remember: one *fundamental* piece of the ultimate goal for the *M&I* mechanics course is that students are able to apply the energy and the momentum principle. To this end, it can be extremely beneficial for the researchers to identify and study the factors, if any, that would possibly help cue students into a spontaneous application of the fundamental principles when answering relevant questions. This, indeed, can be an excellent example of what we tirelessly encourage our students to do—“start from the fundamentals”!

## REFERENCES

- Adkins, C. J. (1981). A national core syllabus for A-level physics. *Physics Education*, 16(3), 128-135.
- Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences*, 3rd ed. Upper Saddle River, N.J.: Prentice Hall.
- Anderson, L., & Krathwohl, D. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Anderson, L., & Sosniak, L. (1994). *Bloom's Taxonomy: A Forty-year Retrospective*. Ninety-third Yearbook of the National Society for the Study of Education. Part II. Chicago: University of Chicago Press.
- Arons, A. (1989). Developing the energy concepts in introductory physics. *The Physics Teacher*, 27, 506-517.
- Arons, A. (1999). Development of energy concepts in introductory physics courses. *American Journal of Physics*, 67(12), 1063-1067.
- Aubrecht, II, G. & Aubrecht, J. (1983). Constructing objective tests. *American Journal of Physics*, 51(7), 613-620.
- Bauman, R. P. (1992). Physics that textbook writers usually get wrong II: heat and energy. *The Physics Teacher*, 30, 353-356.
- Beichner, R. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, 62(8), 750-755.
- Beichner, R. J., Saul, J. M, et al. (2000). *Introduction to SCALE-UP: Student-Centered Activities for Large Enrollment University Physics*. Proceedings of the 2000 Annual meeting of the American Society for Engineering Education.
- Bloom, B., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. Handbook I: Cognitive domain*. New York: Longman.
- Bork, A. (1984). Letter to the editor. *American Journal of Physics*, 52(10), 873-874.
- Chabay, R. & Sherwood, B. (1997). Qualitative understanding and retention. *AAPT Announcer*, 27(2), 96.
- Chabay, R. & Sherwood, B. (1999). Bringing atoms into first-year physics. *American Journal of Physics*, 67(12), 1045-1050.

Chabay, R., & Sherwood, B. (2002). *Matter & interactions I: Modern mechanics*. New York: John Wiley & Sons, Inc.

Chabay, R., & Sherwood, B. (2007). *Matter & Interactions I: Modern mechanics*. New York: John Wiley & Sons, Inc. Second edition.

Chabay, R. & Sherwood, B. (2004a). Modern mechanics. *American Journal of Physics*, 72(4), 439-445.

Chabay, R. & Sherwood, B. (2004b). Fundamental principles in introductory physics. Presented at the 2004 International Conference on Physics Education, Durban, South.

Chabay, R. & Sherwood, B. (2005). Matter & Interactions. To appear in *PER-based reform in university physics*, edited by Redish and Cooney (AAPT).

Crombie, A. C. (1959). *Medieval and early modern science*. Garden City, NY: Doubleday.

Dancy, M. (2000). *Investigating animations for assessment with an animated version of the force concept inventory*, Ph.D. dissertation. North Carolina State University.

Ding, L., Chabay, R., Sherwood, B., & Beichner, R. Evaluating an E&M assessment tool: Brief Electricity & Magnetism Assessment (BEMA). To appear in *Physical Review Special Topics: Physics Education Research*.

Doran, R. (1980). *Basic measurement and evaluation of science instruction*. Washington D.C.: National Science Teachers Association.

Engelhardt, P. & Beichner, R. (2004). Students' understanding of direct current resistive electrical circuits. *American Journal of Physics*, 72(1), 98-115.

Ghiselli, E., Campbell, J., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco: Freeman.

Goldring, H. & Osborne, J. (1994). Students' difficulties with energy and related concepts. *Physics Education*, 29, 26-32.

Haladyna, T. (1997). *Writing test items to evaluate higher order thinking*. Needham Heights, MA: Allyn & Bacon.

Haladyna, T. (1999). *Developing and validating multiple-choice test items 2nd ed*. Mahwah, New Jersey: Lawrence Erlbaum.

Halloun, I., & Hestenes, D. (1995). *The search for conceptual coherence in FCI data*. Retrieved 2005, from <http://modeling.la.asu.edu/R&E/CoherFCI.pdf>

Heller, P. & Huffman, D. (1995). Interpreting the Force Concept Inventory. *The Physics Teacher*, 33, 503-511.

Henderson, C. (2003). Common concerns about the Force Concept Inventory. *The Physics Teacher*, 40, 542-547.

Hestenes, D. & Halloun, I. (1995). Interpreting the Force Concept Inventory: A response to Huffman and Heller. *The Physics Teacher*, 33, 502-506.

Hestenes, D. & Swackhamer G., (unpublished) Energy Concept Inventory.

Hestenes, D. & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher*, 30, 159-166.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, 30, 141-158.

Huffman, D. & Heller, P. (1995). What does the Force Concept Inventory actually measure? *The Physics Teacher*, 33, 138-143.

Kim, J., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it*. Beverly Hills, CA: Sage Publications.

Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. London: Methuen.

Krathwohl, D. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory into Practice*, 41(4), 212-218.

Lawson, R. & McDermott, L. (1987). Student understanding of the work-energy and impulse-momentum theorems. *American Journal of Physics*, 55(9), 811-817.

Leff, H. & Mallinckrodt, A. (1993). Stopping objects with zero external work: Mechanics meets thermodynamics. *American Journal of Physics*, 61, 121-127.

Linn, R., & Gronlund, N. (2001). *Measurement and assessment in teaching*. Columbus, OH: Merrill.

Loverude, M., Kautz, C., & Heron, P. (2002). Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas. *American Journal of Physics*, 70(2), 137-148.

Mallinckrodt, A. & Leff, H. (1992). All about work. *American Journal of Physics*, 60(4), 356-364.

Mallinckrodt, A. & Leff, H. (2002). Work, energy and kinematic car data. *The Physics Teacher*, 40, 516-517.

Maloney, D., O'Kuma, T., Hieggelke, C., & Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics*, 69(7), S12-S23.

Mehrens, W., & Lehmann, I. (1991). *Measurement and evaluation in Education and Psychology*, 4th ed. Fort Worth: Rinehart and Winston.

Meltzer, D. (2004). Investigation of students' reasoning regarding heat, work, and the first law of thermodynamics in an introductory calculus-based general physics course. *American Journal of Physics*, 72(11), 1432-1446.

Mungan, C. E. (2005). A primer on work-energy relationships for introductory physics. *The Physics Teacher*, 43, 10-16.

Mungan, C. (2007a). Defining work. *The Physics Teacher*, 45(5), 261.

Mungan, C. (2007b). Thermodynamics of a block sliding across a frictional surface. *The Physics Teacher*, 45(5), 288-291.

Patton, M. (1990). *Qualitative evaluation and research methods*, 2nd ed. Newbury Park, Calif.: Sage Publications.

Penchina, C. (1978). Pseudowork-energy principle. *American Journal of Physics*, 46(3), 295-296.

Powell, J. & Gillespie, C. Assessment: All tests are not created equally. *Paper Presented at the Annual Meeting of the American Reading Forum (11th, Sarasota, FL, Dec 12-15, 1990)*.

Pride, T., Vokos, S., & McDermott, L. (1998). The challenge of matching learning assessments to teaching goals: An example from the work-energy and impulse-momentum theorems. *American Journal of Physics*, 66(2), 147-157.

Risley, J. D. (1999). WebAssign: Assessing student performance any time any where. *UniServe Science News*, 13. Available at  
<http://science.uniserve.edu.au/newsletter/vol113/risley.html>

Sandin, T. (1985). On not choosing multiple choice. *American Journal of Physics*, 53(4), 299-300.

Scott, B. (1985). A defense of multiple choice tests. *American Journal of Physics*, 53(11), 1035.

- Sherwood, B. (1983). Pseudowork and real work. *American Journal of Physics*, 51(7), 597-602.
- Sherwood, B. & Bernard, W. (1984). Work and heat transfer in the presence of sliding friction. *American Journal of Physics*, 52(11), 1001-1007.
- Sherwood, B. & Chabay, R. (1999). The point-particle system. *AAPT Announcer*, 29(2), 107.
- Sherwood, B. & Chabay, R. (2004). Choice of system and the energy equation. *AAPT Announcer*, 34(2), 163.
- Singh, C. & Rosengrant, D. (2003). Multiple-choice test of energy and momentum concepts. *American Journal of Physics*, 71(6), 607-617.
- Swackhamer, G. & Dukerich, L. (2003). An energy concept inventory. *AAPT Announcer*, 33(2), 90.
- Thornton, R. & Sokoloff, D. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *American Journal of Physics*, 66(4), 338-352.
- Varney, R. (1984). More remarks on multiple choice questions. *American Journal of Physics*, 52(12), 1069.
- Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education*, 1, 205-221.
- Warren, J. (1972). The teaching of the concept of heat. *Physics Education*, 7, 41-44.

## **APPENDICES**

## **Appendix A: Faculty Review**

Dear faculty members:

You are invited to help validate an assessment instrument that aims to evaluate students' understanding of energy. This assessment instrument is a multiple-choice test that covers the basic topics of energy taught in the *M&I* mechanics course, including forms of energy, system selection, work, heat, and the energy principle.

For each question, please check one of following three categories: good, neutral, and poor. If you check "poor" for a specific question, please also select the reason(s) why you think it is a poor question. Reasons include "not covered", "not relevant", "not important", "too hard", "too easy", and "others (please specify)".

Feel free to comment on any question or on the test as a whole. The following questions are some examples:

1. Are there any important concepts missing in the test? If yes, please list the missing concepts.
2. Are there any concepts (topics) covered by too many questions? If yes, please list these concepts (topics) with at least one related question for each concept (topic).
3. Are there any concepts (topics) covered by too few questions? If yes, please list these concepts (topics).

Thank you for your time and input.



## **Appendix B: Current Version of the Energy Assessment**

**1.** An electron was initially traveling at a speed of  $1.4 \times 10^8$  m/s. Some time later, the same electron is observed to be traveling twice as fast. What is the kinetic energy of the electron at this time?

- (a) It is the same as the initial kinetic energy, because it's the same electron
- (b) It is twice the initial kinetic energy, because it's traveling twice as fast
- (c) It is four times the initial kinetic energy, because kinetic energy is proportional to speed squared
- (d) None of the above, because its speed is close to  $3.0 \times 10^8$  m/s, the speed of light

**2.** A proton and an electron are moving at the same speed. Which one has greater rest energy?

- (a) Proton, because it has greater mass
- (b) Electron, because it has less mass
- (c) They have the same nonzero rest energy, because both are traveling at the same speed
- (d) They both have zero rest energy, because neither is at rest
- (e) None of the above

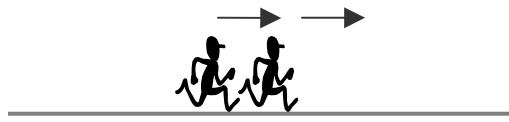
**3a.** Runner A is running due east and runner B is running due west. They have the same mass  $m$  and speed  $v$ . Consider the two runners as a system; what is the kinetic energy of the system?

- (a) 0
- (b)  $\frac{1}{2}mv^2$
- (c)  $mv^2$
- (d)  $mc^2$
- (e)  $2mc^2$



**3b.** Now both runners are running due east with speed  $v$ . What is the kinetic energy of the system?

- (a) 0
- (b)  $\frac{1}{2}mv^2$
- (c)  $mv^2$
- (d)  $mc^2$
- (e)  $2mc^2$

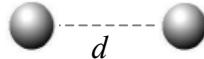


**4.** Consider three moving balls.

Ball 1 moves with a constant speed in a constant direction.  
Ball 2 moves with a constant speed, but its direction changes.  
Ball 3 moves in a constant direction, but its speed changes.  
Which of these balls has (have) a constant kinetic energy?

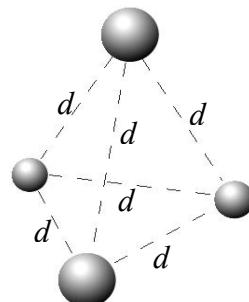
- (a) Ball 1
- (b) Ball 1 and ball 2
- (c) Ball 2 and ball 3
- (d) Ball 1 and ball 3
- (e) Ball 1, ball 2, and ball 3
- (f) None of the above

**5.** When two protons are a distance  $d$  apart, the electric potential energy of the two-proton system is  $U$ .



Now another two protons are taken into account. These four protons form a tetrahedron with a distance  $d$  between any two of them. (See the diagram below.) The electric potential energy of the four-proton system is:

- (a)  $U$
- (b)  $2U$
- (c)  $3U$
- (d)  $4U$
- (e)  $6U$
- (f)  $8U$
- (g)  $12U$



At a location in space far from any other objects, two moving asteroids (asteroid I and asteroid II) pass each other. Consider the two asteroids as a system, and answer the following two questions.

- 6a.** At the instant shown in the diagram, the gravitational potential energy of the system is  $U_1$ . Which of the following is true?

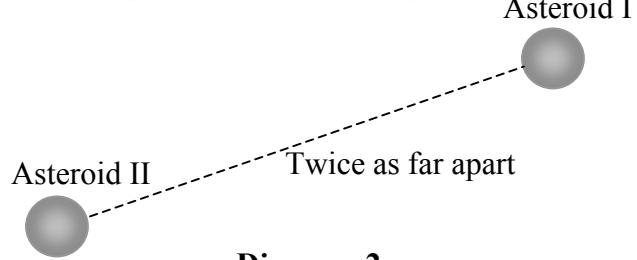
- (a)  $U_1 > 0$
- (b)  $U_1 = 0$
- (c)  $U_1 < 0$
- (d) Not enough information to determine



**Diagram 1**

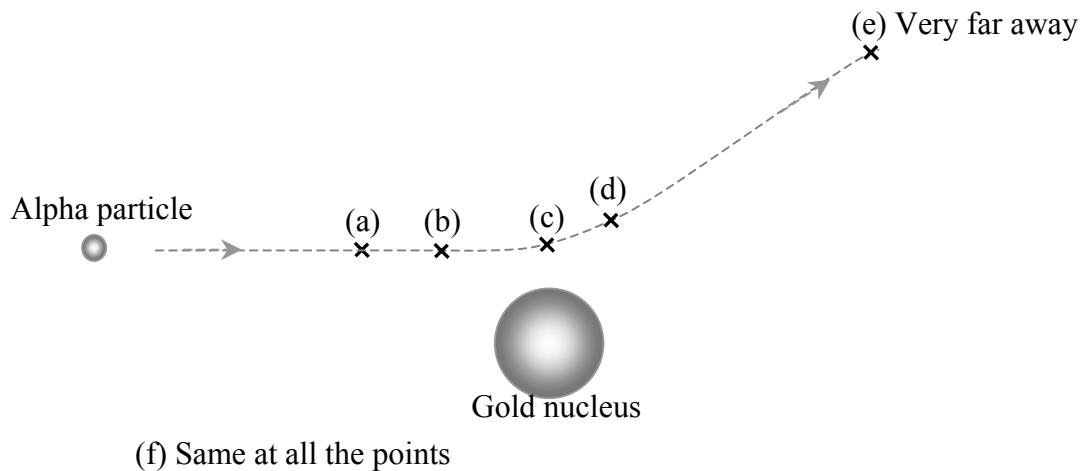
- 6b.** At a later time, the asteroids are twice as far apart, as shown in diagram 2. Consider the two asteroids as a system. Which of the following statements about the gravitational potential energy  $U$  and kinetic energy  $K$  is true, compared to the previous situation in question 6a?

- (a)  $U$  increased,  $K$  decreased
- (b)  $U$  increased,  $K$  increased
- (c)  $U$  increased,  $K$  remained constant
- (d)  $U$  decreased,  $K$  increased
- (e)  $U$  decreased,  $K$  decreased
- (f)  $U$  decreased,  $K$  remained constant
- (g) Both  $U$  and  $K$  remained constant



**Diagram 2**

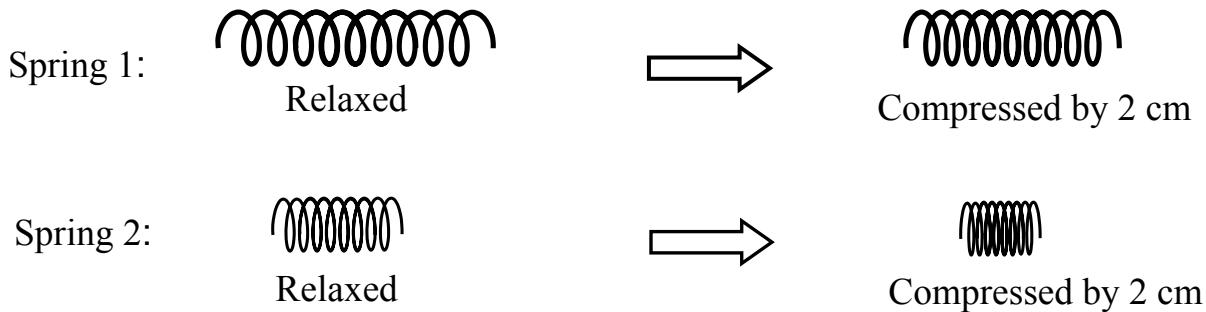
7. An alpha particle (charge  $+2e$ ) is deflected by a heavy gold nucleus (charge  $+79e$ ) that remains nearly stationary during the collision. The following diagram depicts the trajectory of the alpha particle. Consider a system that consists of both the alpha particle and the gold nucleus. At which of the following marked points does the system have the greatest electric potential energy?



(f) Same at all the points

8. Spring 1 and spring 2 are made of different materials but have the same stiffness. Spring 1 has longer relaxed length than spring 2. After both springs are compressed by 2 cm, the spring potential energy of spring 1 is changed by  $\Delta U_1$  and the spring potential energy of spring 2 is changed by  $\Delta U_2$ . Which of the following statements is true?

- (a)  $\Delta U_1 > \Delta U_2$
- (b)  $\Delta U_1 = \Delta U_2$
- (c)  $\Delta U_1 < \Delta U_2$
- (d) Not enough information to determine



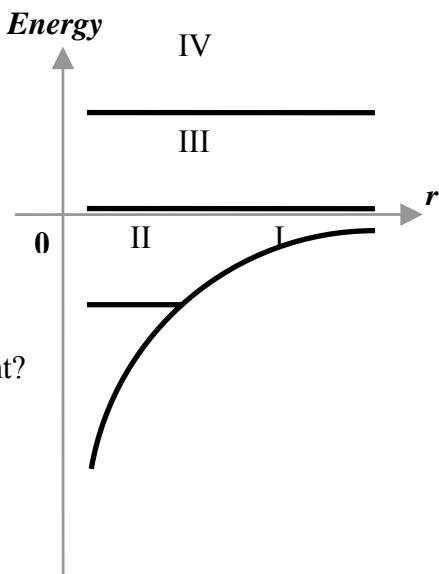
9. Two identical springs A and B have relaxed lengths of 5cm. Spring A is stretched by 2cm. Spring B is compressed by 2cm. Compare the spring potential energy of spring A and spring B. Which one has greater spring potential energy?

- (a) Spring A
- (b) Spring B
- (c) Both springs have the same amount of spring potential energy
- (d) Not enough information to determine

Consider a system that consists of two asteroids in deep space. The following figure plots energy of several different states of the two-asteroid system versus distance  $r$  between them. Questions 12 to 14 refer to this figure.

**10a.** What does the curved line **I** represent?

- (a) Kinetic energy
- (b) Potential energy
- (c) Sum of kinetic energy and potential energy
- (d) Rest energy
- (e) None of the above



**10b.** What do the horizontal lines (**II**, **III**, and **IV**) represent?

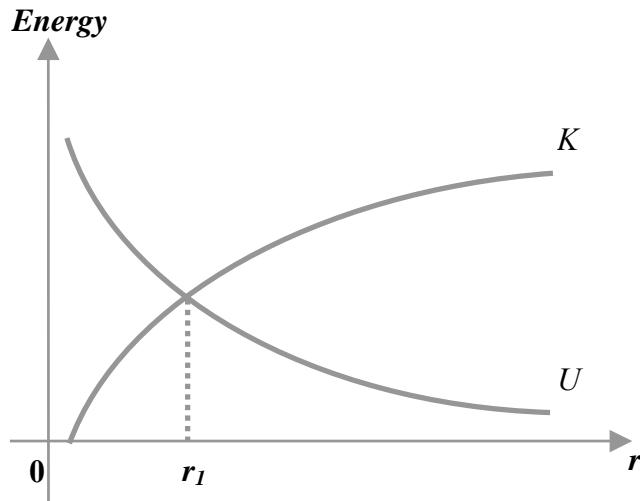
- (a) Kinetic energy
- (b) Potential energy
- (c) Sum of kinetic energy and potential energy
- (d) Rest energy
- (e) None of the above

**10c.** Which of the horizontal lines represents a bound state?

- (a) Line **II**
- (b) Line **III**
- (c) Line **IV**
- (d) All of the above
- (e) None of the above

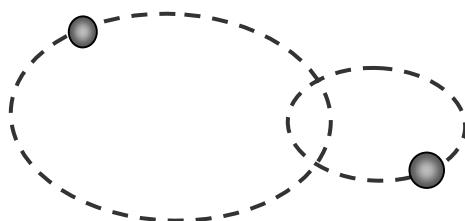
11. The following figure plots kinetic energy  $K$  and potential energy  $U$  of a two-particle system versus distance  $r$  between the two particles. From these curves we can conclude that the interaction between the two particles is:

- (a) Repulsion
- (b) Attraction
- (c) Repulsion when  $r < r_1$ , and attraction when  $r > r_1$
- (d) Attraction when  $r < r_1$ , and repulsion when  $r > r_1$
- (e) Not enough information to determine



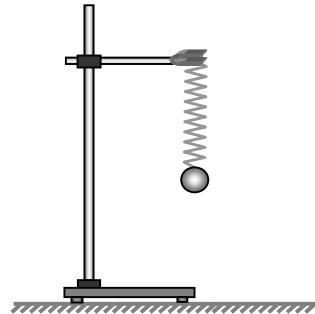
12. In deep space two stars are orbiting around each other. Consider the sum of kinetic energy and gravitational potential energy  $K+U$  of the two-star system. Which of the following statements is true?

- (a)  $K+U$  is positive
- (b)  $K+U$  is zero
- (c)  $K+U$  is negative
- (d)  $K+U$  is either positive or zero
- (e)  $K+U$  is either negative or zero



**13.** In a physics lab, a ball attached to a low-mass spring is oscillating up and down. A student chooses a system for analysis. At a certain time, the student finds the system's kinetic energy  $K$  is 0.3 J, the system's spring potential energy  $U_{\text{spring}}$  is 0.4 J, and the system's gravitational potential energy  $U_{\text{grav}}$  is -0.2 J. What was the student's choice of system?

- (a) Ball
- (b) Ball and spring
- (c) Ball and Earth
- (d) Spring and Earth
- (e) Ball, spring, and Earth

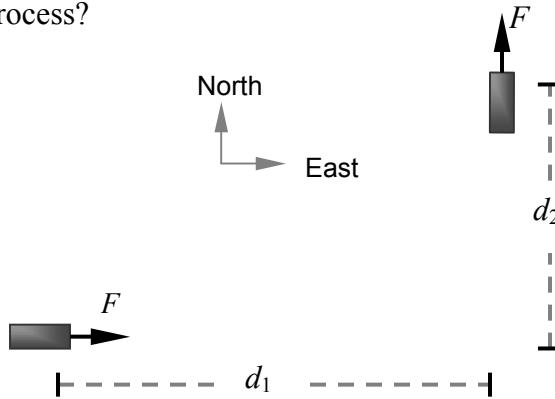


**14.** Consider a falling stone to be a system and answer the following question. At the instant when the stone is still falling down and has not yet hit the ground, which of the following forms of energy does the system of the falling stone have?

- (a) Rest energy only
- (b) Rest energy and kinetic energy
- (c) Rest energy and gravitational potential energy
- (d) Rest energy, kinetic energy and gravitational potential energy
- (e) None of the above

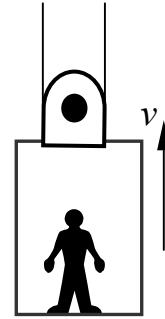
**15.** A truck is pulling a trailer due east with a constant force of magnitude  $F$  over distance  $d_1$ . Then the truck makes a sharp turn and pulls the trailer due north by applying the same magnitude of force  $F$  over a distance  $d_2$ . Considering the trailer as a system, what is the work done on the trailer by the truck during the process?

- (a)  $Fd_1 + Fd_2$
- (b)  $F\sqrt{d_1^2 + d_2^2}$
- (c)  $F|d_1 - d_2|$
- (d)  $F\sqrt{d_1 d_2}$
- (e)  $\langle Fd_1, Fd_2, 0 \rangle$



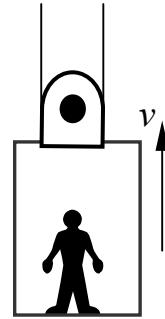
**16a.** A person is standing in an elevator, which is moving upward at a *constant speed*. Consider the person as a system. As the elevator goes between the 3<sup>rd</sup> and 4<sup>th</sup> floor,

- (a) The elevator does nonzero work on the person, and the earth does zero work on the person
- (b) The elevator does zero work on the person, and the Earth does nonzero work on the person
- (c) Both the elevator and the Earth do zero work on the person
- (d) Both the elevator and the Earth do nonzero work on the person
- (e) Not enough information to determine



**16b.** A person is standing in an elevator, which is moving upward at a *constant speed*. Consider the person as a system. As the elevator goes between the 3<sup>rd</sup> and 4<sup>th</sup> floor,

- (a) The total work done on the person by the elevator and the Earth is positive
- (b) The total work done on the person by the elevator and the Earth is zero
- (c) The total work done on the person by the elevator and the Earth is negative
- (d) Not enough information to determine



**17.** A basketball player jumps vertically off a rigid floor. While the player is in contact with the floor:

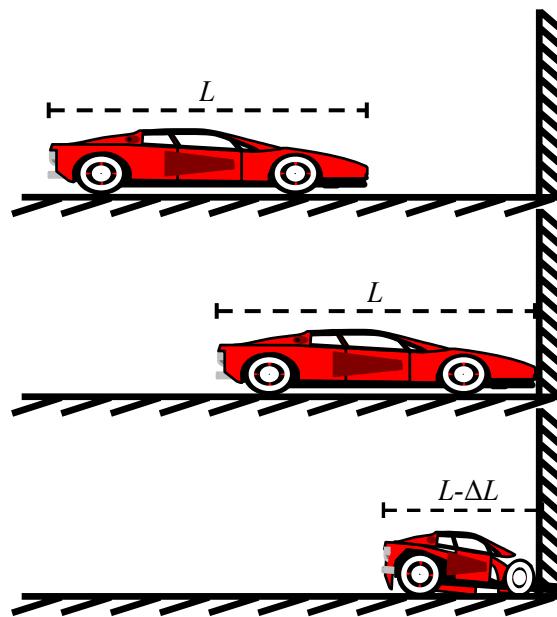
- (a) Neither the floor nor the player does work on the other
- (b) The floor does nonzero work on the player, but the player does zero work on the floor
- (c) The player does nonzero work on the floor, but the floor does zero work on the player
- (d) Both the floor and player do nonzero work on the other, but the floor does greater work on the player
- (e) Both the floor and the player do nonzero work on the other, but the floor does less work on the player
- (f) Both the floor and the player do the same nonzero amount of work on the other

**18.** You pull one end of a spring over a small distance  $d$  with a force of magnitude  $F$  to the left. At the same time, you pull the other end of the spring over a small distance  $d$  with a force of the same magnitude  $F$  to the right. Consider the spring as a real system. The total work done on the spring is:

- (a)  $-2Fd$
- (b)  $-Fd$
- (c) 0
- (d)  $Fd$
- (e)  $2Fd$
- (f) None of the above

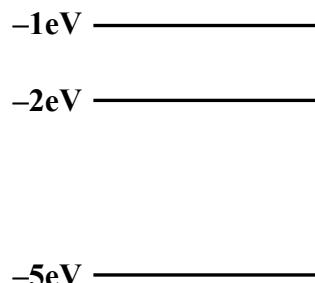
**19.** In a car crash test, a car of original length  $L$  hits a concrete wall and comes to a full stop without bouncing back. The wrecked car is now  $\Delta L$  shorter. During the collision, the average force applied to the car by the wall is  $F$ . Consider the car as a real system. What is the work done on the car by the wall?

- (a)  $F \cdot \Delta L$
- (b)  $F \cdot \Delta L / 2$
- (c)  $-F \cdot \Delta L$
- (d)  $-F \cdot \Delta L / 2$
- (e) 0



**20.** The following diagram shows the energy levels of some identical quantum objects that are known to have only three quantized energy levels:  $-5\text{eV}$ ,  $-2\text{eV}$ , and  $-1\text{eV}$ . When all the quantum objects are kept at a high temperature, what are the possible energies for the emitted photons?

- (a)  $1\text{eV}$ ,  $2\text{eV}$ , and  $5\text{eV}$
- (b)  $1\text{eV}$ ,  $3\text{eV}$ , and  $4\text{eV}$
- (c)  $1\text{eV}$ , and  $3\text{eV}$
- (d)  $1\text{eV}$ , and  $4\text{eV}$
- (e)  $3\text{eV}$ , and  $4\text{eV}$



**21.** There was a collection of some quantum objects, each of whose individual energy levels were  $-10\text{eV}$ ,  $-4\text{eV}$ , and  $-1\text{eV}$ . If all of these identical objects were in the ground state, what would be the energies of dark lines in an absorption spectrum?

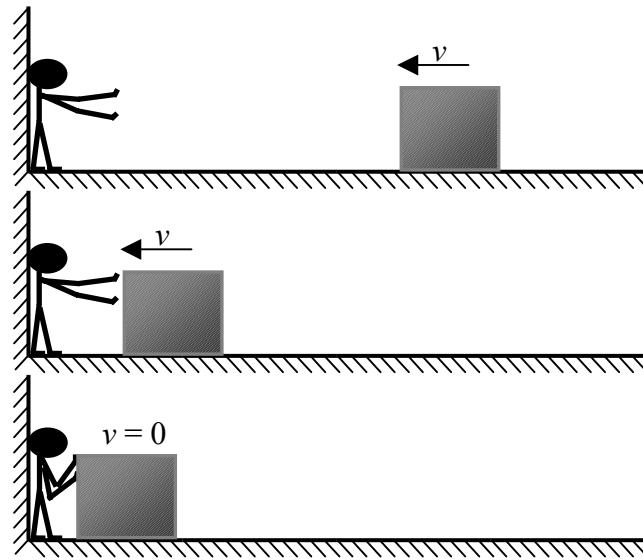
$-1\text{eV}$  \_\_\_\_\_

- (a)  $10\text{eV}$ ,  $4\text{eV}$ , and  $1\text{eV}$
- (b)  $9\text{eV}$ ,  $6\text{eV}$ , and  $3\text{eV}$
- (c)  $9\text{eV}$  and  $6\text{eV}$
- (d)  $6\text{eV}$  and  $3\text{eV}$
- (e) None of the above

$-4\text{eV}$  \_\_\_\_\_

$-10\text{eV}$  \_\_\_\_\_

**22.** A box of mass  $M$  is moving at speed  $v$  toward a person along a surface of negligible friction. A person leans against a wall with both arms stretched and applies a pushing force to stop the box. Finally the person brings the box to a full stop. There is no temperature change in the box at any time. We can conclude that during the process:



- (a) The amount of work done on the box by the person was  $+\frac{1}{2}Mv^2$
- (b) The amount of work done on the box by the person was positive, but there is not enough information to determine the value
- (c) The amount of work done on the box by the person was 0
- (d) The amount of work done on the box by the person was  $-\frac{1}{2}Mv^2$
- (e) The amount of work done on the box by the person was negative, but there is not enough information to determine the value

**23.** In outer space an asteroid was initially moving with a constant speed  $v$  in a certain direction. Some time later, the same asteroid was observed to be moving with the same constant speed  $v$  in the opposite direction. Between the two times, the total external work done on the asteroid was:

- (a) Positive
- (b) Negative
- (c) Zero
- (d) Not enough information to determine

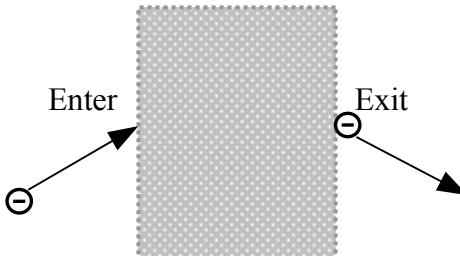
**24.** You warm up some water on a stove. The transfer of energy due to a temperature difference from the fire to the water is 300 J. At the same time, you stir the water with a beater, doing 200 J work on the water. After the stove is turned off, the beater is removed, and the water stops swirling, the thermal energy of the water is increased by:

- (a) 0
- (b) 100 J
- (c) 200 J
- (d) 300 J
- (e) 500 J

**25.** A clay ball moving to the right at a certain speed has kinetic energy of 10 J. An identical clay ball is moving to the left at the same speed. The two balls smash into each other and both come to a stop. What happened to the energy of the two clay-ball system?

- (a) The kinetic energy of the system did not change
- (b) The kinetic energy changed into thermal energy
- (c) The total energy of the system decreased by an amount 20 J
- (d) The initial kinetic energy of the system was zero, so there was no change in energy

- 26.** An electron with speed  $1 \times 10^8$  m/s enters a box along a direction depicted in the diagram. Some time later the electron is observed leaving the box with the same speed  $1 \times 10^8$  m/s but different direction as shown.

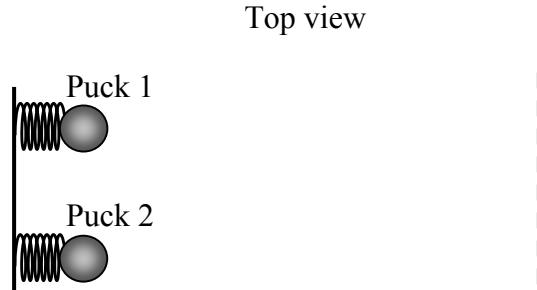


From this, we can conclude that in the box:

- (a) The net force on the electron was nonzero, and the net work done on the electron was nonzero
- (b) The net force on the electron was nonzero, but the net work done on the electron was zero
- (c) The net force on the electron was zero, but the net work done on the electron was nonzero
- (d) The net force on the electron was zero, and the net work done on the electron was zero
- (e) Not enough information to determine

- 27.** On a low-friction table you launch two pucks by pushing them against two identical springs through the same amount. The two pucks have the same shape and size, but the mass of puck 2 is twice the mass of puck 1. Then you release the pucks and the springs propel them toward the finish line. At the finish line, how does the kinetic energy of puck 1 compare to the kinetic energy of puck 2?

- (a) It is the same as the kinetic energy of puck 2
- (b) It is twice the kinetic energy of puck 2
- (c) It is half the kinetic energy of puck 2
- (d) It is four times the kinetic energy of puck 2
- (e) It is one-fourth the kinetic energy of puck 2



**28.** A point particle of mass  $m$  is initially at rest. A constant force is applied to the point particle during a first time interval, while the particle's speed increases from 0 to  $v$ . The constant force continues to act on the particle for a second time interval, while the particle's speed increases further from  $v$  to  $2v$ . Consider the work done on the particle.

During which time interval is more work done on the point particle?

- (a) The first time interval
- (b) The second time interval
- (c) The work done during the two time intervals is the same
- (d) Not enough information to determine

## **Appendix C: Pilot Version of the Energy Assessment**

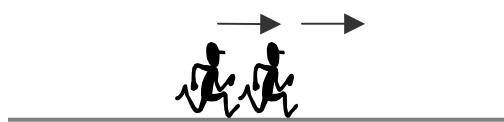
1. Runner A is running due east and runner B is running due west. They have the same mass  $m$  and speed  $v$ . Consider the two runners as a system; what is the kinetic energy of the system?

- (A) 0
- (B)  $\frac{1}{2}mv^2$
- (C)  $mv^2$
- (D)  $mc^2$
- (E)  $2mc^2$



2. Now both runners are running due east with speed  $v$ . What is the kinetic energy of the system?

- (A) 0
- (B)  $\frac{1}{2}mv^2$
- (C)  $mv^2$
- (D)  $mc^2$
- (E)  $2mc^2$



3. Two particles A and B, of masses  $m_A$  and  $m_B$ , have the same magnitude of momentum  $|\vec{p}_A| = |\vec{p}_B|$ . Particle A has greater kinetic energy than B ( $K_A > K_B$ ). Which of the following statements about the masses of particle A and particle B must be correct?

- (A)  $m_A > m_B$
- (B)  $m_A = m_B$
- (C)  $m_A < m_B$
- (D) Not enough information to determine

4. Three balls:

Ball 1 moves with a constant speed in a constant direction.

Ball 2 moves with a constant speed, but its direction changes.

Ball 3 moves in a constant direction, but its speed changes.

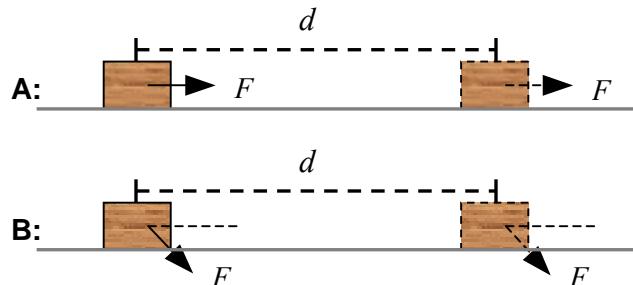
Which of these balls have a constant kinetic energy?

- (a) Ball 1 and ball 2
- (b) Ball 2 and ball 3
- (c) Ball 1 and ball 3
- (d) Ball 1, ball 2, and ball 3
- (e) None of the above

5. An electron of mass  $m_e$  has speed  $v_e$  close to the speed of light. What is the kinetic energy of this fast electron?

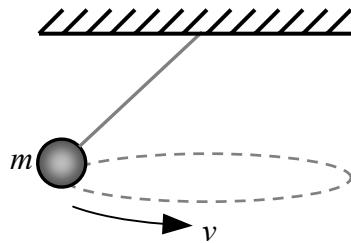
- (A)  $\frac{1}{2}m_e v_e^2$
- (B)  $\frac{1}{2}m_e v_e^2 + m_e c^2$
- (C)  $\frac{\frac{1}{2}m_e v_e^2}{\sqrt{1 - \left(\frac{v_e}{c}\right)^2}}$
- (D)  $\frac{m_e c^2}{\sqrt{1 - \left(\frac{v_e}{c}\right)^2}}$
- (E)  $\frac{m_e v_e^2}{\sqrt{1 - \left(\frac{v_e}{c}\right)^2}}$
- (F)  $\frac{m_e c^2}{\sqrt{1 - \left(\frac{v_e}{c}\right)^2}} - m_e c^2$
- (G)  $\frac{m_e c^2}{\sqrt{1 - \left(\frac{v_e}{c}\right)^2}} + \frac{1}{2}m_e v_e^2$

6. Two *identical* boxes A and B are on a surface. A worker applies a horizontal force to box A, and the box moves a distance  $d$ . Later he applies the *same magnitude* of force at an angle of  $45^\circ$  down to box B, which also moves through distance  $d$ .



- (A) The worker does no work on either box A or box B
- (B) The worker does no work on box A, but does nonzero work on box B
- (C) The worker does no work on box B, but does nonzero work on box A
- (D) The worker does nonzero work on both boxes, but does greater work on box A than on box B
- (E) The worker does nonzero work on both boxes, but does greater work on box B than on box A
- (F) The worker does the same nonzero amount of work on both boxes

7. A circular pendulum of mass  $m$  is moving with a constant speed  $v$  in a *horizontal* plane. Consider the pendulum bob as a system. Which of the following statements about the work done on the system is correct?



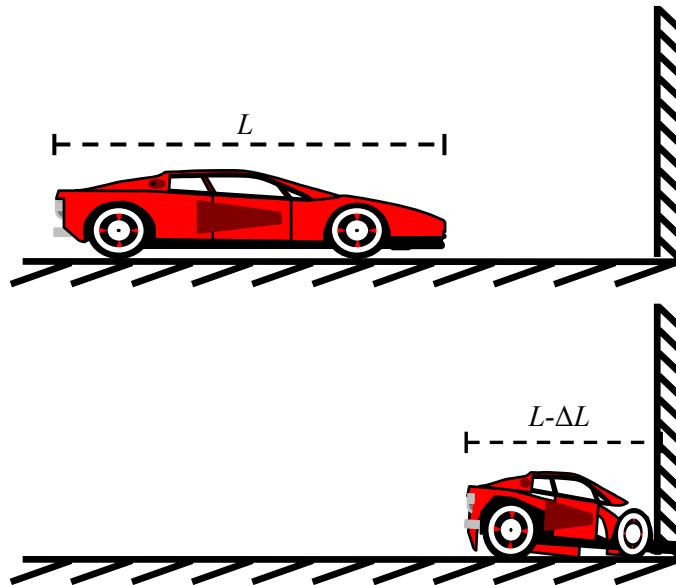
- (A) Neither the gravitational force nor the string does work on the bob
- (B) The gravitational force does nonzero work, but the string does no work on the bob
- (C) The gravitational force does no work, but the string does nonzero work on the bob
- (D) Both the gravitational force and the string do nonzero work, and the total work done by them is nonzero
- (E) Both the gravitational force and the string do nonzero work, and the total work done by them is zero

8. A basketball player jumps vertically off a rigid floor. While the player is in contact with the floor:

- (A) Neither the floor nor the player does work on the other
- (B) The floor does work on the player, but the player does no work on the floor
- (C) The player does work on the floor, but the floor doesn't do work on the player
- (D) Both the floor and player do work on the other, but the floor does greater work on the player
- (E) Both the floor and the player do work on the other, but the floor does less work on the player
- (F) Both the floor and the player do the same nonzero amount of work on the other

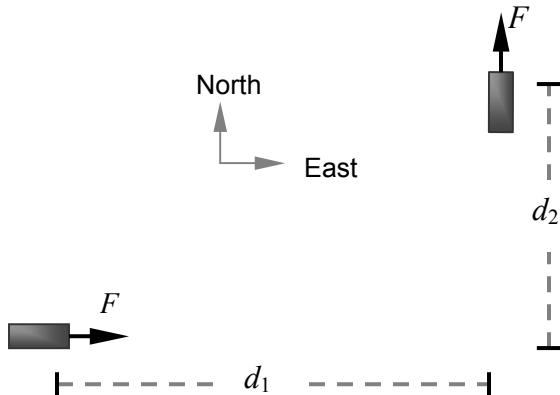
9. In a car crash test, a car of original length  $L$  hits a concrete wall and comes to a full stop without bouncing back. The wrecked car is now  $\Delta L$  shorter. During the collision, the average force applied to the car by the wall is  $F$ . What is the work done on the car by the wall?

- (A)  $F \cdot \Delta L$
- (B)  $F \cdot \Delta L / 2$
- (C)  $-F \cdot \Delta L$
- (D)  $-F \cdot \Delta L / 2$
- (E) Zero work



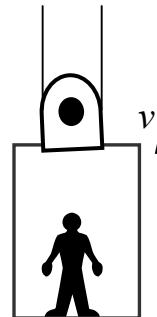
10. A truck is pulling a trailer due east with a constant force (magnitude =  $F$ ) over distance  $d_1$ . Then the truck makes a sharp turn toward north and pulls the trailer due north by applying the same magnitude of force  $F$  over a distance  $d_2$ . Considering the trailer as a system, what is the work done on the trailer by the truck during the process?

- (A)  $Fd_1 + Fd_2$
- (B)  $F\sqrt{d_1^2 + d_2^2}$
- (C)  $F|d_1 - d_2|$
- (D)  $F\sqrt{d_1 \cdot d_2}$
- (E)  $\langle Fd_1, Fd_2, 0 \rangle$



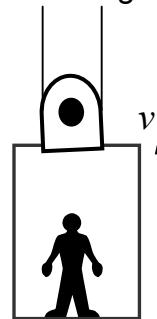
11. A person is standing in an elevator, which is moving upward at a *constant speed*. Consider the person as a system. As the elevator goes between the 3<sup>rd</sup> and 4<sup>th</sup> floor,

- (A) Only the elevator does work on the person
- (B) Only the Earth does work on the person
- (C) Both the elevator and the Earth do work on the person
- (D) Neither the elevator nor the Earth does work on the person
- (E) Not enough information to determine



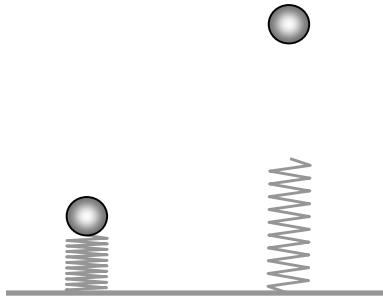
12. A person is standing in an elevator, which is moving upward at a *constant speed*. Consider the person as a system. As the elevator goes between the 3<sup>rd</sup> and 4<sup>th</sup> floor,

- (A) The total work done on the person by the Earth and the elevator is zero
- (B) The total work done on the person by the Earth and the elevator is positive
- (C) The total work done on the person by the Earth and the elevator is negative
- (D) Not enough information to determine



13. A ball is launched vertically up from a low-mass compressed spring. When the ball reaches its maximum height, you claim the system's kinetic energy is 0 ( $K = 0$ ), the spring potential energy is 0 ( $U_{spring} = 0$ ), and the gravitational potential energy is nonzero ( $U_{grav} \neq 0$ ). What was your choice of system?

- (A) Ball
- (B) Ball and spring
- (C) Ball and Earth
- (D) Spring and Earth
- (E) Ball, spring, and Earth



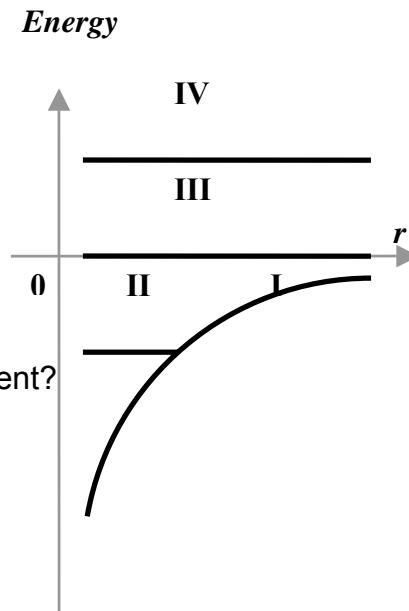
14. Consider a falling stone as a system. At the instant when the stone is several meters above the ground, the stone has:

- (A) Kinetic energy only
- (B) Rest energy only
- (C) Kinetic energy and gravitational potential energy
- (D) Kinetic energy and rest energy
- (E) Kinetic energy, rest energy, and gravitational potential energy

The following figure plots energy graphs of several different states of a two-body system versus distance  $r$  between the two bodies. Questions 15 to 17 refer to this figure.

15. What does the curved line I represent?

- (A) Kinetic energy
- (B) Potential energy
- (C) Sum of kinetic energy and potential energy
- (D) Rest energy
- (E) None of the above



16. What do the horizontal lines (II, III, and IV) represent?

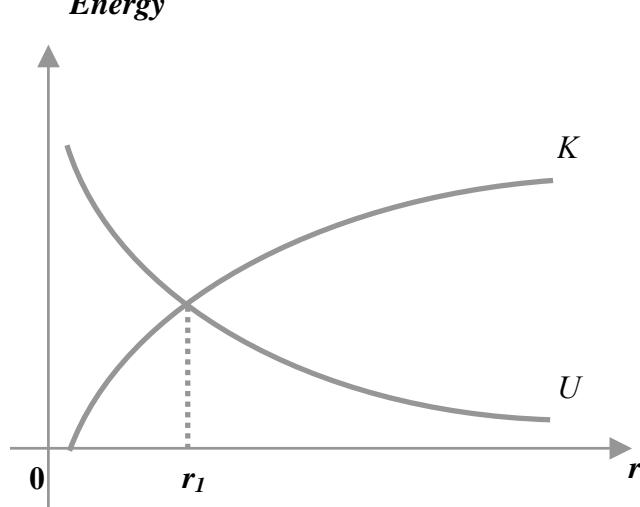
- (A) Kinetic energy
- (B) Potential energy
- (C) Sum of kinetic energy and potential energy
- (D) Rest energy
- (E) None of the above

17. Which of the horizontal lines implies a bound state?

- (A) Line II
- (B) Line III
- (C) Line IV
- (D) All of the above
- (E) None of the above

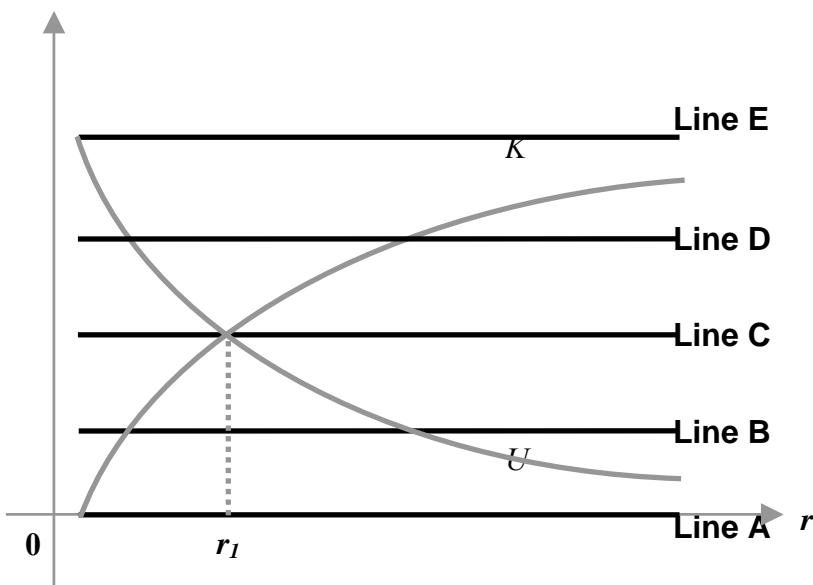
18. In the following figure are plotted kinetic energy  $K$  and potential energy  $U$  of a two-particle system versus distance  $r$  between the two particles. From the potential energy curve we can conclude that the interaction between the two particles is:

- (A) Repulsion always
- (B) Attraction always
- (C) Repulsion when  $r < r_1$ , and attraction when  $r > r_1$
- (D) Attraction when  $r < r_1$ , and repulsion when  $r > r_1$
- (E) Not enough information to determine

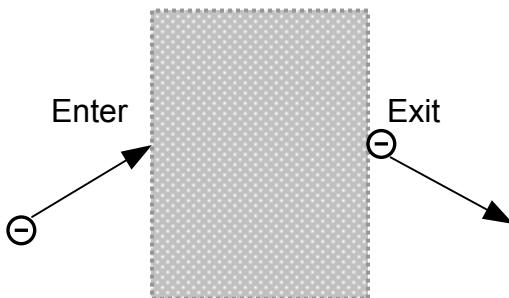


19. In the following figure are plotted kinetic energy  $K$  and potential energy  $U$  of a two-particle system versus distance  $r$ . Which line best represents the sum of kinetic energy and potential energy? **Energy**

- (A) Line A
- (B) Line B
- (C) Line C
- (D) Line D
- (E) Line E



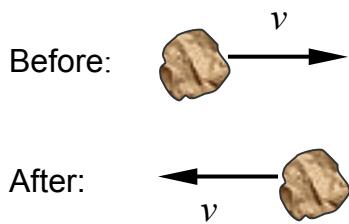
20. An electron with speed  $10^8$  m/s enters a box along a direction depicted in the diagram. Some time later the electron is observed leaving the box with the same speed  $10^8$  m/s but different direction as shown.



From this, we can conclude that in the box the net force on the electron was:

- (A) nonzero, and the net work done on the electron was nonzero
- (B) nonzero, but the net work done on the electron was zero
- (C) zero, but the net work done on the electron was nonzero
- (D) zero, and the net work done on the electron was zero
- (E) Not enough information to determine

21. In outer space an asteroid is moving forward with constant speed  $v$ . Some time later, the same asteroid is observed to be moving backward with the same constant speed  $v$ .

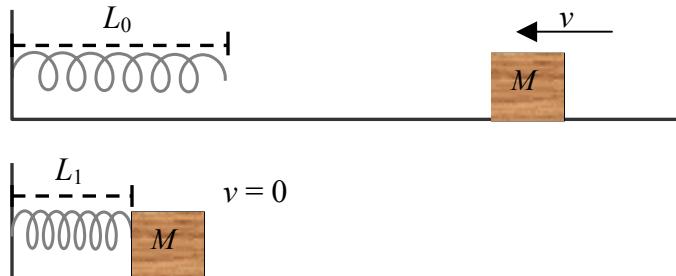


Between the two times:

- (A) No net force acted on the asteroid
- (B) An external impulse was applied to the asteroid
- (C) Nonzero external work was done on the asteroid
- (D) Kinetic energy of the asteroid changed
- (E) Momentum vector of the asteroid did not change

**Questions 22 to 24 are based on the following statement.**

A light spring of stiffness  $k_s$  and relaxed length  $L_0$  has one end attached to a wall. A block of mass  $M$ , moving with speed  $v$  on a surface with negligible friction, approaches and compresses the spring, until it comes to a full stop. The spring is compressed to a length  $L_1$ . During the process there is no temperature change.



22. If we take only the block as system, which of the following expressions is equal to the external work done on the system?

- (A)  $-\frac{1}{2}k_s L_0^2$    (B)  $-\frac{1}{2}k_s L_1^2$    (C)  $-\frac{1}{2}k_s(L_0^2 - L_1^2)$    (D)  $-\frac{1}{2}Mv^2$    (E)  $-M\left(\frac{v}{2}\right)^2$

23. If we take both the spring and block as our system, which statements are correct?

- (1) The system has constant momentum during the process.  
(2) The system has constant total energy during the process.  
(3) The wall does nonzero work on the system during the process.  
(4) The wall applies nonzero impulse to the system during the process

- (A) (1) and (2)
- (B) (1) and (3)
- (C) (2) and (4)
- (D) (3) and (4)
- (E) (1), (2), (3), and (4)

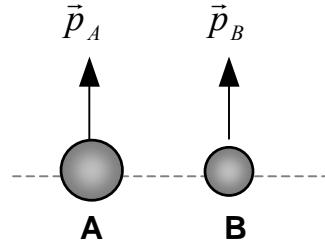


24. If we take the spring and block as the system, which of the following is correct?

- (A)  $\frac{1}{2}k_s(L_0 - L_1)^2 = \frac{1}{2}Mv^2$
- (B)  $\frac{1}{2}k_s(L_0 - L_1)^2 = -\frac{1}{2}Mv^2$
- (C)  $\frac{1}{2}k_sL_0^2 = \frac{1}{2}Mv^2 + \frac{1}{2}k_sL_1^2$
- (D)  $Mv = k_s(L_0 - L_1)$
- (E)  $Mv = k_sL_1$

**Questions 25 and 26 are based on the following statement.**

Two balls A and B are launched vertically up from the same height at the same time. The initial upward momentum of these two balls is the same ( $\vec{p}_A = \vec{p}_B$ ), and ball A is heavier than ball B ( $m_A > m_B$ ). Consider the maximum height  $h_A$  for ball A and  $h_B$  for ball B and ignore air resistance.



25. Which statement about the maximum heights the balls reach is correct?

- (A)  $h_A > h_B$
- (B)  $h_A = h_B$
- (C)  $h_A < h_B$
- (D) Not enough information to determine

26. It takes  $t_1$  seconds for ball A to reach its maximum height after the launch, and  $t_2$  seconds for ball B to reach its maximum height. Which statement about the times is correct?

- (A)  $t_1 > t_2$
- (B)  $t_1 < t_2$
- (C)  $t_1 = t_2$
- (D) Not enough information to determine

**Questions 27 and 28 are based on the following statement.**

A *constant* net force is applied on a particle which is initially at rest.  $\Delta t_1$  seconds later, this particle's speed has increased from 0 to  $v$ . Another  $\Delta t_2$  seconds later, its speed has increased from  $v$  to  $2v$ . The net work done on the particle due to the net force is  $W_1$  during the first time interval  $\Delta t_1$ , and  $W_2$  during the second time interval  $\Delta t_2$ .

27. Which of the following statement about  $W_1$  and  $W_2$  is correct?

- (A)  $W_1 > W_2$
- (B)  $W_1 = W_2$
- (C)  $W_1 < W_2$
- (D) Not enough information to determine

28. Which of the following statement about  $\Delta t_1$  and  $\Delta t_2$  is correct?

- (A)  $\Delta t_1 > \Delta t_2$
- (B)  $\Delta t_1 = \Delta t_2$
- (C)  $\Delta t_1 < \Delta t_2$
- (D) Not enough information to determine