The Effects of Emphasizing Intentional Problem Solving in a Modeling Physics Classroom

Robert Culbertson, Principal Investigator

Brandy Heflin, Ryan Lodes, Scott Pflaumer, and Keith Quigley, Co-Investigators

Arizona State University

Author's Note

Robert Culbertson, Department of Physics, Arizona State University; Brandy Heflin, Department of Physics, Arizona State University; Ryan Lodes, Department of Physics, Arizona State University, Scott Pflaumer, Department of Physics, Arizona State University; Keith Quigley, Department of Physics, Arizona State University

This study completed as Action Research required for the Master of Natural Science degree with concentration in physics.

Correspondence concerning this study should be directed to Robert Culbertson, Department of Physics, Arizona State University.

July 2012

Table of Contents

Abstract 3

Rationale 4

Literature Review 4

Method 12

Procedure for Treatment 16

Data Analysis 18

Timeline 23

Results 23

Conclusion 61

Implications for Further Research 61

References 63

Appendices 67

Abstract

Students begin their education in physics as novice problems solvers. Instead of carefully defining a problem, using qualitative models, and planning a method of solution, students often immediately attempt to find the answer to the problem. The result of this lack of methodical approach is that students are not only unable to solve problems, they are unsure of even the basic steps that lead toward solutions. Previous research has shown that intentionally teaching expert-like strategies increases students' problem solving ability. Other studies have found that Modeling Instruction improves students' expert-like problem solving ability. This study was initiated to evaluate the impact on students' problem solving skills through teaching explicit problem solving strategies in addition to Modeling Instruction. There was no conclusive evidence that the gains from the two methods were additive; however, this approach was reported to be beneficial by study participants. There was substantial evidence that without a solid conceptual understanding, expert-like problem solving ability was limited.

Rationale

Problem solving is an essential skill in order to be successful in the study of physics. Unfortunately, many students are not good at solving problems. All physics teachers instruct their students on how to get answers to physics problems, but the traditional manner of teaching rarely helps most students to develop into effective problem solvers. Research has shown that teaching an explicit, expert-like strategy can lead students toward a more expert way of solving problems. Modeling Instruction has also been shown to produce students who are more expert problem solvers, as a result of teaching conceptual understanding and concept organization.

This study was initiated to evaluate the impact on students' problem solving skills through teaching explicit problem solving strategies within classrooms employing Modeling Instruction. The premise of the study was that, in addition to learning through Modeling Instruction, students would be taught an explicit, expert strategy, and that the students would become more expert-like in their problem solving ability than by using either method by itself. It was theorized that the results from utilizing both methods would be additive, thus providing teachers with another tool to enhance student learning. It would also point to further study into which aspect of cognition each method affects. The outcomes might also inform future research looking at whether the two methods of instruction are redundant or complementary.

Literature Review

Problem solving is one of the most important skills students need in order to be successful in the study of physics, wherein success is defined as more than getting an answer and includes conceptual understanding (Touger, 1964). Unfortunately, a multitude of studies indicate that students solve problems ineffectively (Larkin & Reif, 1979; Chi, Feltovich, & Glaser, 1981; Reif & Heller, 1982; Mestre & Touger, 1989). The traditional manner of teaching, which uses numerous example problems, rarely helps students to effectively problem solve because the teacher typically completes a great deal of the problem solving unconsciously with the majority of instruction focused on the algebra involved in reaching the final answer (Larkin & Reif, 1979). Many students are unaware of or uncomfortable with the critical steps in an effective problem solving strategy which include: carefully defining the problem, deploying hierarchical knowledge, thinking through conceptual aspects of the problem, and checking the final solution (Reif & Heller, 1982; Polya, 1945). Thus, students struggle to take even small steps forward when approaching a new or unfamiliar problem. Without explicit instruction on how to become better problem solvers, students improve very little in this area despite many years of practice (Malone, 2006).

Students encounter several common difficulties that make solving a physics problem quite challenging. They may fail to describe problems adequately (Lederman, 2009), or try to assemble solutions by stringing together miscellaneous mathematical formulas from their repertoire (Reif & Heller, 1982). Yet, the initial description of that problem is crucial to determine the ease with which the problem is solved, and that description is the first task faced by a problem solver (Reif & Heller, 1982). Students often fail to adequately describe a problem qualitatively, concentrating on only the surface structure of the problem (Chi et al., 1981). When attempting to solve problems, most students focus on the numeric terms, as opposed to the problem as a whole, which leads to an inability to detect errors (Malone, 2006). Students entering a physics classroom have a poor foundation of physics knowledge. They need help to integrate their knowledge into a comprehensive structure built around the basic concepts of physics and to develop the techniques for accessing that knowledge when needed (Van Huevelen, 1991). Their knowledge base is often loosely connected and non-hierarchical, which discourages logical reasoning and analysis (Knight, 2004). These factors leave students inadequately prepared to solve problems in a physics classroom.

Much research has been conducted comparing how experts and novices approach problem solving (Chi et al, 1981; Larkin & Reif, 1979; Reif & Heller 1982; Priest & Lindsay, 1992; Reif & Allen, 1992; Walsh, Howard, & Bowe, 2007). Beyond an expert's deeper knowledge and faster pace, experts and novices approach physics problems with fundamentally different strategies (Larkin & Reif, 1979). Novices tend to rush immediately to use equations and work towards a solution without checking their final answer (Reif & Heller, 1982). Experts usually first consider the general type of problem, carefully describe the problem, use qualitative methods to make predictions, and explore outcomes before using quantitative methods to solve mathematically (Reif & Heller, 1982). Additionally, experts tend to deploy knowledge organized in a hierarchical manner and carefully check their solution (Reif & Heller, 1982). Experts gather and store information in clusters or chunks (Larkin & Reif, 1979). The fundamental concepts occupy the highest, most accessible, levels of the hierarchy (Mestre & Touger, 1989). So, the first and most important step is to identify key elements and categorize the problem.

One study (Walsh et al., 2007) described five categories of problem solving approaches used by introductory students, namely: Scientific Approach, Plug-and-Chug Structured, Plug-and-Chug Unstructured, Memory-Based, and No Clear Approach. Key elements in separating students into one of the five approaches included qualitative analysis, planning of solution, systematic manner, and conceptual analysis. Students who employed the Scientific Approach exhibited expert-like thinking as described above, while students who employed the other approaches exhibited varying degrees of novice-like thinking as previously described. The extent to which students used concepts to guide their application of equations separated the four novice-like levels, with Plug-and-Chug Structured students focusing on the relationships between concepts while identifying the variable to be solved and Plug-and-Chug Unstructured students focusing on the variable to be solved while referring to concepts as variables (for example, talking about “t” rather than “time”). Memory-Based students focused on the variable to be solved while also trying to match the problem to previously encountered situations. No Clear Approach students focused on the variables given without relating them to the concepts represented and used the variables haphazardly. Observing student discourse and evaluating student work through this lens, then, gives qualitative evidence suggesting a student’s level of problem solving ability.

The strategies experts and novices use to categorize problems have been studied (Larkin, McDermott, Simon, & Simon, 1980; Chi et al., 1981; Chi, Glaser, & Rees, 1982; Reif & Heller, 1982). In particular, Chi et al. (1981) performed two sorting procedure studies asking participants to organize problems into groups based on similarities of solutions. The problems were typed on 3” × 5” index cards, hence these sorting procedures were referred to as Card Sorting Task (CST). The first study found that experts described the problems based on the laws of physics whereas novices described them based on surface characteristics, and each group sorted the problems accordingly. The second study confirmed those results with the creation of problems with similar surface features, such as objects contained within the problems, but different physical principles such as Newton’s Laws or conservation of mechanical energy. Novices grouped the cards by surface features. Malone (2006) replicated that research and included a quantitative score for the CST. Problems were assigned a number such as 1.3, 2.1, or 2.3, where the first number indicated similarities in deep structure and the second number indicated similarities in surface structure. That made it possible to measure where the student fell on a spectrum from novice to expert in categorizing problems. Both Chi et al. (1981) and Malone (2006) encountered participants of intermediate ability who used a mix of sorting strategies. Consequently, Malone's CST could be utilized to measure the progress of students from novice-like problem solving toward expert-like problem solving during a course of instruction.

One of the primary suggestions of Reif and Heller's (1982) research was for physics teachers to explicitly teach expert-like problem solving strategies. A study on introductory college physics split instruction into two parts: general lecture and problem sessions (Mestre, Dufresne, Gerace, Hardiman, & Touger, 1993). All students took the general lecture, but half of the students attended a traditional (example-based) problem session while the other half of students were taught expert-like problem solving methods during their second session (Mestre et al., 1993). The group that was intentionally taught problem solving strongly outperformed the example-based group (Mestre et al., 1993). Another study (Wright & Williams, 1986) also showed significant gains in student performance by teaching an explicit method of problem solving. It seems reasonable to surmise that one could change a student's problem solving methods with explicit instruction and practice.

One specific method frequently mentioned in research on problem solving is a five-step approach (Heller, Keith, & Anderson, 1992). First, students learn to visualize and translate the problem into language they understand. Second, students are taught to create a qualitative physics description that uses concepts and principles to describe the problem and make qualitative predictions. Third, students are trained to create a solution plan that translates qualitative information into mathematical representations. Fourth, students study how to execute their plan. Finally, students are instructed how to evaluate their solution by checking for reasonableness and consistency.

Another way to improve expert-like problem solving is Modeling Instruction. Hestenes (1987) stated that problem solving in physics is primarily a modeling process, as the complete solution to every physics problem is actually a model that supplies the context to make the numeric answer meaningful (Wells, Hestenes, & Swackhamer, 1995). A model is a way to describe the relationships between objects in a physical system and the processes they undergo. Modeling is the process of developing that description. Hestenes (1987) asserted that a problem is not fully understood until a model of it has been constructed. Wells et al. (1995) explained that Modeling Instruction is a pedagogy based on the modeling cycle, which consists of two phases, model development and model deployment. In the model development phase, students design and conduct a laboratory experiment to determine the qualitative and quantitative relationships between measurable variables in the system under study. In the model deployment phase, students apply the model to various situations in various ways, including pencil-and-paper problems and lab practicums. In Modeling Instruction, physics content is reorganized around a small number of physical models as “units of structured knowledge” (Jackson, Dukerich, & Hestenes, 2008). Each phase lasts at least one week, with a complete cycle concluding in two to three weeks (Wells et al., 1995). In short, the instructor emphasizes student construction of appropriate models for the physical systems they study, stressing development of sound conceptual understanding of problems through graphical and diagrammatic representations prior to developing algebraic representations (Jackson et al., 2008). In fact, in Modeling Instruction problem solving is secondary to Modeling because students are required to justify their conclusions using an explicit model in both problem solving and laboratory activities (Wells et al., 1995).

Hestenes, Wells, and Swackhamer (1992) suggested that students who have been taught using Modeling Instruction are, in fact, better problem solvers than students taught using traditional methods. Wells et al. (1995) assert that this difference is due to the "thorough grounding in basic concepts first," without which students must rely on "rote learning and plug-and-chug problem solving." Hestenes and Wells (1992) correlated conceptual understanding as measured by the Force Concept Inventory (FCI) with problem solving skills as measured by the Mechanics Baseline Test (MBT). The FCI (Hestenes et al., 1992) is a conceptual multiple-choice assessment whose distracters are "commonsense alternatives" to Newtonian explanations for a variety of scenarios involving forces. It measures to what degree students correct their misconceptions and adopt a Newtonian way of thinking when administered as a pre and posttest. The MBT (Hestenes & Wells, 1992) is a problem solving assessment whose distracters are common student mistakes made from deficient conceptual understanding of Newtonian mechanics. Hestenes and Wells (1992) identified two thresholds of understanding in their correlation study: Students who scored below a 60% on the FCI were unlikely to score above a 60% on the MBT, and students who scored above an 80% on the MBT scored at least an 80% on the FCI. They proposed that the minimum level of conceptual understanding necessary for effective problem solving in basic Newtonian mechanics corresponds to a 60% on the FCI, and that a mastery level of basic Newtonian mechanics corresponds to an 80% on the FCI. So, student FCI and MBT scores can be analyzed as quantitative evidence of problem solving ability.

One of the most important differences between experts and novices is how their understanding of physics is organized (Knight, 2004). Where an expert has an organized hierarchy, novices typically have "an unstructured knowledge of loosely connected facts and formulas" (Knight, 2004). Since one of the main objectives of Modeling is to group facts and formulas into overarching models similar to the hierarchy that experts use, it would be expected that Modeling students tend to be stronger problem solvers (Hestenes, 1987). It has been shown that Modeling students outperform traditional students with physics problem solving (Malone, 2006). However, Malone (2006) stated that the Modeling Instruction pedagogy needs to stress and model to a greater extent a number of expert-like traits such as the use of multiple representations, the identifying of a model or principle to help analyze problem statements, and the need to work forward towards a solution.

A common thread between much of the research on how students solve problems is that students are overly focused on finding the answer instead of understanding the whole situation (Chi et al., 1981). Indeed, Charles Kettering articulates "a problem well stated is a problem half solved" (Reif & Heller, 1982). Context-rich problems, that cannot be solved utilizing novice approaches, may be used to encourage students to transition to expert-like methods (Heller et al., 1992).

As outlined above, beginning students often exhibit novice strategies in problem solving. Explicit instruction in problem solving strategies enhances student problem solving ability. Modeling Instruction also improves student use of more expert-like strategies when problem solving. There exists a dearth of evidence examining the combined use of Modeling Instruction and explicit instruction in problem solving. This study examined the concurrent use of explicit problem solving methods and Modeling Instruction in order to assess the possibility of additional improvement in students' ability to solve problems beyond that experienced using exclusively one method.

Method

Subjects and Investigators in Experimental Treatment Group

Table 1 provides condensed demographic information for the subjects and comparison subjects who participated in the study.

Investigator 1. Investigator 1 taught at a medium, suburban, non-religious private school based in a large Midwestern city. The school had an enrollment of approximately 400 students, of whom 78% were Caucasian, 8% were African American, 3% were Asian, 3% were Native American, 3% were Middle Eastern, 3% were Multiracial, 1% were Hispanic and 1% were international students. Approximately 10% of the students received significant financial aid. Investigator 1 worked with approximately 30 freshmen honors physics students. The freshmen honors physics class was equivalent to a junior honors-level class designed to prepare students for AP physics. All students in treatment group 1 were taught using Modeling Instruction.

Investigator 2. Investigator 2 taught at a medium, working-class, public suburban high school in a Western metropolis. The school had an enrollment of approximately 1,500 students in grades 10 through 12. The population was approximately 78% Caucasian, 16% Hispanic, 3% Asian, 2% African American, and 1% Native American and Pacific Islander. Thirty-seven percent of the students received free or reduced lunch. Investigator 2 taught approximately 50 general physics students and 50 basic physics students. All students in treatment group 2 were taught using Modeling Instruction.

Investigator 3. Investigator 3 taught at a small, coeducational independent college preparatory day school in a Midwestern metropolis. The school served 950 students in grades prekindergarten through 12; 426 of them were enrolled at the high school. The population of the

Table 1

Study Subjects and Comparison Subjects Demographic Data

Group	School			Ethnic Composition					Students Receiving Financial Aid
	Enrollment	Type	Urban Setting	Caucasian	African American	Asian	Hispanic	Other	
Investigator 1 (I1)	400	Private	Suburban	78%	8%	3%	1%	10%	10%
Investigator 2 (I2)	1,500	Public	Suburban	78%	2%	3%	16%	1%	37%
Investigator 3 (I3)	426	Private	Suburban	75%	9%	7%	1%	8%	35%
Investigator 4 (I4)	2,500	Public	Suburban	90%	1%	1%	5%	3%	7%
Comparison group 1 (C1)	1,900	Public	Suburban	43%	3%	4%	47%	3%	43%
Comparison group 2 (C2)	1,600	Public	Small town	97%	1%	1%	1%	0%	27%
Comparison group 3 (C3)	426	Private	Suburban	75%	9%	7%	1%	8%	35%
Comparison group 4 (C4)	400	Private	Suburban	78%	8%	3%	1%	10%	10%
Comparison group 5 (C5)	1,000	Public	Rural	5%	0%	0%	94%	0%	85%

high school was 74.6% Caucasian, 8.5% African American, 7% Asian American, 4.5% multiracial, 1.4% international, 1.2% Hispanic, and 0.7% Pacific Islander. Thirty-five percent of the students received need-based financial aid. Investigator 3 worked with approximately 45 regular physics students. Physics was required for graduation and was usually taken in the junior year. All students in treatment group 3 were taught using Modeling Instruction.

Investigator 4. Investigator 4 taught at a large public suburban high school in a major Midwestern city. The high school had an enrollment of approximately 2,500 students in grades 9 through 12. The population of the school was approximately 90% Caucasian, 5% Hispanic, 2% multiracial, and less than 1% African American, Native American and Asian. Seven percent of the students received free or reduced lunch. Investigator 4 worked with approximately 135 general physics students. General physics was populated by about 95% juniors and 5% seniors. All students in treatment group 4 were taught using Modeling Instruction.

Subjects in Comparison Groups

Comparison Group 1. Comparison group 1 was in a large suburban high school in a major Midwestern city. The school had an enrollment of approximately 1,900 students of whom about 47% were Hispanic, 43% Caucasian, 4% Asian, 3% African American, and 2% were multiracial. Approximately 43% of students came from low-income households. Comparison group 1 contained about 80 honors physics students. All students in comparison group 1 were taught using Modeling Instruction.

Comparison Group 2. Comparison group 2 was comprised of students in grades 10 through 12 at a medium high school in a small Midwestern town. The high school had an enrollment of about 1,600 students in grades 9 through12. The population of the school was approximately 97% Caucasian, and 3% Hispanic, Asian and African American. Twenty-seven percent of students were eligible for free or reduced lunch. Comparison group 2 contained 25 physics students. All students in comparison group 2 were taught using Modeling Instruction.

Comparison Group 3. Comparison group 3 was at a small, coeducational independent college preparatory day school in Midwestern metropolis. The school served 950 students in grades prekindergarten through 12; 426 of them were enrolled at the upper school. The population of the high school was 74.6% Caucasian, 8.5% African American, 7% Asian American, 4.5% multiracial, 1.4% international, 1.2% Hispanic, and 0.7% Pacific Islander. Thirty-five percent of the students received need-based financial aid. Comparison group 3 was 10 juniors who took regular physics, taught using traditional methods.

Comparison Group 4. Comparison group 4 was comprised of 90 freshmen physics students at a medium, suburban, non-religious private school based in a large Midwestern city. The school had an enrollment of approximately 400 students, of whom 78% were Caucasian, 8% were African American, 3% were Asian, 3% were Native American, 3% were Middle Eastern, 3% were multiracial, 1% were Hispanic and 1% were international students. Approximately 10% of the students received significant financial aid. Comparison group 4 received traditional instruction covering mechanics, except for circular motion and momentum.

Comparison Group 5. Comparison group 5 was comprised of 9 honors physics students at a medium, rural, public school based in the Southwest. The school had an enrollment of approximately 1,000 students, of whom 5% were Caucasian, 1% were Asian, and 94% were Hispanic. Approximately 85% of the students received free or reduced lunch. All students in comparison group 5 were taught using Modeling Instruction.

Procedure for Treatment

Permission

All participating students, along with their parents or legal guardians, signed an assent/consent form acknowledging their participation in the study. If permission was not received from both a student and parent or guardian, that student was not included in the study. The names of individuals participating in the study remained confidential and anonymous. Additionally, for students participating in the video-recorded case study, students and parents signed a second assent/consent letter.

Pretesting of Student Abilities

The investigators and the comparison group instructors were given the necessary materials at the beginning of the year. The comparison groups only participated in testing at the beginning of the year and after completion of the mechanics units. During the introductory unit (Unit 1), the investigators and comparison group teachers determined the baseline ability of the students' conceptual understanding by administering the FCI and identified their problem solving hierarchy with the CST. This unit varied in length and content because different levels of physics were being taught.

Before the end of the constant velocity kinematics unit (Unit 2), each investigator selected five students for case studies. Students were selected by their willingness to participate in a case study and their problem solving approach. The investigators selected one student from each of the problem solving approaches identified by Walsh et al. (2007). The students' approaches were identified by the individual investigator's field notes and experience with the students. Each student was interviewed multiple times throughout the year while solving a problem in a "think aloud" session.

Treatment

All students received standard instruction for the mechanics units; that is, the comparison instructors who identified as Modelers primarily used Modeling Instruction methods, materials, and sequence (see Table 2), while the comparison instructors identified as non-Modelers primarily used more traditional methods. The investigators administered the same testing at the beginning of the year and after completion of the mechanics unit; in addition, during the instruction of the mechanics units they taught an explicit problem solving method outlined below (hereafter referred to as the Expert Method).

After the test on constant velocity kinematics (Unit 2), the investigators introduced the Expert Method for solving problems. The procedure was a modification of the five step

Table 2

Sequence of Modeling Instruction

Unit	Model
1	Scientific Methods
2	Constant Velocity - kinematics
3	Constant Acceleration - kinematics
4	Static Equilibrium
5	Constant Force - dynamics
6	Projectile Motion - kinematics
7	Energy
8	Central Force – circular motion
9	Impulsive Force

approach by Heller et al. (1992) as described above and used a modified version of van Huevelen's (1991) ALPS worksheet. First, the students classified the problem into general type, (e.g. kinematics, dynamics, energy, etc.) and selected the model that applied to the problem. Second, the students drew a picture to represent the problem and wrote down symbols for the known and unknown quantities. Third, the students drew a diagram and/or graph to represent the process. Fourth, an equation was used and the values then substituted to calculate a solution. Finally, the students evaluated the answer to see if it was consistent within the context of the problem. The investigators provided a template for the explicit use of the Expert Method and had the students individually solve problems with it. The investigators strictly emphasized the Expert Method at every opportunity in class. When working through unit materials, the investigators would work out problems on the board demonstrating the Expert Method as well as encourage students to use it when they were stuck on a problem. At times, the use of the Expert Method was outlined in white boarding sessions involving the entire class.

The students were given an additional problem at the end of each unit where they were asked to emphasize the steps of the Expert Method. These problems were collected to track the usage by each student. The students then white boarded the problems at the end of each unit in groups and presented them to the class.

Data Analysis

Types of Data Collected

The following types of data were collected and analyzed for the study:

1. Pre and posttest FCI scores were used to measure gains in conceptual understanding.
2. Post-MBT scores were used to determine problem solving competency in physics.

3. CST pre and postassessments were used to determine whether and to what degree students shifted from a novice-like problem solving strategy to an expert-like problem solving strategy.
4. The end of unit problems (EUPs) were analyzed by a rubric designed to assess a student's problem solving approach.
5. Selected problems from the end of unit tests from the Modeling Instruction materials were collected. However, these problems were discarded as a source of meaningful data because the problems were written to provide an inherent structure to students' solutions, leaving them little to no opportunity to demonstrate use of the Expert Method.
6. Qualitative data was taken via case study interview videos, student surveys, and investigator field notes. The data were analyzed to determine if qualitative data supported, contradicted or explained quantitative data.

Quantitative Data Analysis Description

1. The experimental groups' pre-FCI scores were contrasted with the comparison groups' scores to determine if all students in the study entered physics with approximately the same conceptual understanding. The post-FCI was used to determine what effect the experimental treatment had on students' conceptual growth.
2. The MBT was administered as a posttest, and treatment scores were analyzed with the comparison group scores to determine if students finished the course with similar problem solving abilities. The posttest scores were used to assess whether the experimental treatment increased students' quantitative problem solving abilities.
3. The experimental groups' pre-CST scores were used to determine how students categorized physics problems prior to instruction. The post-CST was used to see if the

experimental treatment caused students to shift more towards expert-like problem solving strategies when compared to the students in an unaltered Modeling class. The CST yielded three scores for each student for each administration: an Expert score that measured to what degree students sorted the problems based on physical model, a Novice score that measured to what degree students sorted the problems based on a surface feature, and a Question Asked score that measured to what degree students sorted the problems based on the question the problem was asking. Table 3 shows the groupings of subjects tested in the CST.

Each student received six CST scores, three from the preassessment and three from the postassessment, each ranging from 0 to 100. The CST scores are referred to as pre-

Table 3

Groupings of Subjects Tested in CST Categories

Models	Surface Features	Questions Asked
Constant Velocity	Free Fall	Distance
Constant Acceleration	Inclined Plane	Time
Newton’s Second Law	Spring	Speed
Energy	Pulley	Velocity
Momentum	Baseball Bat	Initial Velocity
Circular Motion	Elevator	Final Velocity
	Vehicle	Acceleration
	Circle	Net Force
	Graph	Interpret Graph

Expert (pre-E), pre-Novice (pre-N), pre-Question Asked (pre-QA), post-Expert (post-E), post-Novice (post-N), and post-Question Asked (post-QA). A high score in the Expert category meant that the student sorted the problems largely based on physical models, while a high score in the Novice category signified that the student sorted the problems largely based on surface features; a high score in the Question Asked category indicated that the student sorted the problems largely based on the questions asked.

Upon scoring, many students were discovered to have grouped the constant velocity and constant acceleration problems together and they called them “kinematics” problems (or something similar); therefore, the investigators recalculated both the pre- and post-CST Expert scores based on the more common groupings. No significant change in the correlations with the other data was found, therefore, the rescored data was not used.

4. The EUPs were analyzed for the following characteristics: evidence of structure, use of correct equations, absence of erasures, use of a meaningful diagram or graph, and a reasonable answer. Evidence of student utilization of the characteristics was used to assign a score for each EUP. The scores for the EUPs determined the students’ median problem solving approaches. The approaches were adapted from Walsh et al. (2007), and are characterized as found in Table 4. Appendix B contains a more detailed categorization.
5. Students’ Expert Method rubric scores from the EUPs were compared to other data gathered in order to determine if high usage of the problem solving method correlated with higher FCI, CST, and MBT scores.

Table 4

Characteristics of Problem Solving Approaches

Name of Approach	Description of Approach
No Clear Approach	Analyzes the situation based on the given variables
	Proceeds by trying to use the variables in a random way
	Refers to variables as terms
	Conducts no evaluation
Unstructured Plug-and-Chug	Analyzes the situation based on the required variable
	Proceeds by choosing formulas based on the variables in a trial and error manner
	Refers to concepts as variables
	Conducts no evaluation
Structured Plug-and-Chug	Qualitatively analyzes the situation based on required formulas
	Plans the solution based on the variables and proceeds systematically
	Refers to concepts to guide the solution
	Evaluates the solution
Expert-Like	Qualitatively analyzes the situation
	Plans and carries out solution in a systematic manner based on that analysis
	Refers to concepts to guide the solution
	Evaluates the solution

Qualitative Data Analysis Description

1. Field notes were evaluated to determine if there existed anecdotal support for intentional problem solving.
2. Video recordings of selected students solving additional EUPs were analyzed as case studies to assess the progress of the students during implementation of the Expert Method.

Timeline

Treatment

Treatment began in Unit 2 and continued until the conclusion of Unit 9. During Unit 6 through Unit 9, the investigators encouraged Expert-Like problem solving, but did not require its use or provide explicit structure. Posttesting occurred upon the completion of Unit 9. There were differences in the date when students reached certain points in the treatment because the investigators taught differing amounts of introductory material in Unit 1 (i.e., basic physics students needed a much longer introduction than those in junior-level classes); however, the treatment began after the first week of school and concluded by the middle of May.

Assessment

The CST and FCI were administered within Unit 1 to establish baseline scores. Investigator field notebooks were maintained beginning in Unit 2 and concluding in Unit 9. The CST, FCI, and MBT were administered upon completion of Unit 9.

Results

Quantitative Data Analysis

The FCI provided an indication of a student's conceptual knowledge. Analysis of the pre-FCI suggested differences in the students; consequently, the students were divided into three groups: Advanced, Regular and Fundamental. The post-FCI indicated that students taught with Modeling Instruction gained more conceptual knowledge than traditionally taught students. The post-FCI suggested that the comparison students who were taught using Modeling Instruction exclusively gained more conceptual knowledge than the experimental students who were taught using the combination of Modeling Instruction and explicit problem solving methods.

The CST revealed what students look for when solving problems. Analysis of the pre-CST indicated that all students were similar, regardless of their inclusion in the treatment or comparison groups. The post-CST showed that students taught by Modeling Instruction moved away from using the Questions Asked categorization, while those taught with traditional methods moved toward the Questions Asked categorization.

The MBT measured problem solving ability. Those students who demonstrated a conceptual knowledge (as indicated by their post-FCI score) exhibited stronger problem solving skills as measured by the MBT.

The following statistical tests were performed to see whether different groups were similar or dissimilar, or if two items had a correlation. When comparing the means of two groups, the investigators used the non-directional t-test with $\alpha = .05$ to determine if the null hypothesis was accepted or rejected. When comparing a larger number of means from different groups, the ANOVA test was utilized to determine whether groups were significantly different. The conservative Scheffe test was then employed where differences appeared on the ANOVA test in order to determine which groups were different from each other. In order to address the inherent assumptions in the ANOVA test that all variances of the groups were equivalent, a Levene's test was employed to check that assumption. Due to Levene's test results revealing unequal variances, a combination of the Brown-Forsythe and Welch tests was applied to determine if groups were similar. In cases where variances were not similar, the Tamhane post hoc analysis was calculated because it did not assume equal variances. For correlations, the Spearman rho correlation was determined instead of the Pearson correlation because variances between groups were often different. All correlations were significant at $\alpha = .05$.

Pre-FCI. The FCI was given as a pretest to establish the students' prior knowledge before any instruction within a physics class. Because the students tested were enrolled in differing class types, schools and geographical locations, it was essential to determine a baseline of students' initial conceptual knowledge. Preliminary statistical analysis of pretest scores to determine whether or not students were drawn from the same population indicated that the students did not possess similar prior knowledge and were drawn from diverse populations. The students appeared to fit into one of three statistically distinct populations. Table 5 shows the grouping of students who began the school year in statistically similar levels in conceptual understanding of physics.

Factors such as class size, class culture, and curriculum design varied greatly for the Fundamental group when compared to the Regular and Advanced groups. The Regular group was considered to be typical physics classes with similar curricula and class culture. The Advanced group was comprised primarily of honors classes with the exception of Comparison group 2 (C2), which began with a higher than average FCI pretest scores.

Pre-CST. The CST measured students' ability to categorize physics problems in three ways. The assessment specifically measured to what degree the students categorized problems by the model used to solve the problem, the surface features within the problem, and the question the problem was asking. A baseline was determined for students entering the physics class by

Table 5

Grouping of Students Who Were in Similar Levels at Pretest

Group	Members	Treatment	Modeling	Mean	SD
Fundamental	I4 Fundamental Physics (FT)	Yes	Yes	5.86	2.52
	I4 Regular Physics (FT)	Yes	Yes		
	I2 Fundamental Physics (FT)	Yes	Yes		
Regular	I2 Regular Physics (RT)	Yes	Yes	7.15	3.36
	I3 Regular Physics (RT)	Yes	Yes		
	C4 Regular Physics (RC)	No	No		
	C3 Regular Physics (RC)	No	No		
Advanced	I1 Honors Physics (AT)	Yes	Yes	9.07	4.10
	C2 Regular Physics (AC)	No	Yes		
	C5 Honors Physics (AC)	No	Yes		

Using the Welch $F(2, 167.701) = 21.99$, $p < .001$ and Brown-Forsythe $F(2, 183.39) = 21.36$, $p < .001$ tests the investigators rejected the null hypothesis that all groups were statistically similar. Post hoc comparisons using the Tamhane test indicated that the mean score for each group was statistically different from the other groups (Appendix A, Table A1).

administering the CST as a pretest. Students performed similarly on the pre-CST regardless of group (see Table 6).

CST results indicated that students categorized each problem by the underlying model at an average of about 25%. Caution must be exercised in interpreting the significance of the simple mean score because there exists an inherent Expert score that coincided with the Question Asked score due to the design of the CST. From a Spearman rho correlation, it was apparent that pre-Question Asked scores on the pretest correlated strongly and positively with pre-Expert

Table 6

Categories of Pre-Card Sorting Task

Category	Mean	SD
% Expert	24.55	7.881
% Novice	11.89	9.254
% Question Asked	39.96	22.371

scores $\rho(434) = .859$, $p < .001$. That suggested that students scoring highly in the Question Asked category acquired a parallel higher Expert score.

Students did not seem to categorize problems based on the surface features within a problem. Students only categorized based on surface features about 12% of the time. It was apparent that most students came into the physics classes knowing that the objects within a problem were not the most essential piece of information. Figure 1 exhibits how most students scored about 10% Novice regardless of Expert or Question Asked scores on the pretest; however, students who scored above 10% Novice showed a significant decline in Expert scores.

Figures 2, 3, and 4 show frequencies of student categorizations on the CST. Students tended to categorize problems based on the question being asked rather than by model with a frequency of about 40%. The investigators anticipated that students would initially categorize the problems using a Question Asked organization because to students who have a low conceptual understanding of questions, grouping the problems by the question asked is the simplest method. This approach would be consistent with students' previous experience solving word problems in other math and science classes.

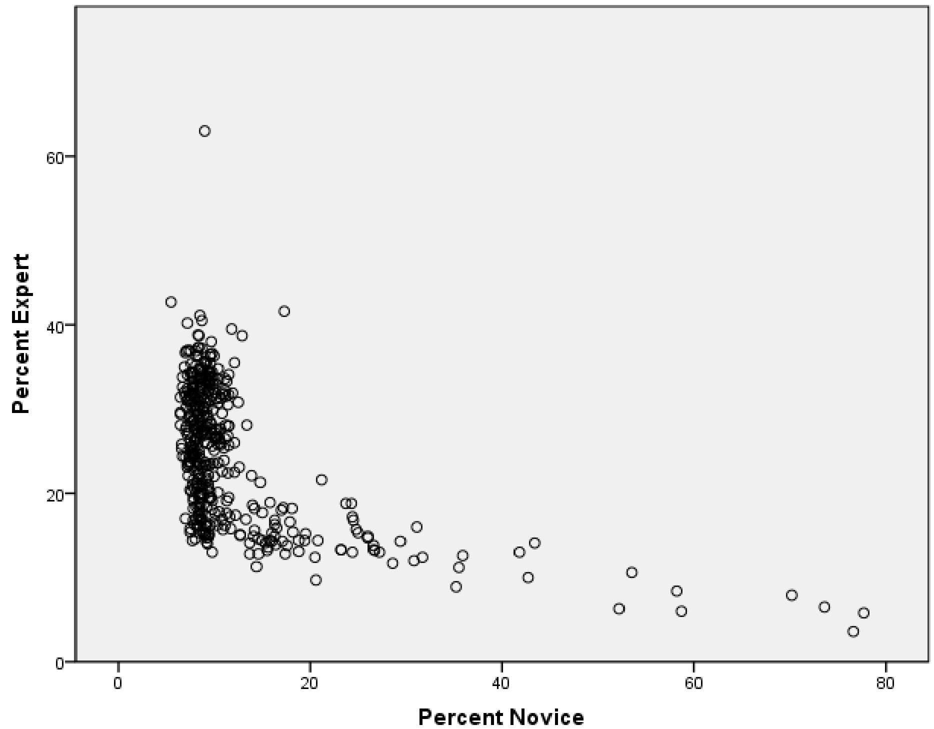


Figure 1. Distribution of student categorizations for Card Sort Task.

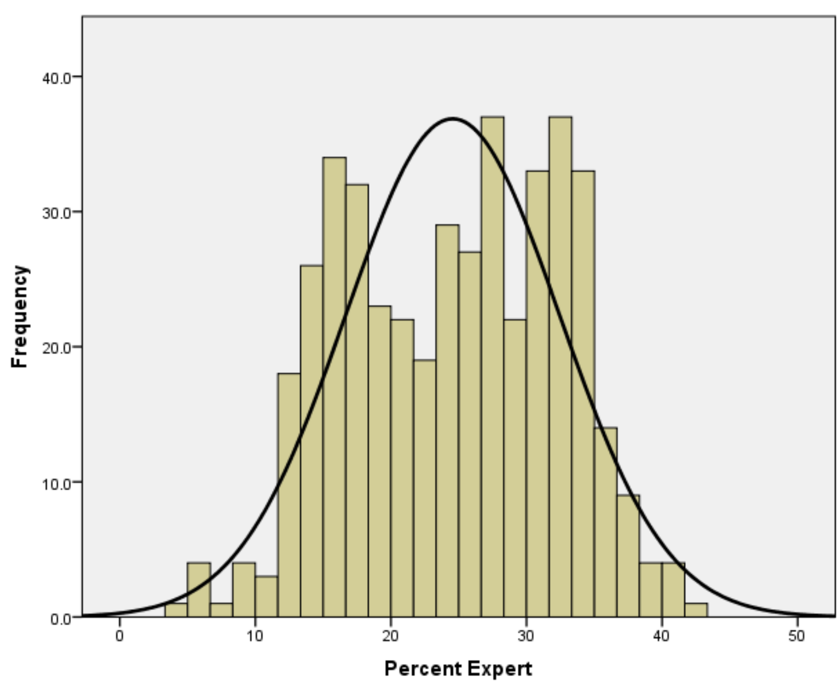


Figure 2. Frequency of Expert categorizations on Card Sort Task.

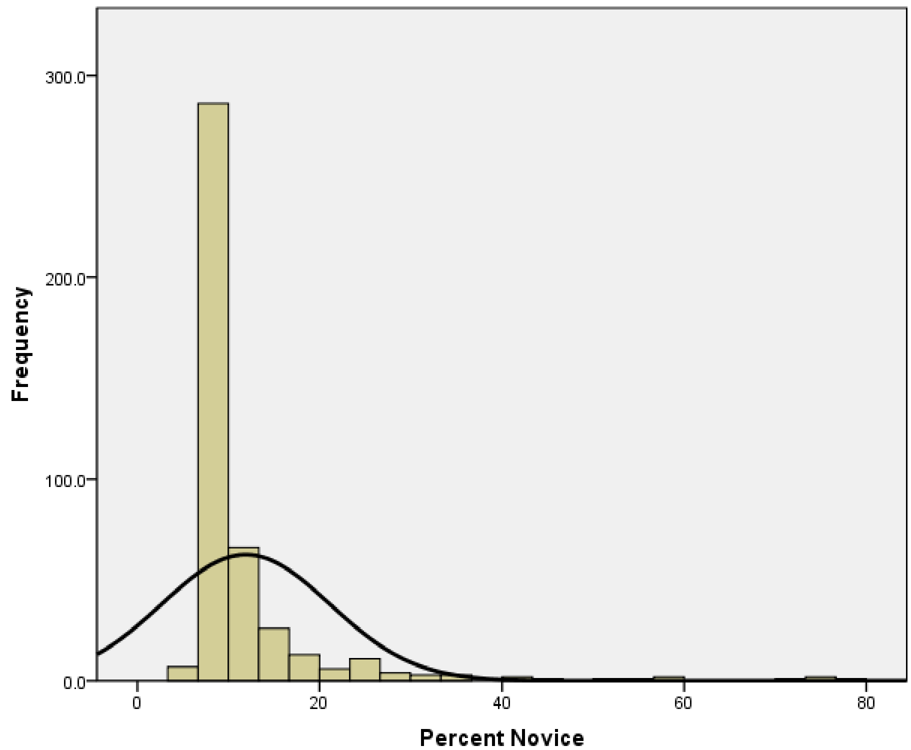


Figure 3. Frequency of Novice categorizations on Card Sort Task.

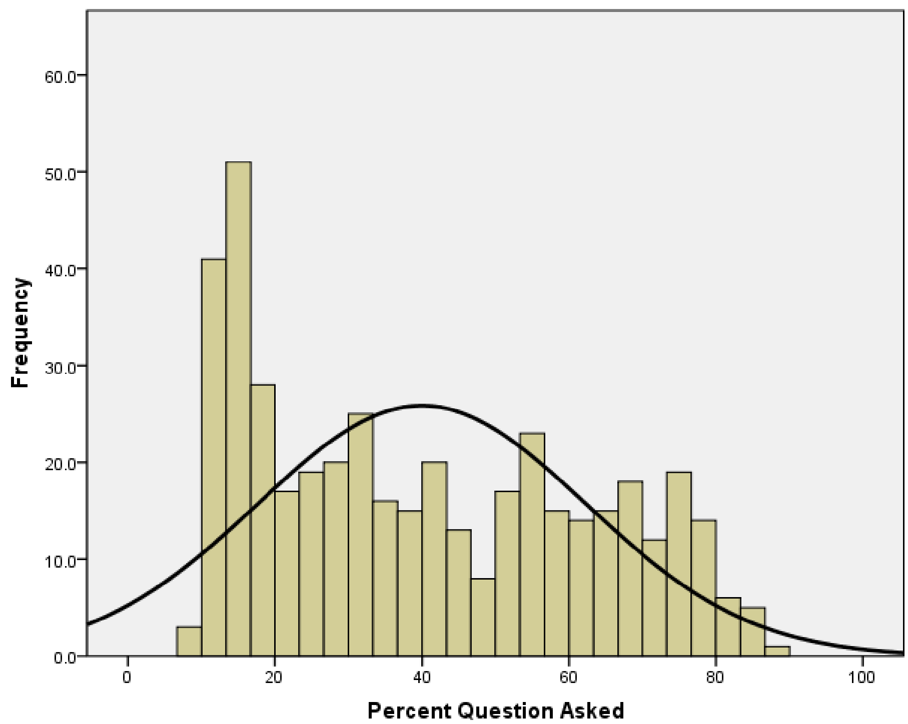


Figure 4. Frequency of Question Asked categorizations on Card Sort Task.

Post-MBT. After assessment of initial student abilities and implementation of the explicit problem solving instruction in concert with the Modeling Instruction, the students' quantitative problem solving ability was assessed using the MBT. The effect of the treatment on the students' quantitative problem solving ability was evaluated (see Table 7). It appeared that the Regular Treatment (RT) group scores were higher, but with a wider range of scores when compared to the Regular Comparison (RC) group (the Regular Comparison group was limited with only 10 students completing the MBT). Statistical analysis of MBT scores for the RT group compared to the RC group implied that the two groups possessed different quantitative problem solving abilities; however, with such a small sample size in the comparison group, conclusions about how the treatment affected the regular group's quantitative reasoning must be approached with care.

To determine how effective the treatment was in developing quantitative skills for the advanced level courses, MBT scores for the Advanced Treatment (AT) group were compared to the results of the Advanced Comparison (AC) group (Note: Comparison group 5 did not take the MBT). The two groups displayed similar mean scores with approximately the same range. Analysis of the scores indicated that both groups possessed similar quantitative problem solving abilities, as measured by the MBT. One can surmise that the treatment did not increase students' quantitative reasoning skills among the more advanced groups. The investigators questioned the possibility that Modeling Instruction was a factor in these results; however, there was not a comparison of Modeling Instruction to traditional instruction because most of the students who took the MBT were enrolled in a Modeling class.

Table 7

Mean Post-MBT Scores by Treatment Group

Treatment Groups	Post-MBT Scores	
	Mean	Standard Deviation
Fundamentals Treatment (FT)	6.55	2.31
Fundamental Comparison (FC)	N/A	N/A
Regular Treatment (RT)	8.88	4.00
Regular Comparison (RC)	6.30	2.35
Advanced Treatment (AT)	11.68	3.975
Advanced Comparison (AC)	10.61	3.65

Using an independent sample, non-directional, t-test the investigators rejected the null hypothesis that the Regular Groups were equal [t(73) = -1.93, p = 0.01]. Using another non-directional t-test, with t(56) = 1.07, p > 0.649, the investigators accepted the null hypothesis that the Advanced Group's means were statistically similar (Appendix A, Table A2).

The investigators examined particular students who were the most and the least impacted by the experimental treatment because no clear conclusions could be drawn from comparing post-MBT scores within groups.

High MBT group. The students used four approaches, which were indicative of a student's problem solving strategy; and the students were tracked using a rubric designed to measure the approach (Walsh et al., 2007). See Table 8 which outlines the approaches and scores used. The rubric was applied to each EUP and the students were given an approach score. The median of each student's scores was useful in assessing their approach.

Table 8

Scores Associated with Problem Solving Approach

Name	Approach Score	Rubric Score
No Clear Approach	0	≤ 1.5
Unstructured Plug-and-Chug	1	1.5 < x ≤ 2.0
Structured Plug-and-Chug	2	2.0 < x ≤ 2.5
Expert-Like	3	> 2.5

There were 21 students, representing approximately 10% of the sample, who scored 14 or above on the MBT (hereafter referred to as the High MBT group). They were categorized as good problem solvers. Within the High MBT group, 25% of the students had a median approach score of 3.0, which meant they were Expert-Like in their approach to the EUPs. While 50% had a median score above 2.65, which categorized those students between Structured Plug-and-Chug and Expert-Like in their approach. Only 25% of the High MBT group scored below 2.0 (meaning that they used No Clear Approach or the Unstructured Plug-and-Chug Approach). See Table 9 for the analysis of the approach used by the High MBT group.

Low MBT group. The investigators also analyzed the group of students who scored 6 or less on the MBT, which included 95 students or about 33% of the sample (hereafter referred to as the Low MBT group). By examining how those students performed on the EUPs, it appeared that they used a different approach from the High MBT group. The Low MBT group showed a greater tendency to use lower functioning approaches. Seventy-five percent of the students in this group performed at or below 1.7 as their median problem solving approach (see Table 10),

Table 9

Median Problem Solving Approach for High MBT Group

n	21	
Median	2.65	
Percentiles	25	2.00
	50	2.65
	75	3.00

Table 10

Median Problem Solving Approach for Low MBT Group

n	95	
Median	1.70	
Percentiles	25	1.30
	50	1.70
	75	1.70

meaning that most of the Low MBT group frequently used an Unstructured Plug-and-Chug approach or No Clear Approach.

All students were given a median problem solving approach and assigned to a specific approach based on their mean score. The investigators assessed each student within a given approach and how they performed on the MBT to determine if the categories showed meaningful differences in student performance. Each approach performed statistically dissimilar in problem solving ability within physics; thus distinctions between No Clear Approach, Unstructured Plug-and-Chug, Structured Plug-and-Chug and Expert-Like appeared to be meaningful (see Table 11 for statistical analysis of Problem Solving Approach MBT Scores).

Post-CST. Multiple posttest measures were studied in order to evaluate improvement of the treatment group weighed against the comparison group. The post-CST, used to measure how students categorized physics problems, was thought to reveal more information than the pre-CST about what students considered when reading a problem. See Table 12 for statistical analysis of the CST scores for the researched populations.

There were some differences in how the Regular Treatment (RT) group performed on the CST when compared to the Regular Comparison (RC) group. The RT and RC groups performed similarly when comparing the post-Expert scores measured by the CST. This score implied that

Table 11

Problem Solving Approach MBT Scores

Approach	Mean	Standard Deviation
No Clear Approach	6.60	2.290
Unstructured Plug-and-Chug	7.03	2.984
Structured Plug-and-Chug	8.63	3.080
Expert-Like	12.32	4.538

Using the Welch $F(3, 82.85) = 16.62$, $p < .001$ and Brown-Forsythe $F(3, 85.20) = 22.30$, $p < .001$ test the investigators rejected the null hypothesis that all groups were statistically similar. Post hoc comparisons using the Tamhane test indicated that the mean score for each approach was statistically different from the other approaches except when comparing the No Clear Approach and Unstructured Plug-and-Chug Approaches. It was clear that these groups performed as if they were guessing on the MBT (Appendix A, Table A3).

Table 12

Statistical Analysis of CST Scores by Group

Group	Post-Expert		Post-Novice		Post-Question Asked	
	Mean	SD	Mean	SD	Mean	SD
Regular Treatment (RT)	30.53	10.49	13.92	13.04	29.35	19.35
Regular Comparison (RC)	29.65	7.56	10.24	7.62	56.01	22.79
Advanced Treatment (AT)	34.77	9.82	10.21	4.88	25.97	15.43
Advanced Comparison (AC)	30.80	6.92	10.04	4.15	38.98	20.26

Independent samples t-test accepted the null hypothesis that the post-Expert scores for the Regular Treatment group (RT) had statistically similar means when compared to the Regular Comparison group (RC) [$t(150) = -.601$, $p = .549$]. Additionally, from a second independent samples t-test where $t(150) = -2.17$, $p = .031$, the investigators rejected the null hypothesis that the post-Novice scores for RT were statistically similar to the post-Novice scores for RC. Finally, the investigators used an independent samples t-test to reject the null hypothesis that RT's post-Question Asked scores were significantly similar to RC's post-Question Asked scores. From $t(147.4) = 7.783$, $p < .001$, the Regular Treatment group scored significantly lower on the post-Question Asked category in comparison to the Regular Comparison group. Independent samples t-test accepted the null hypothesis that the post-Expert scores for the Advanced Treatment group (AT) had statistically similar means when compared to the Advanced Comparison group (AC) [$t(54) = -1.769$, $p = .083$]. Additionally, from a second independent samples t-test where $t(54) = -.131$ $p = .897$, the investigators accepted the null hypothesis that the post-Novice scores for AT were statistically similar to the post-Novice scores for AC. Finally, the investigators used an independent samples t-test to reject the null hypothesis that AT's post-Question Asked scores were significantly similar to AC's post-Question Asked scores. From $t(54) = 2.568$, $p < .014$, the Advanced Treatment group scored significantly lower on the post-Question Asked category in comparison to the Advanced Comparison group (Appendix A, Table A4).

students of both groups had about 30% of their problems sorted by underlying model. When comparing the post-Novice scores for both the groups, the RC group had a slightly lower mean score than the RT group. When looking at the post-Question Asked scores measured by the CST, the RT group had a drastically lower average score than the RC group; in other words, the RT group tended to categorize physics problems by the question being asked less frequently than the RC group. It appeared that the treatment for the RT group tended to move students away from categorizing problems simply by the quantity desired as an answer and into a more model-centered scheme.

When comparing the Advanced Treatment (AT) group and the Advanced Comparison (AC) group, there were similar results found to what occurred in the Regular groups. The AT and AC groups performed similarly in both post-Expert and post-Novice scores measured by the CST. The major difference between the two groups was their performance on the post-Question Asked scores where the investigators noted that students in the AT group categorized physics problems based on the Question Asked with a frequency of 26%, whereas the AC group categorized those same problems with a frequency of 39%. As was the case with the Regular group, it appeared that the experimental treatment encouraged students to group problems based on their model as opposed to the questions being asked. It must be noted that the observed differences were unable to distinguish between the effects of Modeling Instruction and the explicit problem solving methods taught within the same populations.

Expert, Novice, and Question Asked Hake gains were compared between Modeling teachers and traditional teachers, because of considerable starting differences. The Hake gain was a measure of how much a student improved in comparison to total possible improvement. The Hake gains could be positive or negative and were calculated by:

$$\langle g \rangle = \frac{\text{post test score} - \text{pre test score}}{\text{total possible} - \text{pre test score}}$$

The Hake gains for the students in the study are outlined in Table 13.

Comparing the Novice gains of the Modeling students with the traditional students showed that Modeling Instruction caused more students to categorize fewer physics problems based on Surface Features. Traditional instruction appeared to have minimal effect in that realm. When looking at the Question Asked gains, similar results were observed. Modeling Instruction

Table 13

Hake Gains for Students in the Study

Group	Expert Gain		Novice Gain		Question Asked Gain	
	Mean	SD	Mean	SD	Mean	SD
Modeling	4.62	9.8	-1.30	5.43	-4.30	29.198
Traditional	6.90	7.21	0.16	11.66	22.82	23.980

Using a two-tailed t-test at where t(144.219) = 2.270, p = .025, the investigators rejected the null hypothesis that Modeling Expert gain was statistically similar to traditional instruction Expert gain. The investigators used another non-directional independent samples t-test to compare Modeling Instruction vs. traditional instruction Novice gains. With t(384) = -1.572, p = .117, the investigators accepted the null hypothesis that the gains of the Modeling group compared to the traditional group were statistically similar. In comparing Question Asked gain between Modelers and traditional teachers the investigators once more used a non-directional independent samples t-test. Based on t(383) = 7.416, p < .001 the investigators rejected the null hypothesis that Modelers were statistically similar to traditional teachers (Appendix A, Table A5).

showed a decrease in the percentage of students who categorized physics problems by the Question Asked alone; whereas traditional instruction students actually demonstrated an increase in that type of categorization. Of particular interest was the difference in sign in Question Asked gains between Modeling students and traditional students. It appeared that through Modeling, students naturally moved away from organizing their thinking based on what question was being asked; traditional classrooms appeared to encourage that thinking. Also of note in the analysis was that students in the traditional instruction classrooms gained more by the Expert measure than did the students in the Modeling classrooms.

Although gains were modest in the Question Asked and Expert categories, the investigators believed that a possible explanation for higher non-Modeling gains might lie in the fact that higher Question Asked responses yielded an inherently larger Expert score. In looking at how the post-CST differentiated between the Question Asked and Expert categories, the two measures were strongly related. There was a strong positive correlation between the Question

Asked and Expert scores on the post-CST (similar to what was seen on the pre-CST); therefore, the CST did not appear to be a useful tool in distinguishing students who were answering based on Question Asked versus those who relied on Expert strategies to formulate their groups. In order to differentiate between students in the two categories, it would be necessary to look at which teaching style created students with both high Expert scores and low Question Asked scores. See Figures 5, 6, and 7 for the analysis of Post-CST scores for Modeling and traditional instruction students.

Looking at the graphs in the Figures 5, 6, and 7, it was clear that students could take one of two tracks in earning a high Expert score. Students who categorized truly by Question Asked had both a high Question Asked score and high Expert score, whereas students who truly categorized by underlying model had a high Expert score and a relatively low Question Asked score. Modeling Instruction produced both types of students whereas traditional instruction

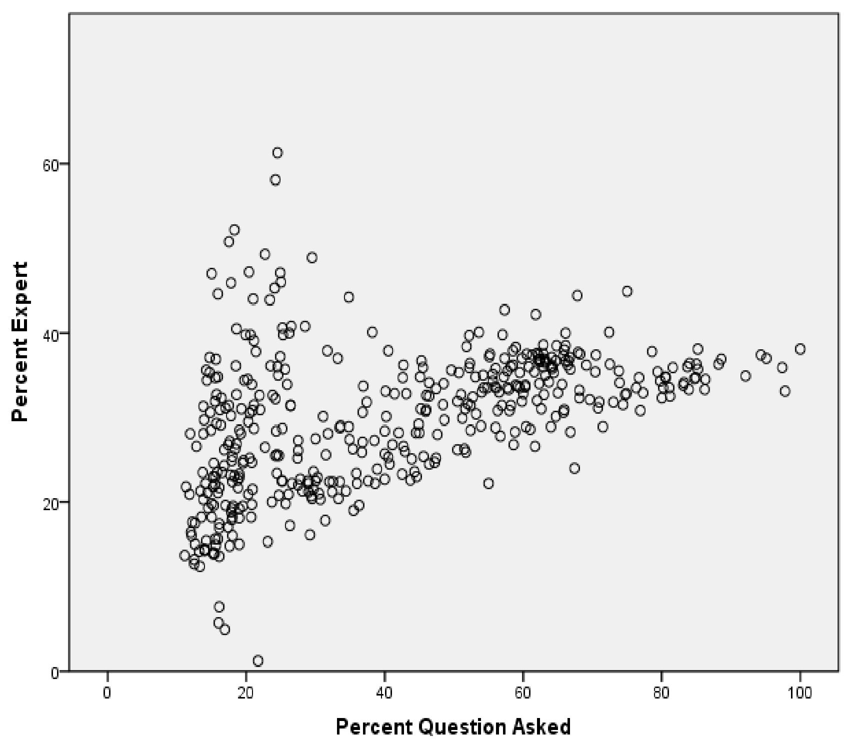


Figure 5. Post-CST Modeling vs. traditional. All results comparing post-Question Asked scores with post-Expert scores using a Spearman rho correlation found another strong positive correlation ρ(431) = .534, p < .001.

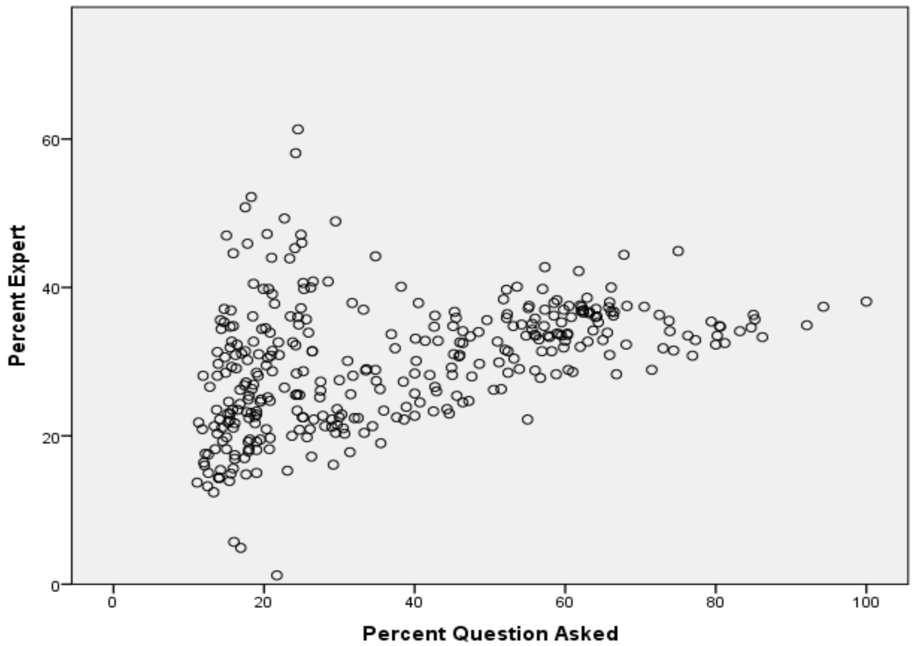


Figure 6. Post-CST Modeling vs. traditional: Modeling results.

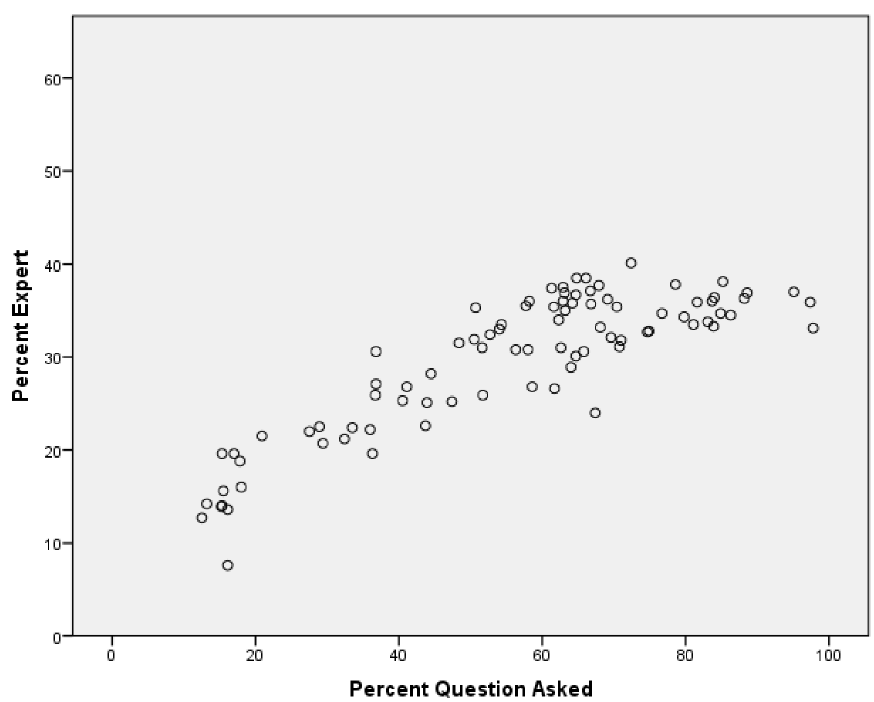


Figure 7. Post-CST Modeling vs. traditional: traditional results.

appeared to produce students who categorized in a Question Asked manner; there seemed to be no students who were truly using an underlying model to categorize physics problems in a traditional class. This difference was significant and it seemed to imply that Modeling students tended to think about conceptual aspects of the problem, as a whole, instead of the specific variable the question was asking for. It did not look as if that happened in the traditional classrooms.

Additionally, students with high post-Expert scores on the CST were evaluated. The investigators questioned how useful a predictor the CST was in terms of differentiating students between No Clear Approach (i.e., Novice), Plug-and-Chug (i.e., Question Asked), and Expert-Like problem solvers. In some instances, there was substantial evidence supporting CST predictive power. In other instances, there was significant evidence that seemed to question CST applicability. To help determine if the CST was useful for the study, the investigators correlated the CST results to established predictors, including the FCI and MBT. Considering all students who took both the post-CST and post-FCI, the two measures correlated on a Spearman rho $\rho(340) = .233$, $p < .001$. There was a statistically significant correlation between post-FCI and post-Expert performance, supporting CST predictive power in designating experts.

Post-FCI.

Regular Group. The investigators were interested in the effect of the experimental treatment on students' conceptual understanding of forces as measured by the FCI. Since students in the Regular Group started out statistically similar, the results of the post-FCI were compared between students within the Regular group. See Figures 8 and 9 for the post-FCI results for RT and RC groups.

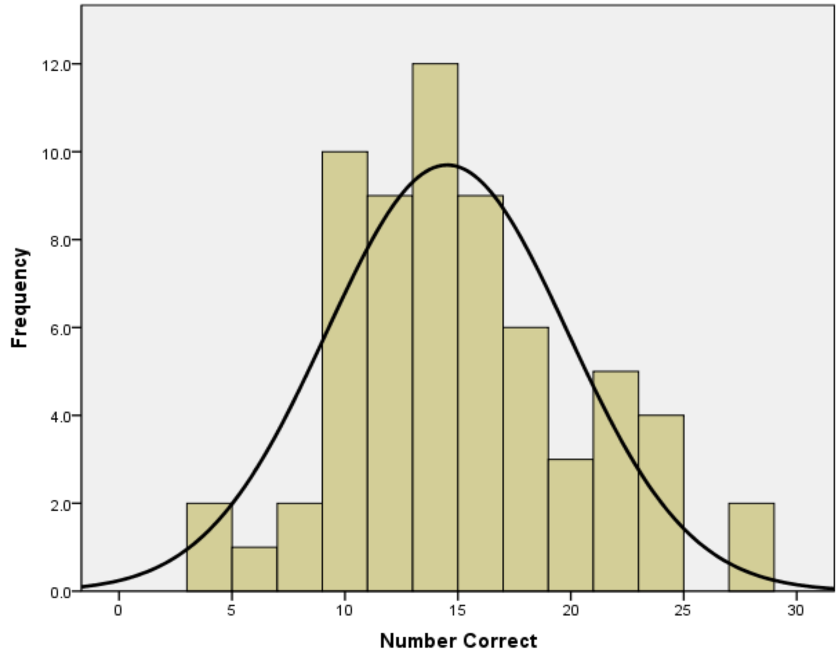


Figure 8. Regular Treatment Post-FCI. Using a non-directional independent samples t-test to compare Regular Treatment (Mean = 14.52, SD = 5.348) and Regular Comparison (Mean = 10.02, SD = 3.479) post-FCI scores t(100.6) = -5.977, p < .001the investigators rejected the null hypothesis indicating the two group had statistically different means (Appendix A, Table A6).

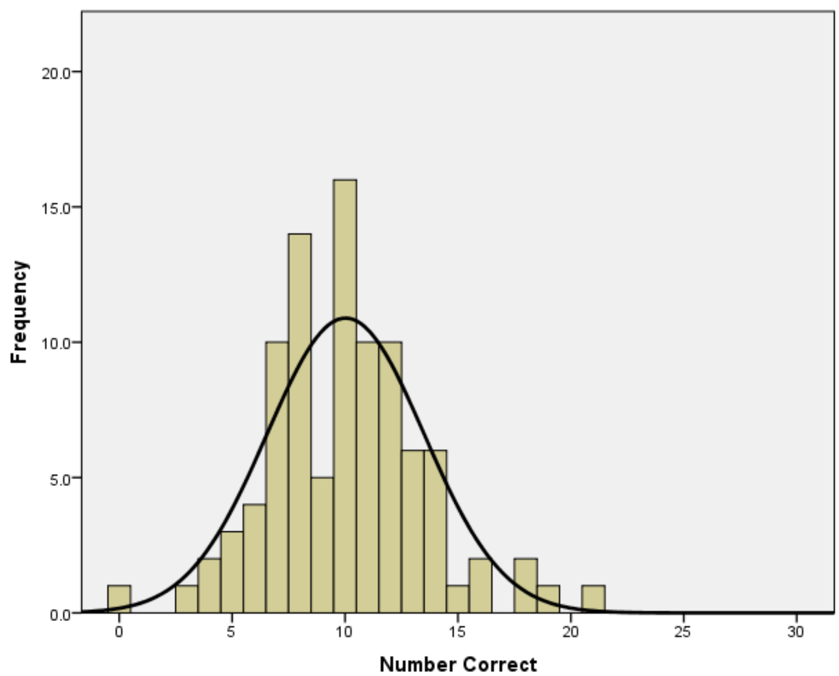


Figure 9. Regular Comparison Post-FCI. Using a non-directional independent samples t-test to compare Regular Treatment (Mean = 14.52, SD = 5.348) and Regular Comparison (Mean = 10.02, SD = 3.479) post-FCI scores t(100.6) = -5.977, p < .001the investigators rejected the null hypothesis indicating the two group had statistically different means (Appendix A, Table A6).

The scores of the Regular Treatment (RT) group were compared to the scores of the Regular Comparison (RC) group. The RT group outperformed the comparison group, implying they possessed a higher conceptual understanding, as measured by the FCI.

Advanced Group. Similarly, the effect of the treatment on the advanced students' conceptual understanding was evaluated. Results of the post-FCI were compared between students within the Advanced Group. There was a stronger performance on the post-FCI within the AT group, than in the AC group, as shown in Table 14. As demonstrated in both cases, the treatment seemed to be of conceptual benefit to the students; however, all treatment teachers within the Regular group used Modeling Instruction, while comparison group students were

Table 14

Mean Post-FCI Scores by Treatment Group

Treatment Groups	Post-FCI Scores	
	Mean	Standard Deviation
Fundamentals Treatment (FT)	9.24	3.761
Fundamental Comparison (FC)	N/A	N/A
Regular Treatment (RT)	14.52	5.348
Regular Comparison (RC)	10.02	3.479
Advanced Treatment (AT)	21.59	5.217
Advanced Comparison (AC)	18.61	4.379

Using a non-directional independent samples t-test to compare Regular Treatment and Regular Comparison post-FCI scores $t(100.6) = -5.977$, $p < .001$ the investigators rejected the null hypothesis indicating the two groups had statistically different means (Appendix A, Table A6). Using a non-directional independent samples t-test to compare Advanced Treatment and Advanced Comparison post-FCI scores $t(63) = -2.477$, $p = .016$ the investigators rejected the null hypothesis indicating the two groups had statistically different means (Appendix A, Table A7).

taught using traditional methods, which may have accounted for the difference. In the Advanced group, all of the teachers used Modeling; thus, it appeared that the experimental treatment had a significant effect on conceptual understanding for Advanced students.

Experimental Treatment versus Comparison Group Analysis. Once again, the investigators were interested in determining whether or not Modeling Instruction was causing the higher post-FCI scores for students in Modeling classrooms. The investigators set a baseline measuring the effect of the treatment on students' improvement of conceptual understanding of forces. In order to compare all study participants, apart from whether if they were similar or not, a Hake gain on FCI scores was used. See Figures 10 and 11 for the Hake Gain analysis.

The Hake gain of students within the Treatment group was higher than the gain of students within the Comparison group, which indicated that the experimental treatment was more successful at instilling the conceptual ideas to students than instruction in Comparison classrooms. There was a significant difference in the improvement of student understanding as measured by a Hake gain for individuals in the Treatment when compared to the Comparison group.

Modeling Instruction versus Traditional Instruction Analysis. The Treatment group consisted of teachers who implemented Modeling Instruction whereas the Comparison group had teachers who implemented Modeling Instruction along with teachers who utilized more traditional approaches. Modeling Instruction may have accounted for part of the gains in conceptual knowledge measured by the FCI. See Figures 12 and 13 for analysis of gains in Modeling classrooms and traditional classrooms.

Modeling and Experimental Treatment Group versus Modeling Comparison Group Analysis. The effect that Modeling Instruction had on a student's improvement of

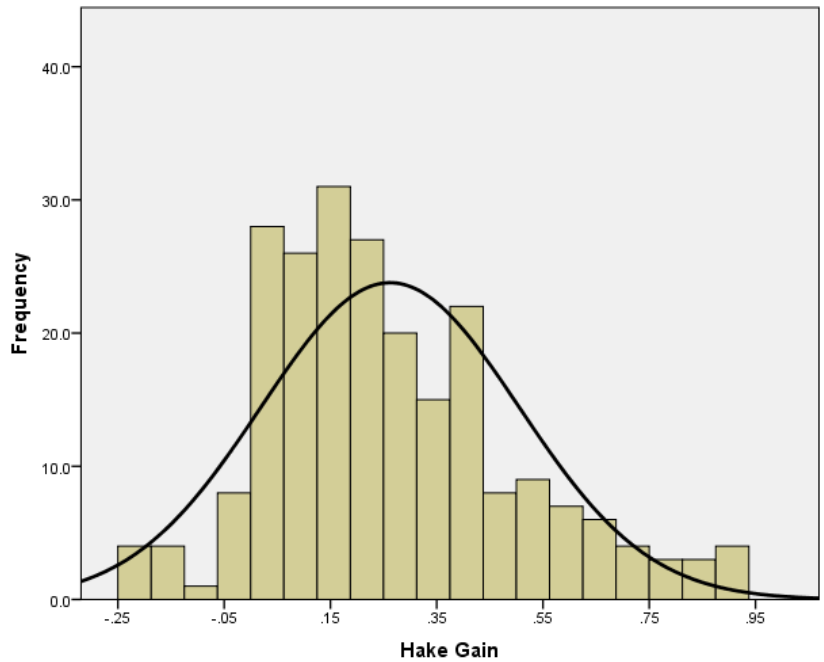


Figure 10. Treatment Hake gain. Based on a non-directional independent samples t-test, the investigators accept the null hypothesis that the mean gain of the treatment group (Mean = .262, SD = .243) was equal to the mean gain (Mean = .224, SD=.213) of the comparison group, t(248.4) = -1.334, $p < .183$. (Appendix A, Table A8).

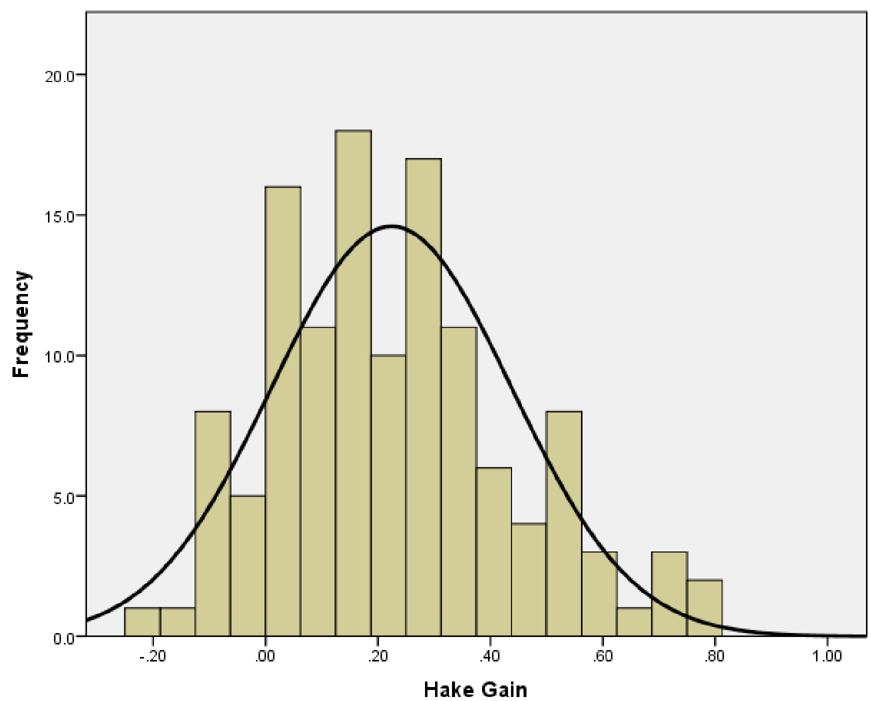


Figure 11. Comparison Hake gain. Based on a non-directional independent samples t-test, the investigators accepted the null hypothesis that the mean gain of the treatment group (Mean = .262, SD = .243) was equal to the mean gain (Mean = .224, SD = .213) of the comparison group, t(248.4) = -1.334, $p < .183$. (Appendix A, Table A8).

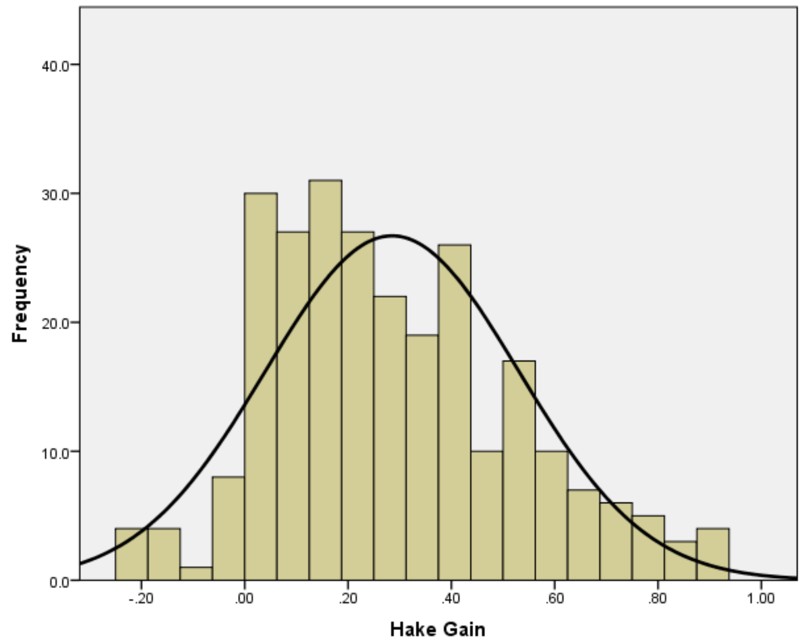


Figure 12. Modeling Hake gains. Based on a non-directional independent sample t-test the investigators rejected the null hypothesis that the mean gain of Modeling Instruction Treatment improvement (Mean = .285, SD = .245) was equal to the mean gain of Traditional Instruction Comparison (Mean = .147, SD = .158), $t(261) = 4.136$, $p < 0.001$ (Appendix A, Table A10).

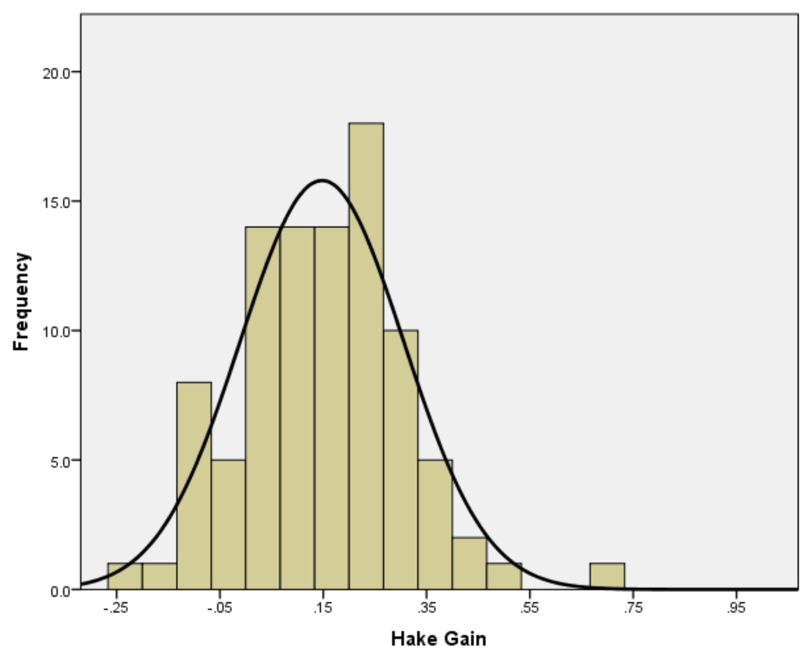


Figure 13. Traditional Hake gains. Based on a non-directional independent sample t-test the investigators rejected the null hypothesis that the mean gain of Modeling Instruction Treatment improvement (Mean = .285, SD = .245) was equal to the mean gain of Traditional Instruction Comparison (Mean = .147, SD = .158), $t(261) = 4.136$, $p < 0.001$ (Appendix A, Table A10).

conceptual understanding of forces as measured by a Hake gain using FCI scores offered insight into the apparent success of the treatment. It appeared that the improvement of students' conceptual understanding of force, as measured by a Hake gain for individuals in a Modeling class, was significantly greater than for the individuals in a traditional class. To assess the influence of the experimental treatment independent of Modeling Instruction, only Modeling Treatment gains were compared to Modeling Comparison gains.

The Comparison group (Mean = .456, SD = .191) made larger improvements in the FCI Hake gain than did the Treatment group (Mean = .262, SD = .243). Other factors, besides the experimental treatment may have been responsible for the difference. There were only 31 students in the Modeling Comparison group and both of the teachers who taught Modeling Comparison students were more experienced in Modeling (and teaching, in general), than the experimental treatment teachers. Additionally, all 31 students in the Modeling Comparison group were previously placed in honors level classrooms; whereas most of the Treatment students were in regular or fundamental level classrooms.

Problem Solving Approach and FCI Score Correlation. Students who lacked conceptual understanding seemed to underperform in nearly every assessment, especially with problem solving tasks. This was most apparent in the analysis of the post-FCI. Hestenes (1987) contended that students needed a strong conceptual basis before approaching problem solving. The student approaches and their achievement on the post-FCI were studied. The students who used No Clear Approach were statistically similar to the students who used an Unstructured Plug-and-Chug approach, but dissimilar from all other groups. Additionally, the students using the Unstructured Plug-and-Chug scored similarly on the post-FCI to the students who used a Structured Plug-and-Chug approach, but dissimilar from the students using an Expert-Like approach. The students using an Expert-Like approach were statistically different from every other group. These results supported Hestenes' assertion regarding conceptual understanding (1987). See Table 15 for the FCI scores analyzed by problem solving approach.

Hestenes offered insight as to why the lower performing students may have had such deficient scores. Student difficulties with the test appeared to stem from real deficiencies in understanding the basic concepts (Hestenes & Wells, 1992). He recognized that the application of a concept to a problem should be the focus of a physics class. Students should be taught that the key to solving a typical physics problem is the application of a model to the given information. Indeed, the problem could not be fully understood until the model had been constructed (Hestenes, 1987).

Table 15

Problem Solving Approach FCI Scores

Approach	Mean	Standard Deviation
No Clear Approach	9.75	3.763
Unstructured Plug-and-Chug	11.54	5.416
Structured Plug-and-Chug	13.32	5.623
Expert-Like	20.78	7.72

Using the Welch $F(3, 80.38) = 19.58$, $p < .001$ and Brown-Forsythe $F(3, 87.08) = 24.059$, $p < .001$ tests the investigators rejected the null hypothesis that all groups were statistically similar. Post hoc comparisons using the Tamhane test indicated that the mean score for the Expert-Like group was statistically different from the other groups. Whereas the test found the Plug-and-Chug groups to be statistically similar but different from No Clear Approach. No Clear Approach was found to test similarly to Unstructured Plug-and-Chug (Appendix A, Table A11).

Qualitative Data Analysis

The investigators gathered a large amount of qualitative data (in the form of talk-aloud interviews, surveys, and investigator field notes). The purpose for this data was threefold: 1) to highlight typical cases of students' problem solving approaches, the extent to which students utilized the Expert Method, and how effective they were while problem solving; 2) to examine traits of students who struggled with quantitative reasoning and attributes of strong problem solvers; and 3) to determine how students perceived the Expert Method. The qualitative data described typical relationships between students' conceptual understanding and problem solving ability. The data also illustrated that the benefits of the Expert Method were dissimilar for high-level and low-level students.

Investigator's Field Note. As would be expected, those students who wrote down the Expert categories generally did better on the FCI and MBT. One exception was SMITH. SMITH was a mainstream Special Education student. The Special Education teacher following his case indicated that SMITH's IQ was quite low. SMITH put in great effort, but was not successful. He would attempt anything he was asked to try. He participated well in his groups; however, his assignments were completed only with extensive assistance from the teacher or another student. He did poorly on all tests. All of the indicators showed that he learned essentially nothing during his year of physics. His FCI scores went from 10 (pre-FCI) to 6 (post-FCI). His MBT score of 4 indicated that he was not even good at guessing answers. His EUPs indicated that he had no structure and no idea how to even begin working on the problems; all but one assessment placed him in the lowest category: No Clear Approach. Even the CST reflected his lack of structure. His pre-CST (E = 12.8%, N = 13.7%, Q = 11.5%) and post-CST (E = 13.7%, N = 13.1%, Q = 11.1%) demonstrated no dominant way of grouping. Yet, his written categories for the post-CST were the expert categories (i.e., physical models). He recognized the categories were what the teacher wanted and they were visible on the board in the classroom, so he copied them onto his paper as his sorting categories. This showed the student was motivated to perform positively, but had no conceptual understanding in order to actually do so. This type of student performance is reiterated in the next section.

Student Interviews. The first video transcript selection was from an interview following Unit 5 (constant force model), presented as an example of the Unstructured Plug-and-Chug approach. The interview took place several weeks after completion of instruction and assessment for the unit (those weeks included spring break). The problem asked students to calculate how much time it would take for a boulder to accelerate down a hill. This student focused her solution on the variables given and correctly identified the variable to be solved, but struggled to combine several concepts in order to solve the problem. This student used only some parts of the Expert Method, namely the pictorial representation and mathematical representation. She omitted identifying the model and problem type and did not create a physical representation. At this point in her solution, she had read the problem, written down the given information, drawn a picture and labeled it inaccurately, attempted to calculate weight components without having analyzed the forces, selected an appropriate kinematics equation, $\Delta y = v_i t + \frac{1}{2} a t^2$, from the list provided on the front board, and was looking between her whiteboard and the front board.

INTERVIEWER: *What are you thinking right now?*

STUDENT: *Um…, I'm just thinking, like, if the net forces should equal m-*a *or* zero*, mass times acceleration.*

INTERVIEWER: *And why do you think that might be important?*

STUDENT: *Um…, I think that's important because if it, um…, I don't know, I think it just, like, tells you which way to go in the problem, like if you need to, I don't know. I mean, usually you could just set things equal to each other if net force equals zero, but if it equals m-a, like if you have a mass and an acceleration, so you have to kind of solve for that.*

INTERVIEWER: *Okay.*

STUDENT: *Okay, I don't think I can go further.*

INTERVIEWER: *Then, in that case, um…, so, how did you choose your problem solving process?*

STUDENT: *Um…, I chose it by, um…, first, I always like to, um…, draw a picture, just to give myself, like, a visual of what I, like, what the problem is asking for, and that kind of helps me, like directs me on how to solve the problem. And then, I kind of just do what I know first, what I, I know that I need to, well, I like, well, I like set it up, like, what I know and then what I need to find out, and after that I just, um…, solve for that.*

INTERVIEWER: *Okay. Um…, how useful did you find your picture in solving the problem?*

STUDENT: *Um…, I find it really useful because I'm, like, I'm more of a visual person, so seeing, um…, a picture, like, helps me, um, understand better, I think.*

INTERVIEWER: *Okay. Um…, what did you think is the physics of this question?*

STUDENT: *Um…, I think the physics of it is it's asking you to, um…, find out how much time the person has to react, to move before, um…, a boulder falls from a, uh…, um…, cliff kind of thing or slope. And, um…, like, everything, like, the degree of the slope is, like, playing, and, like, the height of the, um…, the boulder, and the weight of the rock, it's all playing into, um…, how much time that's going to take. It all has an effect on how fast it's going to fall… and hit the person.*

INTERVIEWER: *Okay, okay. Um…, did you choose one of the experimental models that we've developed in class, and if so, which one?*

STUDENT: *(looks confused) What's the experimental model?*

INTERVIEWER: *Oh, that is…*

STUDENT: *Oh, one of those? (points to front board, where models and equations are listed)*

INTERVIEWER: *Like constant velocity, constant acceleration, free particle, or constant force.*

STUDENT: *Oh, oh, um…, uh…, I didn't choose one, but now thinking about it, um…, it might be constant, I want to say, velocity.*

INTERVIEWER: *Okay. Um…, let me see…um…, do you think you would choose your same problem solving process again?*

STUDENT: *Um…, probably, probab…, probably not. Um…, well, I would definitely draw the picture and all that again, but I probably would, um…, do a different, um…, way to work out the problem and get a, an answer.*

INTERVIEWER: *Okay. Why do you think you got stuck at the, the point that you reached?*

STUDENT: *Um…, I think I got stuck because I didn't know which equation to choose to start it off, and, um…, yeah.*

INTERVIEWER: *Okay. Um…, I… what other information do you think you would need to solve the problem?*

STUDENT: *Um…, I think…some sort of acceleration to plug in.*

The student had focused her solution on variables and rarely tied them to concepts. She selected equations based on the variables given and not on the concepts involved in the problem, as evidenced by her uncertainty whether the forces on the boulder were balanced or not. Her uncertainty arose from trying to manipulate equations instead of conducting a qualitative analysis of the problem and selecting or constructing a model. Her rambling speech about her solution evidenced its lack of structure and coherency, as did her inability to describe how she would approach the problem differently. All in all, this qualitative data suggested a lack of basic conceptual understanding. Her quantitative data corroborated this inference, as her post-FCI score of 9 and MBT score of 4 are well below the threshold of 60% needed for effective problem solving and her EUP median score of 1.7 placed her in the Unstructured Plug-and-Chug category.

Among students classified as using an Unstructured Plug-and-Chug approach, identifying the model and problem type in writing were commonly skipped steps from the Expert Method, but few omitted a physical representation in favor of a mathematical one as the student above did. In fact, it was much more common for students in this category to complete the pictorial and physical representations and to choose an equation but not to solve it. If more than one equation was required, students in this category generally chose only one that was correct; most did not obtain a numeric answer. As a result, students in this category usually showed no evidence of evaluating their solutions, which was an indication that the student applied little to no conceptual understanding while solving the problem.

The previous case was contrasted to a student who performed at a higher level of problem solving. The second video transcript selection was from an interview following Unit 7 (energy model), presented as an example of the Expert-Like approach. The problem asked students to determine the spring constant of a bungee cord used by a stuntman jumping off of a diving board. The student had recognized that he needed to find the spring constant of the cord, written the relationship for elastic potential energy, $E_{elastic} = \frac{1}{2}k\Delta x^2$, calculated the stuntman's gravitational potential energy, and started to write a kinetic energy equation. At this point, he realized he had forgotten to include the fact that the stuntman jumped off the board instead of simply stepping off the board. The student first interpreted the stuntman as jumping down, but then quickly realized that it did not matter whether he jumped up or down.

STUDENT: *So, oh…, wait.*

INTERVIEWER: *So how come you're, like, second-guessing?*

STUDENT: *What?*

INTERVIEWER: *Or what, what changed your mind?*

STUDENT: *Oh, no, I…*

INTERVIEWER: *Oh, nothing? Okay.*

STUDENT: *I just had to add on the, 'cause he's, he's moving at a, he's moving at three meters per second when he first gets off the diving board, so I have to add on the E-k that he gains from jumping. Wait, he is going up (gestures upward and vocally emphasizes it) off the diving board at three meters per second?*

INTERVIEWER: *Um…, you can read it carefully, but sometimes these problems are designed to be ambiguous, so…*

STUDENT: *Oh, well, I guess it wouldn't matter, 'cause if he goes up and then he'd just be going back down again at three meters per second. So, you have to find the extra E-k that he gets for going three meters per second. So…*

INTERVIEWER: *How, how did you know that the, he'd be going three meters per second? Uh…*

STUDENT: *It says he determines that his maximum speed coming off the diving board is three meters per second.*

INTERVIEWER: *Like, how did you determine it doesn't matter if it goes up or down?*

STUDENT: *Because if it goes up three meters per second and then it goes, and then it accelerates down to zero, and then it's going to accelerate with the same acceleration back to three meters per second at the same height, so it doesn't matter if he starts up, if he starts up going three meters per second or he starts down going three meters per second. So, (sotto voce) let me double-check that (checks kinetic energy equation he has begun writing).*

From the fact that the student realized it didn't matter if the person jumped up or down for the inclusion of kinetic energy, it was clear that he applied a strong conceptual understanding of physics. His evaluation of his solution also showed a solid grasp of the spring constant concept. At that point, he had just finished his calculation of the spring constant, determined its units, and written them down.

STUDENT: *So, that's right.*

INTERVIEWER: *How do you know?*

STUDENT: *Um…, well, it's fairly high, because he has to, 'cause he's falling twenty-two point nine meters, and then he has to stop in seven meters, eight meters, so it's going to be fairly high. Well prob… he just… um… and he's starting with the three meters per second velocity, so his velocity's going to get fairly high. And then, all of that has to transfer into E-k over seven-point-six meters, and then… so our coefficient of… and it has to be… Wait… but then it has to be low enough to where it will extend all the way so it doesn't break his leg.*

INTERVIEWER: *So, it's like, it's not too big, it's not too small…*

STUDENT: *Yeah.*

INTERVIEWER: *But just, just where it is. Okay.*

The student's connection between the number he had calculated for the spring constant and its physical interpretation, in the context of how much force would be needed to stop the stuntman's descent in a relatively short distance without being so large as to injure him, showed a deep comprehension of the concept. His quantitative data corroborated the impression that he had a strong conceptual foundation, as his post-FCI score of 30 and MBT score of 19 were well above the threshold of 60% needed for effective problem solving and his EUP median score of 3.0 placed him in the Expert-Like category. In fact, his post-FCI score indicated mastery of basic Newtonian mechanics, and his MBT score was just 7% below the 80% threshold that indicated mastery on that assessment.

This student did not write many steps of the Expert Method on his whiteboard. However, his commentary on his solution and his perceived difficulty of the problem offered some insight.

INTERVIEWER: *Um…, what, uh…, what kind of problem do you think this was?*

STUDENT: *It was an energy problem.*

INTERVIEWER: *Like, how did you recognize it as energy?*

STUDENT: *Because I used energy methods to solve it. And we need to know E, the E elastic.*

INTERVIEWER: *How come?*

STUDENT: *Because it's, uh…, a bungee cord.*

INTERVIEWER: *Oh, okay. Mm-hmm…. Uh… how hard, uh…, do you think this problem was?*

STUDENT: *Um…, it wasn't too hard. It was a little confusing at first, but it kinda, like….*

INTERVIEWER: *What was confusing about it?*

STUDENT: *Um…like, having the three meters per second and… yeah, it wasn't too confusing, but adding the three meters per second was a little, it threw me off a little.*

INTERVIEWER: *Okay.*

STUDENT: *I almost forgot to do it, so….*

INTERVIEWER: *Mm-hmm…. Okay.*

The student was clued into the problem type by an implicit assessment of how he would solve the problem, and he selected the appropriate model that would allow him to do so. Having selected the model, he immediately jumped to the mathematical representation part of the Expert Method, thus he appeared to be using a Structured Plug-and-Chug approach. However, as noted by Walsh et al. (2007), students who could apply an Expert-Like approach would use a Plug-and-Chug approach instead when faced with a problem that lacked sufficient challenge to require a strategic approach; they would conduct a qualitative analysis and then select an appropriate equation to begin their solution. By this student's own admission, this problem was not very challenging for him, so it was presumed that this was why his written work did not show use of the Expert Method. When looking at other factors, the investigators noted that the student was truly using an Expert-Like approach. One such example was the examination of the student's systematic approach in his solution of the problem. The student followed an almost perfectly linear path to the solution and when faced with conceptual obstacles, he applied sound conceptual reasoning to work through them. Finally, the student determined whether his answer was reasonable within the context of the problem by interpreting the conceptual meaning in his solution. Students required a conceptual basis to become effective problem solvers. Student opinion seemed to validate this relationship between conceptual understanding and problem solving ability.

Student Surveys. Students were asked to write a brief comment about how helpful they believed the Expert Method for problem solving was and the survey encouraged them to support their statement with a specific example. Their statements mirror the study's overall findings. The following are some representative statements from students who indicated that they benefited from the Expert Problem Solving Method from student surveys:

> *"Expert method was very effective in helping me problem solve, because one time i did a wrap up problem and it was completely wrong, because i didn't bother to use the expert method, but later that week when I was solving a similar problem i got it right by using the expert method."(sic)*

> *"In some problems, i think that the expert method is very helpful. If there is a bunch of information that you need to sort out in order to solve the problem, then I use the expert method. If the problem isn't very complex then i think it is a pain the use the expert method." (sic)*

> *"Even though it was sometimes a pain to do, I think it helped me a lot. Things would become more clear to me."(sic)*

> *"VERY. Setting out all of the knowns helps tremendously. The space race wrap up problem was solved by me using the expert method."*

> *"It helped when you didn't know where to start. It shows you what you know, and what you need to find out. It sort of relieves some stress from thinking about where to begin.*

> *"I think that the Expert Method is a great tool to use when you have a little grasp on the information at hand, but trying to use it when you have no idea what you are doing will end in failure."*

The last student succinctly summarized the study findings that the Expert Method was most useful for students with a strong conceptual foundation. As the student wrote, without a certain level of understanding, problems could not be solved regardless of the Expert Method. Several students pointed out that the Expert Method involved more work and was not always necessary for simple problems.

It seemed plausible that many of the students with negative opinions of the Expert Method received limited benefit due to their weaker understanding of physics. Other students expressed that the Expert Method was tedious. Below are some representative quotes from student surveys:

> *"The expert method was actually harder than the regular way to solve. If you don't understand the problem how can you make a model and organize the information?"*

> *"Not helpful at all. I felt like it took up more time then necessary and did not help me solve the problem."(sic)*

“I don't think using the expert method sheet was helpful, I focused more on trying to get everything down rather than thinking of a way to solve the problem. I think it did help to look at things though, being able to see something made it easier. But when I know that i have to use the expert method, I try more to put everything down rather than solving the equation.”(sic)

“other than writing out what you are given, Not At ALLL!” (sic)

“not vey helpful. the steps were tedious and most were unnecessary.”(sic)

“The expert method didn't really help me at all. It seemed to me I was just elongating my problem and organizing its different steps into boxes. I did the same thing as I would have even if I hadn't had the expert sheet, I just organizd my steps and gave them names it seems.”(sic)

Overall, both the student comments in favor of and against the Expert Method further support the study’s major contention: When students have the conceptual background, the Expert Method can be a powerful tool. However without a foundation, the Expert Method did not significantly benefit students. Ultimately some of the students built a strong foundation while others did not.

Conclusion

Some research indicated that there were gains in expert-like problem solving when students were taught using Modeling Instruction. Other research indicated gains when students were taught an explicit problem solving method. This study included instruction in an explicit problem solving method even as the teachers were using Modeling Instruction in order to see if the gains from each approach would be additive, providing a way to increase student learning. There was clear evidence that students taught with Modeling Instruction outperformed students taught using traditional educational methods in quantitative problem solving. There was no clear data to indicate that teaching an explicit expert problem solving method, in addition to Modeling Instruction, was better than Modeling Instruction alone; however, there was qualitative support that suggested that the explicit, Expert Method might be beneficial. It was clear that the level of expert-like problem solving was limited by the conceptual understanding of the student; in other words, without understanding concepts, the student did not progress beyond plug-and-chug. Clearly, the most important thing to teach is concepts.

Implications for Further Research

This study was effective in classifying problem solving approaches and demonstrating that a strong conceptual foundation is prerequisite to expert-like problem solving; however, it was not possible to determine whether or not gains from Modeling and explicit emphasis of expert-like problem solving methods augmented the gains in student expertise seen with only one of the methods. Further research in this area would require a sizeable population of treatment and comparison students who could reach a high level of conceptual understanding. It would also be beneficial to explore how the problem solving strategy and the related conceptual understanding might change over time. Lastly, there exists a significant need to understand how students improve their problem solving approach by considering other factors beyond conceptual foundation.

References

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152

Chi, M.T.H., Glaser, R. & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, (Vol. 1)*, Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: verbal reports as data*. Cambridge, MA: MIT Press.

Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping, part 1: group versus individual problem solving. *American Journal of Physics, 60*, 627.

Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics, 55*(5), 440.

Hestenes, D. & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher, 30*(3), 159-166.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*, 141-158.

Jackson, J., Dukerich, L., & Hestenes, D. (2008). Modeling instruction: an effective model for science education. *Science Educator, 17*(1), 10-17.

Knight, R. D. (2004). *Five easy lessons: strategies for successful physics teaching*. San Francisco, CA: Addison Wesley.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335.

Larkin, J., & Reif, F. (1979). Understanding and teaching problem-solving in physics. *European Journal of Science Education, 1*(1), 191-201.

Lederman, E. (2009). Journey into problem solving: a gift from Polya. *The Physics Teacher*, *47*(2), 94.

Malone, K. (2006). *A comparative study of the cognitive and metacognitive differences between modeling and non-modeling high school physics students* (Unpublished doctoral thesis). Carnegie Mellon University, Pittsburgh, PA.

Mestre, J., Dufresne, R., Gerace, W., Hardiman, P., & Touger, J. (1993). Promoting skilled problem-solving behavior among beginning physics students. *Journal of Research in Science Teaching, 30*(3), 303.

Mestre, J. & Touger, J. (1989). Cognitive research–what's in it for physics teachers? *The Physics Teacher, 27*, 447-456.

Polya, George. (1945). *How to solve it*. Princeton, NJ: Princeton University Press.

Priest, A. G., & Lindsay, R. O. (1992). New light on novice-expert differences in physics problem-solving. *British Journal of Psychology, 83*(3), 389.

Reif, F., & Allen, S. (1992). Cognitions for interpreting scientific concepts. *Cognition and Instruction, 9*(1), 1-44.

Reif, F. & Heller, J. I. (1982). Knowledge structure and problem solving in physics. *Educational Psychologist, 17*, 102-127.

Shekoyan, V. (2009). *Using multiple-possibility physics problems in introductory physics courses* (Unpublished doctoral dissertation). Graduate School of New Brunswick, Newark, NJ.

Touger, J. (1964). Open-ended problem instruction in general physics. *The Physics Teacher, 27*(12), 927.

University of Minnesota Physics Department. *Context rich problems online archive*. Retrieved from http://groups.physics.umn.edu/physed/Research/CRP/on-lineArchive/ola.html.

Van Huevelen, A. (1991). Overview: case study physics. *American Journal of Physics, 59*(10), 898-907.

Walsh, L., Howard, R., & Bowe, B. (2007). Phenomenographic study of students' problem solving approaches in physics. *Physical Review Special Topics–Physics Education Research 3.*

Wells, M., Hestenes, D., & Swackhamer, G. (1995). A modeling method for high school physics instruction. *American Journal of Physics, 63*, 606-619.

Wright, D. S. & Williams, C. D. (1986). A WISE strategy for introductory physics. *The Physics Teacher*, *24*(4), 211- 216.

Appendix A: Detailed Statistical Calculations

Table A1

Pre-FCI Descriptive Statistics

Group	n	Mean	Standard Deviation	Standard Error	95% Confidence Interval for Mean		Min
					Lower Bound	Upper Bound	
Fundamental	170	5.86	2.522	.193	5.48	6.24	1
Regular	170	7.15	3.361	.258	6.64	7.66	0
Advanced	69	9.07	4.106	.494	8.09	10.06	0
Total	409	6.94	3.376	.167	6.61	7.26	0

Table A2

Pre-FCI Test of Homogeneity of Variances

Levene Statistic	df 1	df 2	Sig.
6.469	2	406	.002

Table A3

Pre-FCI Robust Tests of Equality of Means

Test	Statistic[a]	df1	df2	Sig.
Welch	21.991	2	167.701	.000
Brown-Forsythe	21.359	2	183.390	.000

a. Asymptotically F distributed.

Table A4

Pre-FCI Multiple Comparisons Tamhane

(I) Group	(J) Group	Mean Difference (I-J)	Standard Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Fundamental	Regular	-1.288*	.322	.000	-2.06	-.51
	Advanced	-3.214*	.531	.000	-4.51	-1.92
Regular	Fundamental	1.288*	.322	.000	.51	2.06
	Advanced	-1.925*	.558	.002	-3.28	-.57
Advanced	Fundamental	3.214*	.531	.000	1.92	4.51
	Regular	1.925*	.558	.002	.57	3.28

*. The mean difference is significant at the 0.05 level.

Table A5

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Standard Error Difference	95% Confidence Interval of the Difference: Lower	95% Confidence Interval of the Difference: Upper
Equal variances assumed	3.676	.059	-1.973	72	.052	-2.575	1.305	-5.177	.027
Equal variances not assumed			-2.866	18.401	.010	-2.575	.899	-4.460	-.690

Table A6

MBT Descriptive Statistics

	n	Mean	Standard Deviation	Standard Error	95% Confidence Interval for Mean: Lower Bound	95% Confidence Interval for Mean: Upper Bound	Min	Max
No Clear Approach	85	6.60	2.290	.248	6.11	7.09	3	13
Unstructured Plug-and-Chug	106	7.03	2.984	.290	6.45	7.60	2	21
Structured Plug-and-Chug	41	8.63	3.080	.481	7.66	9.61	3	15
Expert-Like	28	12.32	4.538	.858	10.56	14.08	4	21
Total	260	7.71	3.457	.214	7.29	8.13	2	21

Table A7

MBT Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
8.216	3	256	.000

Table A8

MBT Robust Tests of Equality of Means

Test	Statistic[a]	df1	df2	Sig.
Welch	16.624	3	82.847	.000
Brown-Forsythe	22.302	3	85.203	.000

a. Asymptotically F distributed.

Table A9

MBT Multiple Comparisons Tamhane

(I) Approach	(J) Approach	Mean Difference (I-J)	Standard Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
No Clear Approach	Unstructured Plug-and-Chug	-.428	.382	.840	-1.44	.59
	Structured Plug-and-Chug	-2.034*	.541	.002	-3.51	-.56
	Expert-Like	-5.721*	.893	.000	-8.23	-3.22
Unstructured Plug-and-Chug	No Clear Approach	.428	.382	.840	-.59	1.44
	Structured Plug-and-Chug	-1.606*	.562	.033	-3.13	-.09
	Expert-Like	-5.293*	.905	.000	-7.82	-2.76
Structured Plug-and-Chug	No Clear Approach	2.034*	.541	.002	.56	3.51
	Unstructured Plug-and-Chug	1.606*	.562	.033	.09	3.13
	Expert-Like	-3.687*	.983	.003	-6.40	-.98
Expert-Like	No Clear Approach	5.721*	.893	.000	3.22	8.23
	Unstructured Plug-and-Chug	5.293*	.905	.000	2.76	7.82
	Structured Plug-and-Chug	3.687*	.983	.003	.98	6.40

*. The mean difference is significant at the 0.05 level.

Table A10

Group Statistics

Test	Treatment	n	Mean	Standard Deviation	Standard Error Mean
Post-E	Regular Comparison	22	30.80	6.921	1.476
	Regular Treatment	34	34.77	8.921	1.530
Post-N	Regular Comparison	22	10.04	4.158	.887
	Regular Treatment	34	10.21	4.883	.837
Post-QA	Regular Comparison	22	38.98	20.257	4.319
	Regular Treatment	34	25.97	15.436	2.647

Table A11

Independent Samples Test Statistics

Samples		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Differ-ence	Standard Error Differ-ence	95% Confidence Interval of the Difference	
									Lower	Upper
Post-E	Equal variances assumed	1.208	.277	-1.769	54	.083	-3.969	2.244	-8.468	.530
	Equal variances not assumed			-1.867	52.104	.067	-3.969	2.126	-8.234	.296
Post-N	Equal variances assumed	.000	.998	-.131	54	.897	-.165	1.263	-2.696	2.366
	Equal variances not assumed			-.135	49.907	.893	-.165	1.219	-2.615	2.285
Post-QA	Equal variances assumed	7.936	.007	2.721	54	.009	13.008	4.780	3.425	22.591
	Equal variances not assumed			2.568	36.467	.014	13.008	5.066	2.739	23.277

Table A12

Group Statistics

	Modeling	n	Mean	Standard Deviation	Standard Error Mean
E-gain	Traditional	74	6.90	7.212	.838
	Modeling	313	4.62	9.800	.554
Q-gain	Traditional	74	22.82	23.980	2.788
	Modeling	311	-4.30	29.198	1.656
N-gain	Traditional	73	-1.30	5.543	.649
	Modeling	313	.16	11.661	.659

Table A13

Independent Samples Test

Samples		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Differ-ence	Standard Error Differ-ence	95% Confidence Interval of the Difference Lower	Upper
E-gain	Equal variances assumed	4.053	.045	1.885	385	.060	2.281	1.210	-.099	4.661
	Equal variances not assumed			2.270	144.219	.025	2.281	1.005	.295	4.267
Q-gain	Equal variances assumed	2.090	.149	7.416	383	.000	27.124	3.657	19.933	34.315
	Equal variances not assumed			8.366	129.783	.000	27.124	3.242	20.710	33.539
N-gain	Equal variances assumed	4.220	.041	-1.037	384	.300	-1.454	1.401	-4.209	1.302
	Equal variances not assumed			-1.572	238.696	.117	-1.454	.925	-3.275	.368

Table A14

Group Statistics

	Treatment	n	Mean	Standard Deviation	Standard Error Mean
Post-FCI	Regular Comparison	95	10.02	3.479	.357
	Regular Treatment	65	14.52	5.348	.663

Table A15

Independent Samples Test

Samples		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Differ-ence	Standard Error Differ-ence	95% Confidence Interval of the Difference Lower	Upper
Post-FCI	Equal variances assumed	12.621	.001	-6.453	158	.000	-4.502	.698	-5.880	-3.124
	Equal variances not assumed			-5.977	100.692	.000	-4.502	.753	-5.996	-3.008

Table A16

Group Statistics

	Treatment	n	Mean	Standard Deviation	Standard Error Mean
Post-FCI	Advanced Comparison	31	18.61	4.379	.787
	Advanced Treatment	34	21.59	5.217	.895

Table A17

Independent Samples Test

Samples		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Differ-ence	Standard Error Differ-ence	95% Confidence Interval of the Difference	
									Lower	Upper
Post-FCI	Equal variances assumed	3.132	.082	-2.477	63	.016	-2.975	1.201	-5.375	-.575
	Equal variances not assumed			-2.498	62.593	.015	-2.975	1.191	-5.356	-.594

Table A18

Group Statistics

	Treatment	n	Mean	Standard Deviation	Standard Error Mean
FCI Hake Gain	Comparison	126	16.7990	17.33560	1.54438
	Treatment	232	19.3247	16.68755	1.09559

Table A19

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Differ-ence	Standard Error Differ-ence	95% Confidence Interval of the Difference: Lower	95% Confidence Interval of the Difference: Upper
FCI Hake Gain	Equal variances assumed	.116	.733	-1.349	356	.178	-2.52567	1.87223	-6.20769	1.15636
	Equal variances not assumed			-1.334	248.43	.183	-2.52567	1.89352	-6.25507	1.20374

Table A20

Group Statistics

	Treatment	n	Mean	Standard Deviation	Standard Error Mean
FCI Hake Gain	Traditional	95	11.6841	14.46343	1.48392
	Modeling	263	20.8746	17.12717	1.05611

Table A21

Independent Samples Test

Samples		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Differ-ence	Standard Error Differ-ence	95% Confidence Interval of the Difference	
									Lower	Upper
FCI Hake Gain	Equal variances assumed	7.019	.008	-4.663	356	.000	-9.19046	1.97098	-13.0667	-5.3142
	Equal variances not assumed			-5.046	195.36	.000	-9.19046	1.82136	-12.7825	-5.5984

Table A22

Post-FCI Descriptive Statistics

	n	Mean	Standard Deviation	Standard Error	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
No Clear Approach	81	9.75	3.763	.418	8.92	10.59	4	21
Unstructured Plug-and-Chug	103	11.54	5.416	.534	10.49	12.60	2	28
Structured Plug-and-Chug	41	13.32	5.623	.878	11.54	15.09	4	28
Expert-Like	27	20.78	7.723	1.486	17.72	23.83	3	30
Total	252	12.25	6.154	.388	11.48	13.01	2	30

Table A23

Post-FCI Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
9.436	3	248	.000

Table A24

Post-FCI Robust Tests of Equality of Means

Test	Statistic[a]	df1	df2	Sig.
Welch	19.584	3	80.379	.000
Brown-Forsythe	27.059	3	87.080	.000

a. Asymptotically F distributed.

Table A25

Post-FCI Multiple Comparisons Tamhane

(I) Approach	(J) Approach	Mean Difference (I-J)	Standard Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
No Clear Approach	Unstructured Plug-and-Chug	-1.791	.678	.053	-3.59	.01
	Structured Plug-and-Chug	-3.564*	.973	.003	-6.21	-.92
	Expert-Like	-11.025*	1.544	.000	-15.37	-6.68
Unstructured Plug-and-Chug	No Clear Approach	1.791	.678	.053	-.01	3.59
	Structured Plug-and-Chug	-1.773	1.028	.427	-4.55	1.01
	Expert-Like	-9.234*	1.579	.000	-13.65	-4.82
Structured Plug-and-Chug	No Clear Approach	3.564*	.973	.003	.92	6.21
	Unstructured Plug-and-Chug	1.773	1.028	.427	-1.01	4.55
	Expert-Like	-7.461*	1.726	.001	-12.22	-2.70
Expert-Like	No Clear Approach	11.025*	1.544	.000	6.68	15.37
	Unstructured Plug-and-Chug	9.234*	1.579	.000	4.82	13.65
	Structured Plug-and-Chug	7.461*	1.726	.001	2.70	12.22

*. The mean difference is significant at the 0.05 level.

Appendix B: Rubric, End of Unit Problems, and Student Examples

End of Unit Problem Rubric

EUPs were scored using a flow chart loosely based on the approaches recognized by Walsh et al. (2007). Two investigators scored each problem. Although rare, disagreements between scores were resolved through discussion by the investigators.

Structure

1 – Solution can be followed.

0 – Solution work cannot be followed. Random numbers or arithmetic are present.

Equation

1 – At least one equation is written in the solution.

0 – No equations are written in the solution.

Correct Equations

1 – All equations are correct and fully complete.

0 – One equation is missing, incomplete, or incorrect.

Erasures

1 – Erasures or crossed out work is not present.

0 – Erasures or crossed out work is present.

Reasonable Answer

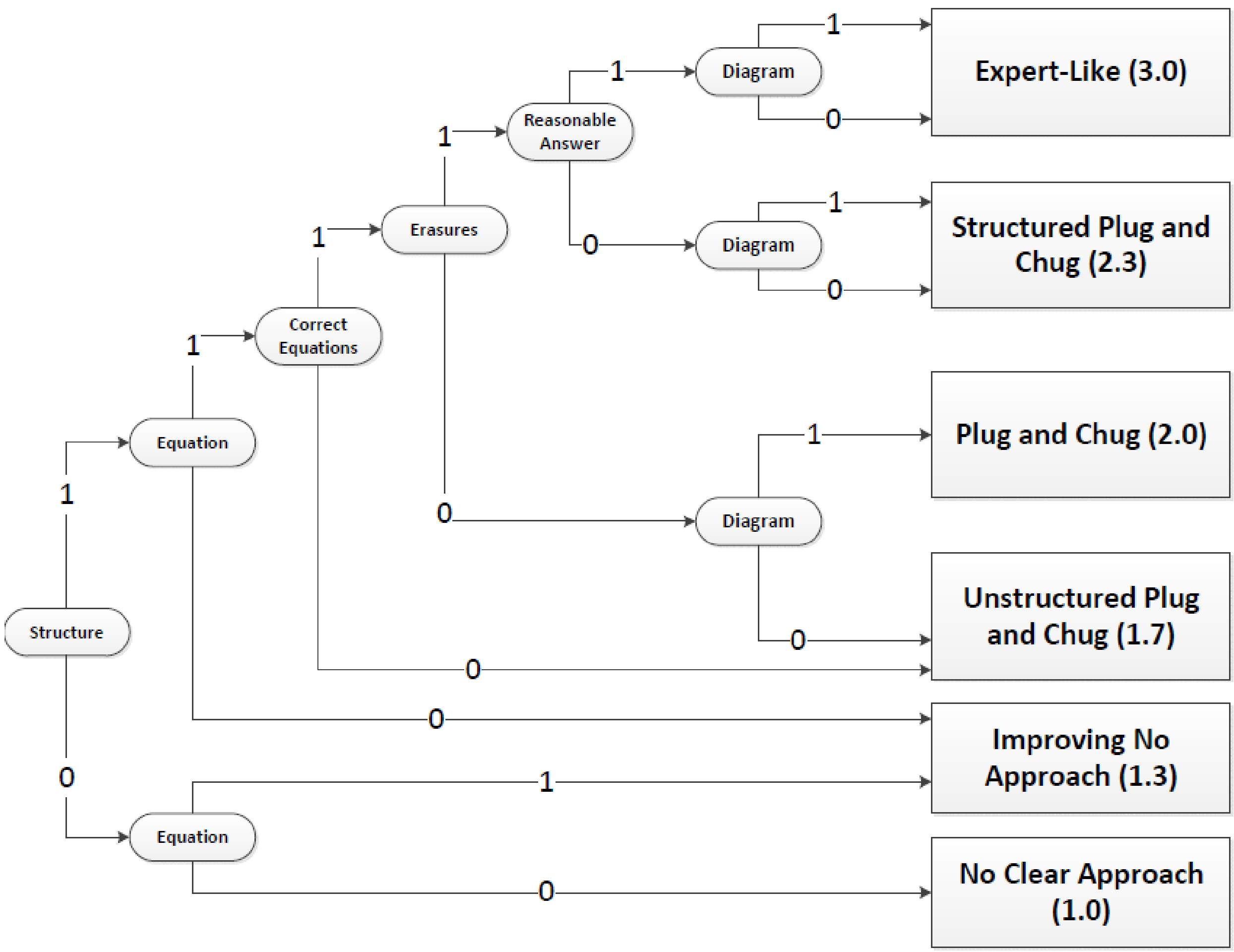
1 – Answer is close to the correct one. Units are appropriate. Answer is interpreted correctly.

0 – Answer is far from the correct one, or units are incorrect, or answer is interpreted incorrectly.

Diagrams

1 – Diagrams are incorporated in solution method.

0 – No diagrams are incorporated in solution method.



End of Unit Problems (EUPs)

Each of the EUPs was given at the conclusion of a unit. All are content rich problems from the University of Minnesota PER group (University of Minnesota Physics Department).

Unit 2

It's a sunny Sunday afternoon, about 65 °F, and you are walking around Lake Calhoun enjoying the last of the autumn color. The sidewalk is crowded with runners and walkers. You notice a runner approaching you wearing a tee-shirt with writing on it. You read the first two lines, but are unable to read the third and final line before he passes. You wonder, "Hmm, if he continues around the lake, I bet I'll see him again, but I should anticipate the time when we'll pass again." You look at your watch and it is 3:07 p.m. You recall the lake is 3.4 miles in circumference. You estimate your walking speed at 3 miles per hour and the runner's speed to be about 7 miles per hour.

Unit 3

Because parents are concerned that children are learning "wrong" science from TV, you have been asked to be a technical advisor for a science fiction cartoon show on Saturday morning. In the plot, a vicious criminal (Natasha Nogood) escapes from a space station prison. The prison is located between galaxies far away from any stars. Natasha steals a small space ship and blasts off to meet her partners somewhere in deep space. The stolen ship accelerates in a straight line at its maximum possible acceleration of 30 m/sec2. After 10 minutes all of the fuel is burned up and the ship coasts at a constant velocity. Meanwhile, the hero (Captain Starr) learns of the escape while dining in the prison with the warden's daughter (Virginia Lovely). Of course he immediately (as soon as he finishes dessert) rushes off the recapture Natasha. He gives chase in an identical ship, which has an identical maximum acceleration, going in an identical direction. Unfortunately, Natasha has a 30 minute head start. Luckily, Natasha's ship did not start with a full load of fuel. With his full load of fuel, Captain Starr can maintain maximum acceleration for 15 minutes. How long will it take Captain Starr's ship to catch up to Natasha's?

Unit 4

You are taking advantage of an early snow to go sledding. After a long afternoon of going up and down hills with your sled, you decide it is time to go home. You are thankful that you can pull your sled without climbing any more hills. As you are walking home, dragging the sled behind you by a rope fastened to the front of the sled, you wonder what the coefficient of friction of the snow on the sled is. You estimate that you are pulling on the rope with a 2 pound force, that the sled weighs 10 pounds, and the rope is taught and level to the ground.

Unit 5

You have always been impressed by the speed of the elevators in the IDS building in Minneapolis (especially compared to the one in the physics building). You wonder about the maximum acceleration for these elevators during normal operation, so you decide to measure it by using your bathroom scale. While the elevator is at rest on the ground floor, you get in, put down your scale, and stand on it. The scale reads 130 lbs. You continue standing on the scale when the elevator goes up, carefully watching the reading. During the trip to the 50th floor, the greatest scale reading was 180 lbs.

Unit 6

The Police Department has hired you as a consultant in a robbery investigation. A thief allegedly robbed a bank and, to escape the pursing security guards, took the express elevator to the roof of the building. Then, in order to not be caught with the evidence, the thief allegedly threw the money bag to a waiting accomplice on the roof of the next building. The defense attorney contends that in order to reach the roof of that next building, the defendant would have had to throw the money bag horizontally with a minimum velocity of 10 meters/second. However, in a test, the accused could throw the bag with a maximum horizontal velocity of no more than 5 meters/second. How will you advise the prosecuting attorney? You determine that the bank building is 250 meters high, the next building is 100 meters high and the distance between them is 20 meters.

Unit 7

In a weak moment you have volunteered to be a human cannonball at an amateur charity circus. The "cannon" is actually a 3-foot diameter tube with a big stiff spring inside which is attached to the bottom of the tube. A small seat is attached to the free end of the spring. The ringmaster, one of your soon to be ex-friends, gives you your instructions. He tells you that just before you enter the mouth of the cannon, a motor will compress the spring to 1/10 its normal length and hold it in that position. You are to gracefully crawl in the tube and sit calmly in the seat without holding on to anything. The cannon will then be raised to an angle such that your speed through the air at your highest point is 10 ft/s. When the spring is released, neither the spring nor the chair will touch the sides of the 12-foot long tube. After the drum roll, the spring is released and you will fly through the air with the appropriate sound effects and smoke. With the perfect aim of your gun crew, you will fly through the air over a 15-foot wall and land safely in the net. You are just a bit worried and decide to calculate how high above your starting position you will be at your highest point. Before the rehearsal, the cannon is taken apart for maintenance. You see the spring, which is now removed from the cannon, is hanging straight down with one end attached to the ceiling. You determine that it is 10 feet long. When you hang on its free end without touching the ground, it stretches by 2.0 feet. Is it possible for you to make it over the wall?

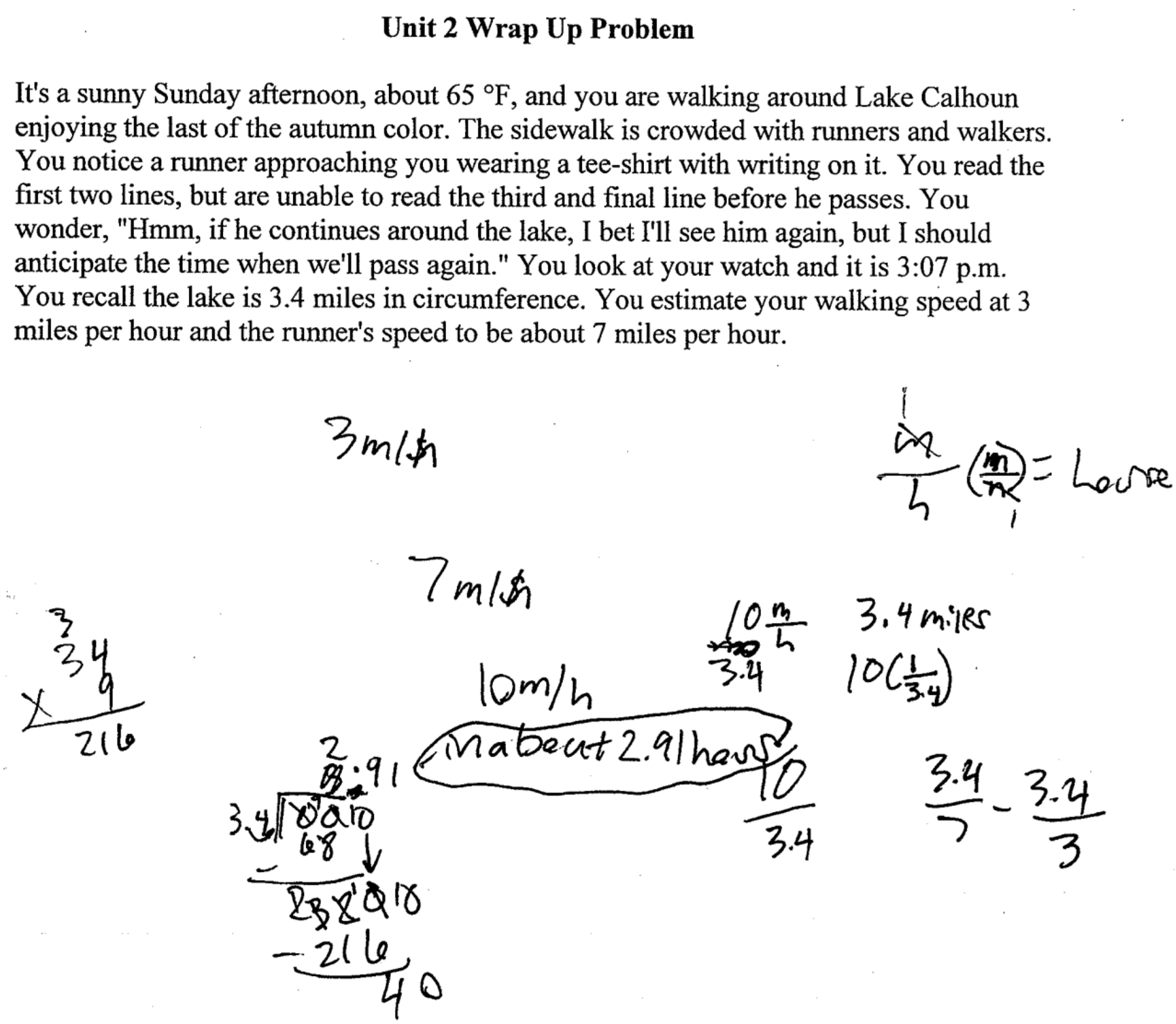
Unit 8

A hammer thrower in the Olympics uses a 40 N force to swing an 18 kg hammer in a circle radius of 1.2 m. Does she break the work record distance of 50 m if she lets go at an angle of 30 degrees to the horizontal?

Unit 9

You have been hired to check the technical correctness of an upcoming made-for-TV murder mystery that takes place in the space shuttle. In one scene, an astronaut's safety line is cut while on a space walk. The astronaut, who is 200 meters from the shuttle and not moving with respect to it, finds that the suit's thruster pack has also been damaged and no longer works and that there is only 4 minutes of air remaining. To get back to the shuttle, the astronaut unstraps a 10-kg tool kit and throws it away with a speed of 8 m/s. In the script, the astronaut, who has a mass of 80 kg without the toolkit, survives, but is this correct?

Student Examples



Unit 2 Wrap Up Problem

It's a sunny Sunday afternoon, about 65 °F, and you are walking around Lake Calhoun enjoying the last of the autumn color. The sidewalk is crowded with runners and walkers. You notice a runner approaching you wearing a tee-shirt with writing on it. You read the first two lines, but are unable to read the third and final line before he passes. You wonder, "Hmm, if he continues around the lake, I bet I'll see him again, but I should anticipate the time when we'll pass again." You look at your watch and it is 3:07 p.m. You recall the lake is 3.4 miles in circumference. You estimate your walking speed at 3 miles per hour and the runner's speed to be about 7 miles per hour.

Figure B1. Example of No Clear Approach Unit 2 EUP.

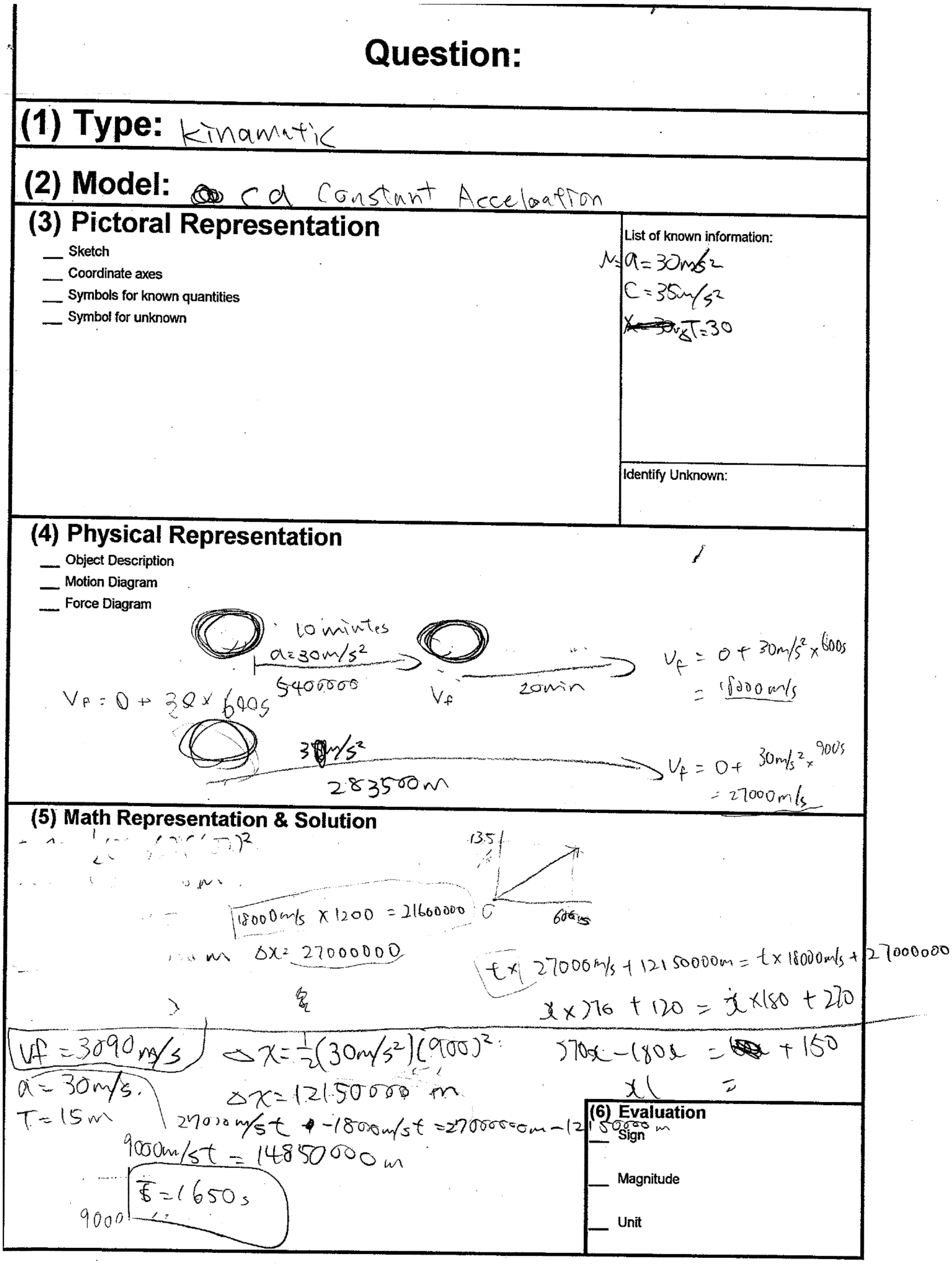
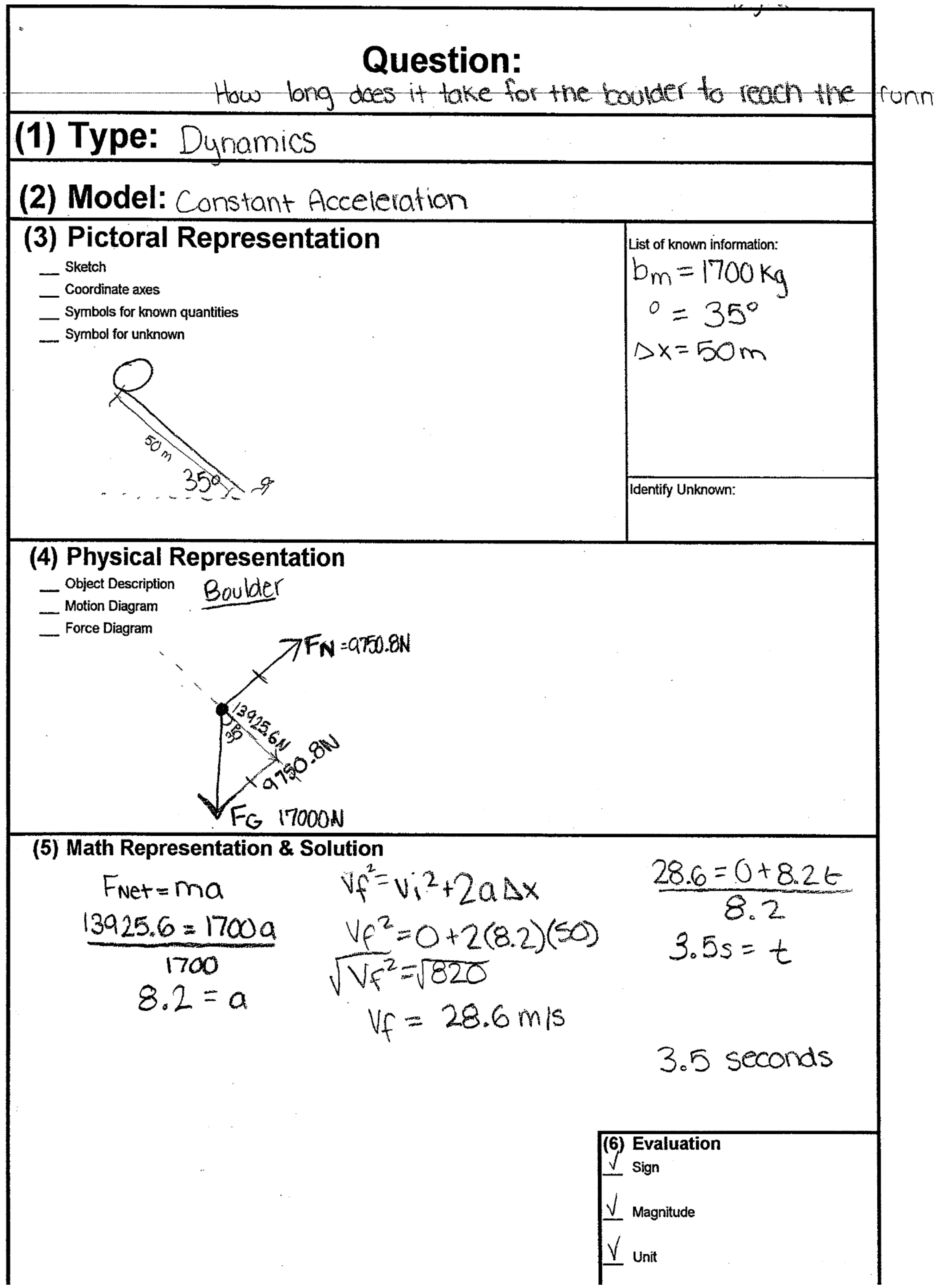


Figure B2. Example of Unit 3 Unstructured Plug and Chug EUP.



Question:

How long does it take for the boulder to reach the ~~form~~

(1) Type: Dynamics

(2) Model: Constant Acceleration

(3) Pictoral Representation

__ Sketch
__ Coordinate axes
__ Symbols for known quantities
__ Symbol for unknown

List of known information:

$b_m = 1700$ Kg
$\theta = 35°$
$\Delta x = 50$ m

Identify Unknown:

(4) Physical Representation

__ Object Description Boulder
__ Motion Diagram
__ Force Diagram

(5) Math Representation & Solution

$F_{Net} = ma$
$\frac{13925.6 = 1700a}{1700}$
$8.2 = a$

$V_f^2 = V_i^2 + 2a\Delta x$
$V_f^2 = 0 + 2(8.2)(50)$
$\sqrt{V_f^2} = \sqrt{820}$
$V_f = 28.6$ m/s

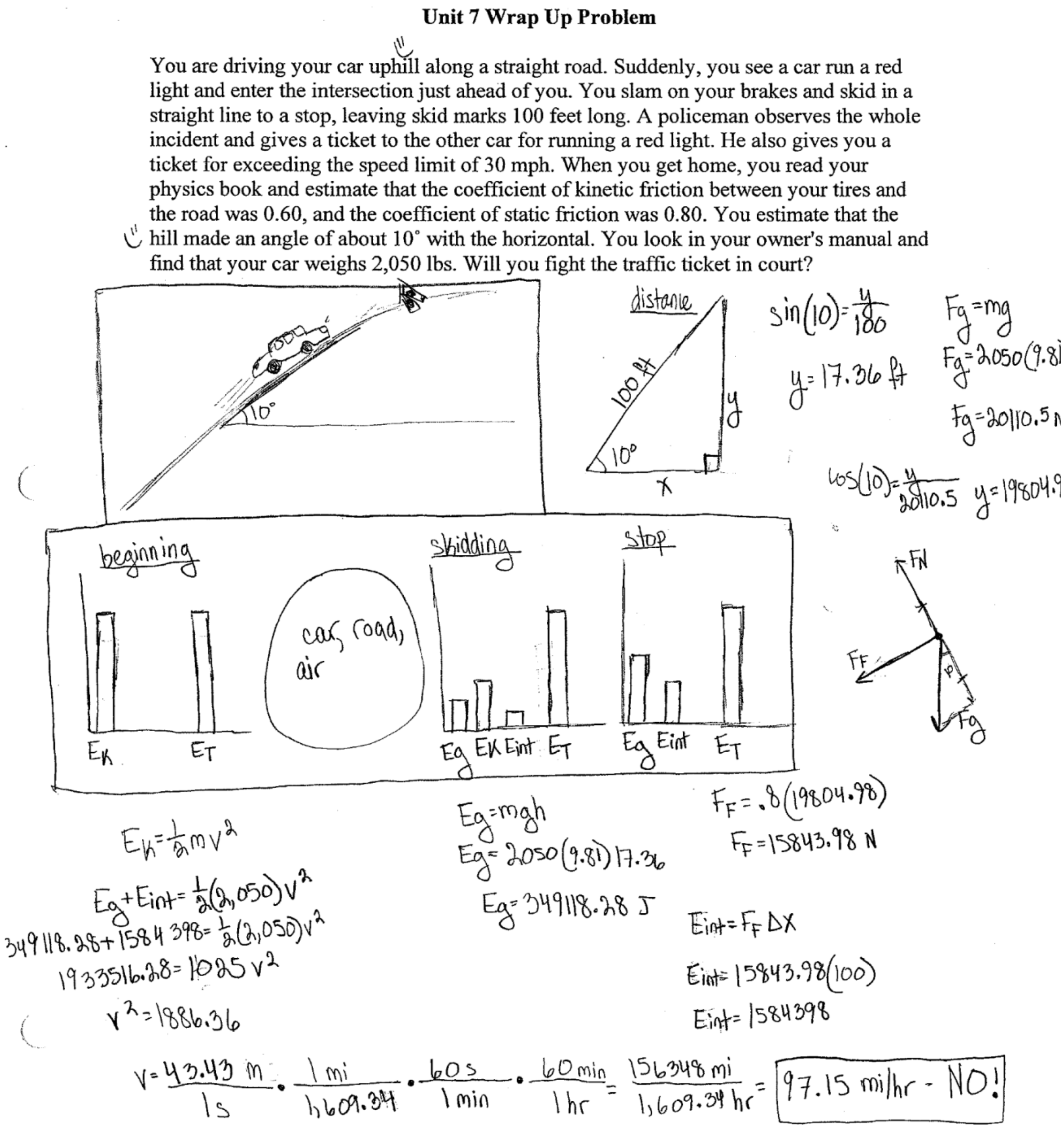
$\frac{28.6 = 0 + 8.2t}{8.2}$
$3.5s = t$

3.5 seconds

(6) Evaluation

√ Sign
√ Magnitude
√ Unit

Figure B3. Example of Structured Plug and Chug Unit 5 EUP.



Unit 7 Wrap Up Problem

You are driving your car uphill along a straight road. Suddenly, you see a car run a red light and enter the intersection just ahead of you. You slam on your brakes and skid in a straight line to a stop, leaving skid marks 100 feet long. A policeman observes the whole incident and gives a ticket to the other car for running a red light. He also gives you a ticket for exceeding the speed limit of 30 mph. When you get home, you read your physics book and estimate that the coefficient of kinetic friction between your tires and the road was 0.60, and the coefficient of static friction was 0.80. You estimate that the hill made an angle of about 10° with the horizontal. You look in your owner's manual and find that your car weighs 2,050 lbs. Will you fight the traffic ticket in court?

Figure B4. Example of Expert-Like Unit 7 EUP.

Appendix C: Interview Problems

Interview Problems

These interview problems were taken from the University of Minnesota PER group website (University of Minnesota Physics Department) containing content rich problems.

Unit 3

In your new job, you are the technical advisor for the writers of a gangster movie about Bonnie and Clyde. In one scene Bonnie and Clyde try to flee from one state to another. (If they got across the state line, they could evade capture, at least for a while until they became Federal fugitives.) In the script, Bonnie is driving down the highway at 108 km/hour, and passes a concealed police car that is 1 kilometer from the state line. The instant Bonnie and Clyde pass the patrol car, the cop pulls onto the highway and accelerates at a constant rate of 2 m/s^2. The writers want to know if they make it across the state line before the pursuing cop catches up with them.

Unit 5

In an effort to catch the elusive Road Runner, Wylie Coyote hoists 1700 kg boulder to the top of a 35 degree slope. He then secures the boulder with a twig and waits in hiding to release the trap on the bird 50 m down the slope. How long does the Road Runner, who is busily eating food below, have to react and move out of the way before being crushed?

Unit 7

Super Dave has just returned from the hospital where he spent a week convalescing from injuries incurred when he was "shot" out of a cannon to land in an airbag which was too thin. Undaunted, he decides to celebrate his return with a new stunt. He intends to jump off a 100-

foot tall tower with an elastic cord tied to one ankle, and the other end tied to the top of the tower. This cord is very light but very strong and stretches so that it can stop him without pulling his leg off. Such a cord exerts a force with the same mathematical form as the spring force. He wants it to be 75 feet long so that he will be in free fall for 75 feet before the cord begins to stretch. To minimize the force that the cord exerts on his leg, he wants it to stretch as far as possible. You have been assigned to purchase the cord for the stunt and must determine the elastic force constant which characterizes the cord that you should order. Before the calculation, you carefully measure Dave's height to be 6.0 ft and his weight to be 170 lbs. For maximum dramatic effect, his jump will be off a diving board at the top of the tower. From tests you have made, you determine that his maximum speed coming off the diving board is 10 ft/s. Neglect air resistance in your calculation -- let Dave worry about that.

Unit 9

You are helping your friend prepare for the next skateboard exhibition by determining if the planned program will work. Your friend will take a running start and then jump onto a heavy-duty 15-lb stationary skateboard. The skateboard will glide in a straight line along a short, level section of track, then up a sloped concrete wall. The goal is to reach a height of at least 10 feet above the starting point before coming back down the slope. Your friend's maximum running speed to safely jump on the skateboard is 23 feet/second. Your friend weighs 150 lbs.

Appendix D: Card Sorting Task (CST)

Card Sorting Task (CST)

The investigators followed Malone's (2006) procedures for assigning numeric scores to the CST. Each of the 58 problems had a "real" number that denoted its model and surface feature, in the format m.sf such that the tens digit (m) indicated the model and the tenths and hundredths digits (sf) indicated the surface feature. The problems given to the students were randomly assigned a number from 1 through 58 so that the students would not be influenced to group them a certain way based on the problem number (like putting all the 1.##s together or putting all the #.01s together).

The investigators set up a confusion matrix for each of the three scores they wished to obtain: Expert, Novice, and Question Asked. Each matrix listed the problems down the left side and across the top, grouped by the sorting scheme being tested. That is, the Expert matrix had all the 1.##s grouped together, followed by the 2.##s, and so on, while the Novice matrix had all the #.01s grouped together, followed by the #.02s, and so on. If all the 1.##s were constant velocity questions, while all the #.01s were free fall questions, it was expected that an Expert would group all the constant velocity questions together, while a Novice would group all the free fall questions together. Given that the Question Asked groups were a mix of models and surface features, that organization was less obvious from reading the groups of problem numbers.

An "X" was entered into each column under a number in the row for each problem that had been grouped with that number. The investigators did not want to count a problem as being grouped with itself, so the boxes were shaded as a visual reminder not to enter data into those cells that were arranged along the diagonal of the matrix and thus showed where the column and

	1.01	1.02	1.03	2.01	2.02	2.03	3.01	3.02	3.03
1.01									
1.02									
1.03									
2.01									
2.02									
2.03									
3.01									
3.02									
3.03									

Figure D1. Example of an Expert scoring matrix.

row corresponded to the same problem number. The darker borders ("big boxes") in the matrix delineated all the cells for an expected grouping for a given problem. A purely expert sorting based strictly on physical model yielded big boxes full of X's and no X's anywhere else on the Expert scoring matrix. A numeric score was determined by counting the total number of X's inside the big boxes, dividing by the total number of X's entered in the matrix, and then multiplying by 100%; that is, the Expert score showed the total number of problems that had been grouped by model divided by the total number of problems sorted, reported as a percentage. A purely expert sorting, with all X's in big boxes, yielded an Expert score (E) of 100.

$$E = \frac{\text{number of X's in big boxes}}{\text{total number of X's in matrix}} \times 100\%$$

Figure D2 is an example of what a purely Expert sorting looked like on the Expert scoring matrix.

The purely expert sorting looked very different on the Novice matrix, however. Instead of all the X's falling in boxes, the X's were placed on diagonals and there were no X's in boxes. Determining the Novice score for this sorting was done in the same way as determining the Expert score. Counting the X's in big boxes, dividing by the total number of X's, and multiplying by 100% yielded a Novice score of 0 for a purely expert sorting based on physical model.

$$N = \frac{\text{number of X's in big boxes}}{\text{total number of X's in matrix}} \times 100\%$$

The converses of these two examples were also true. A purely novice sorting based strictly on surface feature entered into the Expert scoring matrix had all of the X's on diagonals

	1.01	1.02	1.03	2.01	2.02	2.03	3.01	3.02	3.03
1.01		X	X						
1.02	X		X						
1.03	X	X							
2.01					X	X			
2.02				X		X			
2.03				X	X				
3.01								X	X
3.02							X		X
3.03							X	X	

Figure D2. Example of a purely Expert sorting on the Expert matrix.

and no X's in the big boxes, yielding an Expert score of 0. A purely novice sorting entered into the Novice scoring matrix had all X's in the big boxes and no X's anywhere else in the matrix, yielding a Novice score of 100%.

Of course, the actual student data was not this clear-cut. The students used a mix of the three sorting schemes. The Question-Asked score (QA) was calculated in the same way as the Expert and Novice scores, but the big boxes that corresponded to the expected groupings for that scheme contained different problem numbers than the expected groupings for Expert and Novice sorting. Malone (2006) noted that the Question Asked groupings overlapped to some degree with both the expert groupings and the novice groupings; that is, within a given model or surface feature group, a number of the problems asked about the same variable. Thus, a student who

	1.01	2.01	3.01	1.02	2.02	3.02	1.03	2.03	3.03
1.01				X			X		
2.01					X			X	
3.01						X			X
1.02	X						X		
2.02		X						X	
3.02			X						X
1.03	X			X					
2.03		X			X				
3.03			X			X			

Figure D3. Example of a completed Novice matrix for purely an Expert sorting used a strictly Question Asked sorting had a Question Asked score of 100 and still had non-zero Expert and Novice scores. Malone (2006) also pointed out that the Question Asked sorting was a somewhat more sophisticated scheme than the novice sorting scheme, but was still more Novice-Like than Expert-Like because it was based on a literal feature of the problem.