REFLECTION ON PROBLEM SOLVING IN INTRODUCTORY AND ADVANCED PHYSICS

by

Andrew J. Mason

B.S. Physics, State University of New York at Buffalo, 2004

M.S. Physics and Astronomy, University of Pittsburgh, 2005

Submitted to the Graduate Faculty of

the Department of Physics and Astronomy in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

DEPARTMENT OF PHYSICS AND ASTRONOMY

This dissertation was presented

by

Andrew J. Mason

It was defended on

July 7, 2009

and approved by

Dissertation Advisor: Dr. Chandralekha Singh, Associate Professor, Department of Physics and Astronomy

Dr. Peter F. M. Koehler, Professor, Department of Physics and Astronomy

Dr. Adam Leibovich, Associate Professor, Department of Physics and Astronomy

Dr. Robert P. Devaty, Associate Professor, Department of Physics and Astronomy

Dr. Christian Schunn, Associate Professor, Department of Psychology

REFLECTION ON PROBLEM SOLVING IN INTRODUCTORY AND ADVANCED PHYSICS

Andrew J. Mason, PhD

University of Pittsburgh, 2009

Reflection is essential in order to learn from problem solving. This thesis explores issues related to how reflective students are and how we can improve their capacity for reflection on problem solving. We investigate how students naturally reflect in their physics courses about problem solving and evaluate strategies that may teach them reflection as an integral component of problem-solving. Problem categorization based upon similarity of solution is a strategy to help them reflect about the deep features of the problems related to the physics principles involved. We find that there is a large overlap between the introductory and graduate students in their ability to categorize. Moreover, introductory students in the calculus-based courses performed better categorization than those in the algebra-based courses even though the categorization task is conceptual. Other investigations involved exploring if reflection could be taught as a skill on individual and group levels. Explicit self-diagnosis in recitation investigated how effectively students could diagnose their own errors on difficult problems, how much scaffolding was necessary for this purpose, and how effective transfer was to other problems employing similar principles. Difficulty in applying physical principles and difference between the self-diagnosed and transfer problems affected performance. We concluded that a sustained intervention is required to learn effective problem-solving strategies. Another study involving reflection on problem solving with peers suggests that those who reflected with peers drew more diagrams and had a larger gain from the midterm to final exam. Another study in quantum mechanics involved giving common problems in midterm and final exams and suggested that advanced students do not automatically reflect on their mistakes. Interviews revealed that even advanced students often focus mostly on exams rather than learning and building a robust knowledge structure. A survey was developed to further evaluate students' attitudes and approaches towards problem solving. The survey responses suggest that introductory students and even graduate students have different attitudes and approaches to problem solving on several important measures compared to physics faculty members. Furthermore, responses to individual survey questions suggest that expert and novice attitudes and approaches to problem solving may be more complex than naively considered.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE - ACKNOWLEDGMENTS

Here I acknowledge the following individuals for their help and support over these past several years.

First of all, I would like to thank my adviser, Dr. Chandralekha Singh, for all her help and support as my adviser. I would also like to thank the members of the committee: Dr. Peter Koehler, Dr. Adam Leibovich, Dr. Robert P. Devaty, and Dr. Chris Schunn for their insight and critiques as members of my committee.

In addition, I would like to thank Dr. Edit Yerushalmi and Dr. Elisheva Cohen for their substantial advice, ideas, and contributions as collaborators in the self-diagnosis study.

I would also like to thank the faculty of the University of Pittsburgh for allowing Dr. Singh and me to work with their classes as part of my projects, as well as for their advice on several matters. In addition, I would like to thank Dr. Jeremy Levy for his insight and enthusiasm for the peer reflection study and discussions. More additional thanks to Dr. Devaty for his aid and suggestions in reading over some of our papers and also checking for errors.

I am grateful to the many faculty and students who participated in all of the studies in this work. It is regretfully impossible to name everyone as so many people participated.

I am happy to thank my parents, Janice and Joseph Mason, and my sister, Maryrose Mason, for all their love and support through all these years of study. I am incredibly grateful to them for encouraging and advising me through the thick and thin of it all.

1.0 INTRODUCTION

Physics education research (PER) focuses on improving the teaching and learning of physics. Physics education researchers have content knowledge and access to physics students, both of which provide advantages to studying domain-specific learning. Furthermore, most benefits from PER studies are accrued to physics departments (e.g. McDermott and Redish 1999).

Cognitive research suggests that we cannot pour knowledge into students' brains. Students must be actively engaged in the learning process for learning to be meaningful. A generic model of learning is stated as follows. The learner begins learning with an initial knowledge state, and through the process of learning will reach a final knowledge state (Reif 1986). The purpose of instruction, therefore, is to guide the progress of learning from the initial knowledge state to a desired final knowledge state. Learning is meaningful only if the instruction takes into account the learner's prior knowledge and builds on it. One way to measure the learner's initial and final knowledge states is to analyze student performances on assessment activities both before and after instruction. Of course, the performance on these activities will reflect whether the goals of instruction have been achieved only if the assessment activities are carefully designed commensurate with the goals of the course. In order to evaluate the effectiveness of instruction, it is indeed important to be able to consistently measure and analyze the learner's gain (e.g., Hake 1998).

Two examples of how an education researcher would examine knowledge gain are the pretest-posttest approach and interviews with students. Pretests and posttests are tests given to examine the knowledge of students both before and after an instruction sequence, e.g. the Force Concept Inventory test (Hestenes et al. 1992; Hestenes 1995) can be given to students before and after instruction of force concepts to measure learning gain without fear of biasing the posttest scores by exposure to the pretest (Henderson 2002). In contrast to the quantitative nature of pretest-posttest data, interviews with individual students are more qualitative but offer more in-depth insight into the students' problem-solving approaches. Surveys given at the beginning and end of a course may also work to this end, e.g., the MPEX survey (Redish et al. 1998) used to evaluate changes in students' epistemology about physics learning.

The field of physics education research attempts to improve student learning by developing and evaluating strategies that build on the initial knowledge of students. Different concentrations within the field of PER are based around different aspects of this goal. Examples include (but are certainly not limited to) the following: pedagogical approaches to instruction, assessment of students' content knowledge, understanding epistemological beliefs of students about physics, and so forth.

1.1 PROBLEM SOLVING: DEFINITION

The concentration within PER that pertains directly to the main topic of this thesis is that of problem solving. Problem solving can be defined as any purposeful activity where one is presented with a novel situation and one must devise strategies to reach the goal (Larkin and Reif 1979, Reif 1981, Eylon and Reif 1984, Heller and Reif 1984, Reif 1986, Reif and Larkin 1991, Reif and Allen 1992, Reif 1995). A physics problem can be primarily quantitative or conceptual in nature. The understanding of how to solve a problem is an integral part of moving from the initial state of learning to the desired final state of learning.

In addition to the definite interest that physics departments would have in stressing the value of problem solving in physics to their students, physics problem solving is also considered important from the perspective of the fields of cognitive science and science education. The reason cognitive scientists often use physics problems to study cognitive issues and problem solving in general is because physics problems are very well-defined and have distinct solutions. Thus, other fields, which share some of the same objectives as physics education researchers in pursuing research on problem solving, may also have some different motivations (McDermott, 1990).

1.2 INFLUENCE FROM COGNITIVE SCIENCE: THE CONCEPT OF MEMORY

Problem solving is a cognitive process, and as such, much work has been done in the realm of cognitive science on the general topic of problem solving. This work has run parallel to physics education research and is an important source of defining theory used as background in PER today.

Human memory, i.e., the human information processing system, has two broad components at a coarse-grained level (Simon 1974). Long-term memory (LTM) is where prior learned knowledge is stored and can be drawn upon in future problem solving. Short-term memory (STM), or working memory, is where information is processed. While solving problems, the STM receives input from the sensory buffers (eyes, hands, ears, etc.) and from the

LTM. Both components of memory are critical for conscious problem solving; STM is used to rearrange and synthesize ideas to reach the goal, and LTM may provide relevant knowledge that short-term memory can draw upon.

1.2.1 Schemata

The concept of schemata (Cheng and Holyoak 1985), or mental representations of an aspect of the world, is useful to consider when explaining “novice” vs. “expert” behavior in solving problems. These knowledge hierarchies or mental representations can be viewed as a form of categorization in order to make sense of the world. The physics knowledge schema of expert physicists is more hierarchical than that of novices who are striving to be experts. In the expert schema, which can be thought to have a pyramid like structure, the most fundamental concepts are stored on the top followed by the secondary concepts and then the tertiary concepts.

1.2.2 Chunking and Cognitive Load during Problem Solving

A major revolution in cognitive science happened when George Miller showed that the capacity of LTM or Working Memory is limited to 7 ± 2 distinct units or bits or pieces of information (i.e., 5 to 9 bits) for any individual (Miller, 1956). The limit on STM is related to the concept of cognitive load, or the demand placed on a person’s working memory. If a person needs to process too many distinct bits of information at once while solving a problem, the person experiences cognitive “overload”, i.e., he/she is simply unable to process all the information (without the assistance of a tool, e.g., a calculator or pencil and paper or another person, that can reduce the cognitive load).

Later research showed that as the expertise of an individual increases in a particular field, the person's cognitive load while solving problems decreases (Sweller 1988). This is because as one's expertise expands, knowledge becomes more "automatic" or "compiled" so that several pieces of information may be stored in a single chunk and can be accessed together while solving problems. Experts may thus employ a "forward" approach to problem solving. Anderson (1982) defines a "declarative" stage of learning, in which facts about the knowledge are learned, and a "procedural" stage in which the facts have been compiled into a procedural structure.

Studies have convincingly shown that chunking is related to expertise. An example of this is a famous study performed on mental imaging via remembering chess positions (Chase and Simon 1973, Simon 1974). This study revealed that, when a chess board was dismantled and individuals were asked to reproduce it, experts at chess relied heavily upon recognizing attack and defense patterns between pieces in order to recall the pieces' positions when the chess board corresponded to a good game of chess, but did no better than novices at recalling randomized chess pieces. The inference made from this study was that chess experts' knowledge was "chunked" and they knew the position of one piece with respect to another piece so that more than 9 chess pieces could be reproduced. This study demonstrated that experts tend to organize their knowledge hierarchically in chunks.

Expert physics problem solvers, for example, may have a schema for mechanics in which the most fundamental principles are at the top of the hierarchy followed by the secondary and tertiary concepts in mechanics. While solving problems, experts can bring from their LTM into the STM large "chunks" of knowledge so that mental resources are still available for processing the information and solving problems successfully. On the other hand, novices may have to allocate memory for individual variables and concepts, e.g., vector, subtraction, displacement, velocity, acceleration, force, etc. and, due to the limited capacity of STM, can have a cognitive overload while solving problems if sufficient scaffolding is not provided.

Experts have already compiled ready approaches upon recognizing constraints, e.g., it may not take any time for a physics expert to recognize a conservation of momentum problem. Experts also have superior strategic knowledge from experience in problem solving, i.e. they are able to process and regulate their knowledge more efficiently. While it is necessary to teach students both content knowledge and problem solving skills, prior research has been done to recognize the necessary interaction between domain-specific knowledge and strategic knowledge (e.g. Alexander and Judy 1988).

Much work has been done in defining novice vs. expert behavior in this context in the realm of categorizing or grouping together physics problems based upon similarity of solutions (e.g. Chi et al. 1981, Schoenfeld 1985) as well as in understanding a categorizer's knowledge structure (e.g. Larkin et al. 1980, Fergusson-Hessler and de Jong 1987). Physics experts have been shown to group and solve problems according to physical principles and concepts (Hardiman et al. 1989), analogous to chess experts' reliance on elements of chess strategy. Novices in physics, on the other hand, tend to start their approach using "surface" features of physics problems, e.g., elements of the presented physical situation or target and intermediate variables presented in the problem. Procedural differences between novices and experts are also highly recognized (e.g., Dhillon 1998). Models of expert problem solving have thus been developed (e.g., Chi et al. 1982). The work of Chi et al. (1981) will be discussed later in further detail with regard to part of the research in this thesis.

1.3 METACOGNITION

Metacognition is defined traditionally as the knowledge and experiences obtained about one's own cognitive processes (Flavell 1979). It may be more simply described as thinking about thinking. In addition, metacognition is one of the most actively investigated cognitive processes in contemporary psychological research (Tobias et al. 1999). This definition has been elaborated upon to focus on specific elements of metacognition: knowledge, monitoring, and experiences. For example, Schoenfeld (1987) describes three ways of talking about metacognition in the context of mathematical problem solving: the student's beliefs and intuition about the subject material, the student's knowledge of his or her own thought processes, and self-awareness, i.e. keeping track of one's own actions. When students come to the classroom, explicit support and guidance must be provided to develop their metacognitive skills. Many factors must be considered by the instructor in order to develop problem solving, reasoning, and metacognitive skills (e.g. Yerushalmi and Eylon 2003, Scott et al. 2007). The development of these more expert-like thinking skills must happen while students are acquiring content knowledge. Conceptual understanding is necessary in order to understand physics problems, and it is highly useful to model qualitative strategies in an instructional strategy to emphasize this to learners (Leonard et al. 1996).

1.3.1 Metacognition, heuristics, and the problem-solving process

A process-oriented, metacognitive approach is highly beneficial to the problem solving process (Berardi-Coletta et al. 1995). Necessary components of this process have been defined as follows: creating a mental representation of the problem, establishing a strategy for solving the problem, and being able to handle difficulties in the actual solution process (Davidson et al. 1994). These steps involve intuitive heuristic problem-solving choices that must be learned as the material is learned.

The creation of mental representations, i.e., structures based upon available knowledge that are relevant to a problem, is a vital component of problem solving. Numerous studies have shown that development of knowledge representations improves problem solving ability in students. Examples of knowledge representations are as follows. Analogies (Gick and Holyoak 1980) may make use of hierarchical structuring to allow more thorough understanding of a physical concept even if, for example, the content knowledge specific to the domain of physics is not well understood (Gentner 1983). Diagrams may synthesize a large amount of textual description of a physical situation (e.g., Larkin and Simon 1989, van Meter et al. 2006). Active self-explanations are useful when a problem solver can generate a high rate of meaningful explanations (Chi et al. 1989, Chi et al. 1994). Due to their hierarchical knowledge, problem solving experts are able to create more sophisticated and more abstract representations than novices can (Larkin et al. 1980).

The term "heuristic" refers to methods that are useful in problem solving. Heuristic problem solving approaches are generally "rules of thumb" that are used to efficiently obtain a solution. Rather than an algorithmic method used to solve a problem, heuristic methods are usually based on inferences and educated guesses based upon problem solving experience and are very useful to quickly reach a solution to a problem. Polya (1945) describes many different kinds of heuristic methods, e.g., analogies and drawing a picture as described previously, and suggests a general technique to solve problems: understand the problem, suggest a reasonable plan, carry out the plan, and finally reflect on your work to make it better, e.g. if the plan was not successful. A good example of a heuristic is a means-ends analysis (e.g. Newell and Simon 1963, Newell and Simon 1972, Larkin et al. 1980). A problem solver first envisions the target quantity for which to solve and actively considers the most effective strategy to obtain that goal at each step of the problem.

However, there are potential weaknesses with heuristic models of problem solving based upon the fact that by definition they are not guaranteed to lead to the correct solution. Intuitive methods do not necessarily lead to a solution. For example, heuristic methods themselves may be influenced by biases which can lead to specific mistakes (Tversky and Kahneman 1974). Therefore one must take care when employing heuristics as part of the problem solving and learning process. While an intuitive approach may lead more directly to a solution and thus be more effective, such an approach may not always work even for expert problem solvers (e.g. Singh 2002). For example, it has been argued that the means-ends analysis is novice-like (Larkin et al. 1980), inefficient and may conflict with schemata acquisition in terms of cognitive load (Sweller 1988, Sweller et al. 1990).

1.4 PHYSICIST'S PERSPECTIVE: AN OVERVIEW OF PROBLEM SOLVING IN PER

Across-domain differences, e.g. differences between mathematics and physics in terms of knowledge content, make it necessary to consider the relative importance of content and strategy knowledge (Schoenfeld and Herrmann 1982). For example, it is important that physicists are able to conduct research in problem solving because physicists have access to both an understanding of physics and knowledge of a successful analytical problem solving approach.

This is not to say that transfer of problem solving skills between physics and, for example, mathematics cannot occur at all (e.g. Halloun and Hestenes 1985), but rather to indicate that transfer of problem solving skills can be difficult (e.g. Bassok and Holyoak 1989).

One must also consider the role of a general set of skills found in scientific reasoning, as opposed to domain-specific knowledge (Schunn and Anderson 1999) and the extent to which these skills are properly taught in a classroom setting (Schunn and Anderson 2001). In addition to having content knowledge to properly understand a problem's solution, physicists also understand the importance of a systematic problem solving strategy based upon understanding physical principles and concepts. Physics reasoning is therefore incredibly important in terms of both qualitative representations of a problem and hierarchical knowledge structures (van Heuvelen 1991).

Hsu et al. (2004) provide an excellent literature overview of research in problem solving, as well as a brief glossary of terms that are alluded to in this thesis. In addition, McDermott and Redish (1999) mention an essential sample of work done in problem solving. The resources listed are by no means exhaustive and, in addition to some highlights, a couple of separate sources will be described here as well.

Reif pioneered the examination of problem solving within physics education (Reif et al., 1976). He was one of the first physics education researchers to conduct investigations of issues central to problem solving and how the issues are related to physics (e.g. Reif, 1986). Furthermore, he described how mechanisms recognized in cognitive science ought to be emphasized in order to improve upon the traditional state of physics education, and science education generally (Reif 1981). In the latter reference, Reif addressed the need for students to learn a useful problem solving strategy, background content knowledge, and an efficient, hierarchical organization of knowledge (e.g., Eylon and Reif 1984).

From here, much has been developed in the realm of pedagogical, instructional approaches to problem solving and implementation of these approaches during instruction. The following are only a few examples of what exists within physics education research. A basic modeling approach for problem solving (Hestenes 1987) emphasizes a mathematical modeling of the physical world and its role in curricular instructional design, and has been shown to improve student performance (Halloun and Hestenes 1987). It is therefore possible to implement the teaching of problem solving into curricula and improve students' problem solving approach by focusing on specific elements of a systematic method (e.g. Heller and Reif 1984, Mestre et al. 1993, Leonard et al. 1996; also see Zhu and Simon (1987) in the context of mathematics rather than physics).

Alternate problem constructions and approaches (e.g., van Heuvelen 1996) have been developed from analysis of qualitative problem-solving strategies (e.g., Leonard et al. 1996) that take advantage of a direct strategic problem solving approach. An example of an alternate problem approach is the development of context-rich problems (Heller and Hollabaugh 1992) which are similar to real-life situations and induce the student to read problems carefully and adapt a good strategic approach to solving the problem. Context-rich problems have been shown to be beneficial to students' problem-solving approach, for example, in cooperative group problem solving (Heller et al. 1992, Duch 1997). This topic receives special attention here because one of the studies described in a later chapter will employ context-rich problems. A good library of context-rich problems has been developed by the University of Minnesota

Physics Education Research Group which can be found hosted at the following website: http://groups.physics.umn.edu/physed/Research/CRP/crintro.html.

Another example of an alternate problem construction is that of isomorphic problems (e.g., Singh 2008a and 2008b). The concept entails two problems that involve the same physical principles in different contexts. Novices may categorize these problems differently while experts will likely categorize them similarly. It is also of interest to note that some contexts are more accessible to novices than others in that novices may solve one problem more easily than another isomorphic problem. We will make use of isomorphic problems. In one study discussed in this thesis, we will even consider the extreme approach of asking the exact same problem twice with some time-interval in between.

Analysis of expert-novice performance in physics problem solving has also been developed. A direct invocation of an expert problem solving approach (Dufresne et al. 1992) focused on constraining students to an expert's problem-solving approach based upon schema and improved their skill by shifting their focus onto deeper structure contained within problems, e.g., physical principles. Intuitive knowledge also plays a role in transition from novice to expert problem solving ability (Sherin 2006). As stated previously (Singh 2002), sometimes this intuition can even fail expert problem solvers who use a systematic approach towards more difficult, novel problems.

Computational work in education was initially used in modeling human problem solving by mapping to artificial intelligence (e.g. Larkin et al. 1980, Newell and Simon 1972, Feigenbaum and Simon 1984), and some progress continues in this direction (e.g. Gobet and Simon 2000, Waters and Gobet 2008). However, active assistance to students has recently been the more commonly used function of computers in the field of physics education research and specifically in problem-solving instruction. Specifically, the intended utility is to offer students feedback and guidance when working on individual physics problems based on heuristic problem-solving methods. This is another example of how effective problem-solving techniques may be implemented. Personal Assistants for Learning, i.e., PALs (Reif and Scott 1999, Scott 2001), for example, focus on teaching systematic scientific thinking skills using the reciprocal teaching method employed by Palinscar and Brown (also see section 1.5.1). Following the reciprocal teaching method, PAL sometimes takes the role of an instructor and asks the person using it to carry out certain steps in problem solving while taking the role of a student and letting the person using it ask questions. The PAL will alternate between the role of a student and the traditional role of the instructor in order to reinforce the student's learning. Andes (e.g., VanLehn et al. 2002) is a more intensive program designed to both guide and track student progress through individual steps of a problem while being minimally invasive.

1.5 REFLECTION ON PROBLEM SOLVING

One metacognitive task that may help students learn from problem solving is reflection, namely the act of monitoring one's own work on a problem. This is a key element of the problem solving learning process. Reflection allows the learner to detect mistakes and misconceptions in an initial attempt at a problem. Davidson et al. (1994) argue that the complexity of problem solving implies the need for attention to planning, monitoring and regulating cognition. Experts in a particular field reflect upon problem solving as they solve a problem because chunking of knowledge leaves some cognitive resources for reflection. However, novices may have all of their cognitive resources used up by the demands of the problem and may be able to reflect upon problem solving only at the end of the problem. These reflection activities include asking questions such as “what did I learn by solving this problem?”, “how does it connect with what I already know?”, “how will I be able to recognize that the same principle should be used to solve another problem in the future?,” etc.

While the reflection task has been investigated extensively, it has only been fairly recently that PER has turned its attention to the task (e.g. Yerushalmi and Eylon, 2003; Scott et al. 2007). Reflection on one’s work is essential to learning from mistakes made in problem-solving.

1.5.1 The Cognitive Apprenticeship Model

In the traditional apprenticeship model, an expert at a trade will model a task for his student, and coach the student’s attempts to follow that model. During this process, the expert eventually weans the student to independent performance by “fading,” or phasing out of, this instruction as the student becomes more proficient at the task. Cognitive apprenticeship (Collins et al. 1990) was built by using this model to promote learning in the context of formal educational instruction.

The idea of modeling, coaching and fading from an instructor’s perspective is meant to teach a robust method for problem-solving which breaks down the problem solving process so that students can more readily digest it. In other words, the cognitive apprenticeship provides scaffolding to this end. Reflection is important in these situations as the student must develop independence and confidence in problem-solving skills.

In addition to modeling, coaching and scaffolding, Collins et al. argue the usefulness of reflection. Students can “compare their own problem-solving processes with those of an expert, another student, and ultimately, an internal cognitive model (p. 482).” Therefore, it would be of interest to employ cognitive apprenticeship to investigate reflection as the latter is inherent in the task.

Collins et al. cite examples of cognitive apprenticeship in action. Reciprocal teaching (e.g. Palinscar and Brown 1984), in which a student must explain material to an instructor, is a form of active learning that emphasizes dialogue between student and instructor. Students are required to do more than sit and listen to a lecture (i.e. “passive” learning). The original study and follow-up studies by Palinscar and Brown were focused on reading comprehension. Scardamalia and Bereiter (1985) introduced the idea of procedural facilitation, i.e. breaking down the planning aspect of problem solving (here, the context of writing) into goals facilitated by specific prompts so that students may learn to organize their planning as experts do. Schoenfeld (1985) emphasized heuristic strategies employed by experts in teaching mathematics to students using different formats, e.g. group problem-solving sessions. These examples center on analyzing knowledge and strategies required for expertise.

The cognitive apprenticeship model provides scaffolding, i.e., a general structural framework, for students to learn a specific reflection task. In addition to content understanding, it is important to check students’ problem-solving method. A prescribed method, for example, should include the students’ ability to describe the problem, create and execute a plan, and assess the solution (Heller and Reif, 1984).

1.6 A STUDY OF REFLECTION IN THE CONTEXT OF PROBLEM SOLVING

Five different studies that involve reflection on problem solving are described in the following chapters. In the first investigation, we revisit the categorization study of Chi et al. (1981), which was conducted with only eight introductory physics students, with a large number of introductory students in a classroom setting. The original Chi study came up with some excellent qualitative results that could perhaps be compared to more quantitative data. Categorization may be interpreted as a form of reflection in that students must think about how to solve the problem in order to properly categorize it. It is possible that first-year physics students will exhibit reflection on a higher level than simply describing "surface" features of the problem. We also compare the categorization by calculus-based and algebra-based introductory physics students with physics graduate students and faculty from a previously conducted study (Singh 2009). We find that calculus-based students seem to categorize better than algebra-based students, possibly from reduced cognitive load based on a more structured understanding of the role between mathematics and physical concepts. In addition, we find that there is an overlap between introductory students and graduate students, suggesting that we may not automatically consider graduate students as experts or introductory students as novices.

The second study examines reflection directly as an individual exercise in the format of a recitation activity whose goal is to diagnose one's own errors, as well as transfer of this knowledge to an isomorphic problem. Here, a rubric is designed based upon heuristic considerations in order to properly evaluate students' self-diagnostic approaches. We investigate the level of scaffolding necessary for students to effectively self-diagnose and we also investigate whether transfer of knowledge occurs from a self-diagnosed quiz problem to a paired isomorphic problem on a midterm exam. We will examine the concepts of near and far transfer

(Etkina and Mestre 2004) by examining two sets of problems in this context. We find that in terms of physics content, students with minimal scaffolding may benefit from experiencing more cognitive engagement with regard to transferring the knowledge to a paired problem, but if the problem is sufficiently difficult, students may be unable to learn from minimal scaffolding. In contrast, superficial review of one's work may require no cognitive engagement and may not be effective for learning. We also find little improvement on problem-solving approaches, suggesting that one or two interventions are not enough to improve problem solving habits.

The third study observes whether advanced undergraduate students automatically reflect and learn from their mistakes. In this study, undergraduate students in advanced quantum mechanics are asked the same question twice during two different exams. The rubric designed in the second study is adapted for evaluating students in this situation. We find that one may not assume that advanced undergraduate students have implicitly learned to reflect on their work as a result of further experience in formal coursework on even a direct transfer, and in fact exhibit tendencies that may be considered novice-like.

The fourth study examines reflection as a peer exercise in the form of actively discussing homework problem solutions in an in-class recitation setting. We seek to find out whether orienting the recitation around this activity will cause students to more frequently invoke basic elements of effective problem solving strategy, as well as the effect on exam grades. We find that while no quantitative difference occurs on the final exam score between students who participated in the peer exercise and students who did not (although there is some evidence that those who participated in the peer exercise were somewhat weaker at the beginning of the course), basic problem-solving strategies such as drawing diagrams seem to be encouraged by the peer activity.

The last study investigates students' epistemological views towards problem solving. We adopted a survey (Cummings et al. 2004) based upon the more general Maryland Physics Expectations Survey (Redish et al. 1998) and added several of our own questions which focus on several topics, one of which is reflection on problem solving. One goal is to compare graduate students with algebra-based and calculus-based introductory students and faculty. The graduate students' responses are compared when they answer these survey questions about graduate-level problem solving and problem solving in introductory physics. We focus on each of these groups' responses on individual questions and analyze cases where major differences are found between groups. The results focus on graduate students and seem to indicate two themes. First, graduate students are seen as in a transition from novice-like expectations and approaches to those of experts. Second, graduate students often have less expert-like views towards graduate-level questions than towards introductory-level questions. In addition, interviews suggest that there are different reasons for novices and experts to express views that do not seem expert-like, and therefore answers to the survey questions should be treated more carefully in light of this.

1.7 CHAPTER REFERENCES

Alexander, P., & Judy, J. (1988). "The interaction of domain-specific and strategic knowledge in academic performance." *Review of Educational Research* 58(4), 375-404.

Anderson, J. (1982). "Acquisition of Cognitive Skill." *Psychological Review* 89(4), 369-406.

Bassok, M. & Holyoak, J. (1989). "Interdomain transfer between isomorphic topics in algebra and physics." *Journal of Experimental Psychology: Learning, Memory and Cognition* 15, 153-166.

Berardi-Coletta, B., Dominowski, R., Buyer, L., & Rellinger, E. (1995). "Metacognition and Problem Solving: A Process-Oriented Approach." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21 (1), 205-223.

Chase, W., & Simon, H. (1973). "Perception in chess." *Cognitive Psychology* 4, 55-81.

Cheng, P. & Holyoak, K. "Pragmatic reasoning schemas." *Cognitive Psychology* 17, 391-416.

Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). "Self-Explanations: How students study and use examples in learning to solve problems." *Cognitive Science* 13, 145-182.

Chi, M., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). "Eliciting self-explanations improves understanding." *Cognitive Science* 18, 439-477.

Chi, M., Feltovich, P., Glaser, R. (1981). "Categorization and representation of physics knowledge by experts and novices." *Cognitive Science* 5, 121-152.

Chi, M., Glaser, R., & Rees, E. (1982). "Expertise in problem solving." In *Advances in the Psychology of Human Intelligence, Vol. 1*, edited by R. Sternberg. (Erlbaum, Hillsdale, NJ), 7-75.

Collins, A., Brown, J., & Newman, S. (1989). "Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing and Apprenticeship." In *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser.* Edited by R. Glaser and L. Resnick. Hillsdale, NJ: Lawrence Erlbaum Associates, 453-494.

Cummings, K., Lockwood, S., & Marx, J. (2004). "Attitudes toward problem solving as predictors of student success." In *2003 Physics Education Research Conference Proceedings*. Edited by J. Marx, S. Franklin, and K. Cummings. *American Institute of Physics,* 133-136.

Davidson, J., Deuser, R. & Sternberg, R. (1994). "The Role of Metacognition in Problem-Solving." In *Metacognition: Knowing About Knowing.* Edited by J. Metcalfe and A. Shimamura. Cambridge, MA: MIT Press, 207-26.

Dhillon, A. (1998). "Individual differences within problem-solving strategies used in physics." *Science Education* 82, 379-405.

Duch, B. (1997). "Problem-based learning in physics: Making connections with the real world." In *AIP Conference Proceedings: The Changing Role of Physics Departments in Modern Universities,* 399, 557-565. Edited by E. Redish and J. Rigden.

Dufresne, R., Gerace, W., Hardiman, P. & Mestre, J. (1992). "Constraining novices to perform expert-like problem analyses: Effects on schema acquisition." *Journal of the Learning Sciences* 2, 307-331.

Etkina, E. & Mestre, J. (2004). *Implications of Learning Research for Teaching Science to Non-Science Majors.* Harrisburg, PA: SENCER.

Eylon, B. & Reif, F. (1984). "Effects of knowledge organization on task performance." *Cognition and Instruction* 1(1), 5-44.

Feigenbaum, E., & Simon, H. (1984). "EPAM-like models of recognition and learning." *Cognitive Science* 8, 305-336.

Ferguson-Hessler, M., & de Jong, T. (1987). "On the quality of knowledge in the field of electricity and magnetism." *American Journal of Physics* 55 (6), 492-497.

Flavell, J. (1979). "Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry." *American Psychologist* 34(10), 906-911.

Gentner, D. (1983). "Structure-Mapping: A Theoretical Framework for Analogy." *Cognitive Science* 7, 155-170.

Gick, M. & Holyoak, K. (1980). "Analogical Problem Solving." *Cognitive Science* 12, 306-355.

Gray, K., Adams, W., Weimann, C., & Perkins, K. (2008). "Students know what physicists believe, but they don't agree: A study using the CLASS survey." *Phys. Rev. ST Phys. Educ. Res.* 4(2), 56-65.

Gobet, F., & Simon, H. (2000). "Five seconds or sixty? Presentation time in expert memory." *Cognitive Science, 24,* 651-682.

Hake, R. (1998). "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses." *American Journal of Physics* 66, 64.

Halloun, I., & Hestenes, D. (1985). "The initial knowledge state of college physics students." *American Journal of Physics* 53, 1043-1055.

Halloun, I. & Hestenes, D. (1987). "Modeling instruction in mechanics." *American Journal of Physics* 55(5), 455-462.

Hardiman, P., Dufresne, R., & Mestre, J. (1989). "The Relation between Problem Categorization and Problem Solving among Novices and Experts." *Memory & Cognition* 17(5), 627-638.

Heller, J. and Reif, F. (1984). "Prescribing effective human problem-solving processes: Problem description in physics." *Cognition and Instruction* 1, 177-216.

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, 627-636.

Heller, P. & Hollabaugh, M. (1992). "Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups." *American Journal of Physics* 60, 637-644.

Henderson, C. (2002). "Common Concerns about the Force Concept Inventory." *The Physics Teacher* 40, 542-547.

Hestenes, D. (1987). "Towards a Modeling Theory of Physical Instruction." *American Journal of Physics* 55(5), 440-454.

Hestenes, D. (1995). "Interpreting the force concept inventory." *Physics Teacher* 33, 502.

Hestenes, D., Wells, M., and Swackhamer, G. (1992). "Force Concept Inventory." *Physics Teacher* 30, 141.

Hsu, L., Brewe, E., Foster, T., & Harper, K. (2004). "Resource Letter RPS-1: Research in problem solving." Edited by Stuewer, R. *American Journal of Physics* 72 (9), 1147-1156.

Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980). "Expert and novice performance in solving physics problems." *Science* 208, 1335-1362.

Larkin, J. &. Reif, F. (1979). "Understanding and teaching problem solving in physics." *Eur. J. Sci. Ed.* 1(2), 191-203.

Larkin, J. & Simon, H. (1989). "Why a diagram is (sometimes) worth ten thousand words." In *Models of Thought, Vol. II* by H. Simon. New Haven, CT: Yale University Press, 413-437.

Leonard, W., Dufresne, R., & Mestre, J. (1996). "Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems." *American Journal of Physics* 64, 1495-1503.

McDermott, L. (1990). "A View from Physics." In *Towards a Scientific Practice of Scientific Education: Part I: View from the Disciplines.* Edited by Gardner, M., Stage, E. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., 3-30.

McDermott, L. & Redish, E. (1999). "RL-PER1: Resource Letter on Physics Education Research." *American Journal of Physics* 67(9), 755-767.

Mestre, J., Dufresne, R., Gerace, W., Hardiman, P., & Touger, J. (1993). "Promoting skilled problem-solving behavior among beginning physicists." *Journal of Research in Science Teaching* 30, 303-317.

Miller, G. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological Review* 63, 81-97.

Newell, A., & Simon, H. (1972). *Human Problem Solving.* Englewood Cliffs, NJ: Prentice Hall.

Palinscar, A., & Brown, A. (1984). "Reciprocal teaching of Comprehension-Fostering and Comprehension-Monitoring Activities." *Cognition and Instruction* 1(2), 117-175.

Polya, G. (1945). *How to Solve It.* Princeton, NJ: Princeton University Press.

Redish, E., Steinberg, R., & Saul, J. (1998). "Student Expectations in Introductory Physics." *American Journal of Physics* 66, 212-224.

Reif, F. (1981). "Teaching problem solving – a scientific approach." *Physics Teacher* 19, 310-316.

Reif, F. (1986). "Scientific approaches to science education." *Physics Today* 39(11), 48-54.

Reif, F. (1995). "Millikan Lecture 1994: Understanding and teaching important scientific thought processes." *Am. J. Phys.* 63, 17-32.

Reif, F., & Allen, S. (1992). "Cognition for interpreting scientific concepts: A study of acceleration." *Cognition and Instruction* 9, 1-44.

Reif, F., & Larkin, J. (1991). "Cognition in scientific and everyday domains: Comparison and learning implications." *Journal of Research in Science Teaching* 28, 733-760.

Reif, F., Larkin, J., & Brackett G. (1976). "Teaching general learning and problem solving skills." *American Journal of Physics* 44(3), 212-217.

Reif, F. & Scott, L. (1999). "Teaching scientific thinking skills: Students and computers coaching each other." *American Journal of Physics* 67, 819-831.

Scardamalia, M., & Bereiter, C. (1985). "Fostering the development of self regulation in children's knowledge processing." In *Thinking and Learning Skills: Research and Open Questions* (pp. 563-577). Edited by S. Chipman, J. Segal, and R. Glaser. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schoenfeld, A. H. (1985). *Mathematical problem solving.* Orlando, FL: Academic Press.

Schoenfeld, A. H. (1987). "What's all the fuss about metacognition?" In *Cognitive science and mathematics education* (pp. 189-215). Edited by A. H. Schoenfeld. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schunn, C. & Anderson, J. (1999). "The Generality/Specificity of Expertise in Scientific Reasoning." *Cognitive Science* 23(3), 337-370.

Schunn, C. & Anderson, J. (2001). "Acquiring Expertise in Science; Explorations of What, When, and How." In *Designing For Science: Implications from Everyday, Classroom, and Professional Settings.* Edited by K. Crowley, C. Schunn, & T. Okada. Mahwah, NJ: Lawrence Erlbaum Associates, 84-114.

Scott, L. (2001) "Design and assessment of an interactive physics tutoring environment." Dissertation, University of Pittsburgh School of Education.

Scott, M., Stelzer, T., & Gladding, G. (2007). "Explicit Reflection in an Introductory Physics Course." In *Physics Education Research Conference, AIP Conference Proceedings 2007*. Edited by L. Hsu, C. Henderson, & L. McCullough. *AIP Conference Proceedings* 951, 188-191.

Sherin, B. (2006). "Common sense clarified: The role of intuitive knowledge in physics problem solving." *Journal of Research in Science Teaching* 43(6), 535-555.

Simon, H. (1974). "How big is a memory chunk?" *Science* 183(4124), 482-488.

Singh, C. (2002). "When physical intuition fails." *American Journal of Physics* 70, 1103-1109.

Singh, C. (2008a). "Assessing student expertise in introductory physics with isomorphic problems. I. Performance on nonintuitive problem pair from introductory physics." Physical Review Special Topic Physics Education Research 4, 010104.

Singh, C. (2008b). "Assessing student expertise in introductory physics with isomorphic problems. II. Effect of some potential factors on problem solving and transfer." Physical Review Special Topic Physics Education Research 4, 010105.

Singh, C. (2009). "Categorization of problems to assess and improve proficiency as teachers and learners." *American Journal of Physics* 77, 73-80.

Sweller, J. (1988). "Cognitive load during problem solving: effects on learning." *Cognitive Science* 12(2), 257-285.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). "Cognitive load as a factor in the structuring of technical material." *Journal of Experimental Psychology* 119(2), 176-192.

Tobias, S., Everson, H., Laitusis, V., & Fields, M. (1999). "Metacognitive Knowledge Monitoring: Domain Specific or General?" *Paper presented at the Annual meeting of the Society for the Scientific Study of Reading, Montreal*

Tversky, A., & Kahneman, D. (1974). "Judgment under uncertainty: Heuristics and biases." *Science* 185, 1124-1131.

Van Heuvelen, A. (1991). "Learning to think like a physicist: A review of research-based strategies." *American Journal of Physics* 59, 891-897.

Van Heuvelen, A. (1996). "Active Learning Problem Sheets." Plymouth, Michigan: Hayden-McNeil, 1996.

VanLehn, K., Lynch, C., Taylor, L., Weinstein, A., Shelby, R., Schulze, K., Treacy, D., & Wintersgill, M. (2002). "Minimally invasive tutoring of complex physics problem solving." In *Intelligent Tutoring Systems: 6th International Conference.* Edited by S. Cerri, G. Gouardees, and F. Paraguacu. Berlin: Springer, 367-376.

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). "Learner-generated drawing as a strategy for learning from content area text." *Contemporary Educational Psychology* 31, 142-166.

Waters, A., & Gobet, F. (2008). "Mental imagery and chunks: Empirical and computational findings." *Memory & Cognition* 36 (3), 505-517.

Yerushalmi, E. & Eylon, B. (2003). "Assessing reflection on practice: a problem solving perspective." *2003 Physics Education Research Conference. AIP Conference Proceedings* 720, pp. 153-156.

Zhu, X., & Simon, H. (1987). "Learning mathematics from examples and by doing." *Cognition and Instruction* 4(3), 137-166.

2.0 CATEGORIZATION REVISITED: ASSESSING STUDENT ENTERPRISE IN INTRODUCTORY PHYSICS

2.1 ABSTRACT

We revisit the classic study by Chi, Feltovich, and Glaser (1981) related to categorization of introductory mechanics problems. However, ours is an "in vivo" study conducted in a classroom environment rather than their "in vitro" study. We asked introductory physics students in both calculus- and algebra-based courses to categorize or group mechanics problems based upon similarity of solution. To evaluate the effect of the nature of problems on students' ability to categorize, two sets of problems were developed for categorization. Some problems in one set included those given in the prior study by Chi et al. We also compare the performance of introductory students with physics graduate students and faculty. We find that introductory physics students in the calculus-based courses perform significantly better than those in the algebra-based courses, even though the categorization task is conceptual. Moreover, there was a large overlap between the students in the calculus-based courses and physics graduate students with regard to categorizations that were assessed as "good". This large overlap contrasts our findings from that of Chi et al. and suggests that it is misleading to label all introductory physics students "novices" and all physics graduate students "experts".

2.2 INTRODUCTION

A crucial difference between the problem solving strategies used by experts in physics and beginning students lies in the interplay between how their knowledge is organized and how it is retrieved to solve problems (Meltzer 2005, Ozimek et al. 2004, Larkin 1981, Larkin et al. 1980, Maloney 1994, Singh 2002, McDermott 2001). Categorizing or grouping various problems based upon similarity of solution can be a useful tool for teaching and learning (Chi et al. 1981, Dufresne et al. 2005, Hardiman et al. 1989). Here, we revisit the classic study on the categorization of introductory mechanics problems based upon similarity of solution by introductory students in the calculus-based and algebra-based physics courses in an "in vivo," i.e. in-classroom, environment. We also interviewed a subset of students using a think-aloud protocol (Chi 1997).

In the categorization study conducted by Chi et al. (called Chi study in this chapter for convenience), eight introductory physics students were asked to categorize introductory mechanics problems based upon similarity of solution. They were also asked to explain the reasons for choosing a particular category. Unlike experts who categorize problems based on the physical principles involved in solving them, introductory students categorized problems according to surface features (e.g. inclined planes in one category and pulleys in a separate category).

The Chi study was an "in vitro," or out-of-classroom, study in which 24 problems from introductory mechanics were given to eight introductory physics student volunteers (called novices) and eight physics graduate student volunteers (called experts). There were no gross quantitative differences in the number of categories produced by the two groups. There were no differences in the number of categories produced (approximately 8.5 categories by each group on average) and four of the largest categories produced by each student from both groups captured the majority of the problems (80% for experts and 74% for novices). Immediately after the first categorization, each student was asked to re-categorize the same problems. The second categorization matched the first categorization very closely, based upon which it was concluded in the Chi study that both experts and novices were able to categorize problems into groups that were meaningful to them.

Analysis of data in the Chi study showed that experts and novices categorize problems differently in categories. It was found that the introductory physics students (novices) are sensitive to the context and surface features of a physics problem and based their categorization on the problem's literal features. On the other hand, physics graduate students (experts) were able to distill physics principles applicable in a situation and categorize the problems based upon those principles. For example, 75%, 50%, 50%, and 38% of the novices had “springs,” “inclined plane,” “kinetic energy,” and “pulleys” as one of the categories respectively, while 25% of the experts used “springs” as a category but “inclined plane,” “kinetic energy,” and “pulleys” were not chosen as category names by experts.

2.2.1 Goals and Motivation

In this study, we investigated the answers to the following questions:

- Does the ability to categorize introductory mechanics problems by introductory physics students depend strongly upon the nature and context of the questions that were asked?
- How do hundreds of introductory physics students in an in-class study categorize introductory mechanics problems compared to the eight introductory student

volunteers in the Chi study three decades ago at the same University? Are there any qualitative differences between the two studies?

- How does the performance of introductory physics students in the algebra-based courses compare with the performance of students in the calculus-based courses?
- How does the categorization by the introductory physics students compare with the categorization by the physics graduate students and the physics faculty?

2.3 CATEGORIZATION TASK AND DETERMINATION OF GOOD CATEGORIES

All students and faculty members who performed the categorization were provided the following instructions at the beginning of the problem set:

- Your task is to group the 25 problems below based upon similarity of solution into various groups on the sheet of paper provided. Problems that you consider to be similar should be placed in the same group. You can create as many groups as you wish. The grouping of problems should NOT be in terms of ``easy problems", ``medium difficulty problems" and ``difficult problems" but rather it should be based upon the features and characteristics of the problems that make them similar. A problem can be placed in more than one group created by you. Please provide a brief explanation for why you placed a set of questions in a particular group. You need NOT solve any problems.
- Ignore the retarding effects of friction and air resistance unless otherwise stated.

The sheet on which the individuals were asked to perform the categorization of problems had three columns. The first column asked them to use their own category name for each of their categories, the second column asked them for a description of the category that explains why those problems may be grouped together, and the third column asked them to list the problem numbers for the questions that should be placed in a category. Apart from these directions, students were not given any examples or other suggestions about the category names they should choose.

We were unable to obtain the problems from the Chi study except for 7 problems for which the problem numbers from the third edition of the introductory physics textbook by Halliday and Resnick (1974 edition) were mentioned in their paper. We developed two versions of the problem set, each with 25 problems. However, as discussed later in detail, administration of the categorization task using version I of the problem set showed major differences with respect to the Chi study. In particular, we found a large overlap between the introductory physics students and physics graduate students with regard to categorizations that were assessed as "good." Therefore, we decided to investigate the sensitivity of the student performance to the questions in the categorization task and developed version II of the problem set for categorization.

We mostly chose our own mechanics problems on sub-topics similar to those chosen in the Chi study. Fifteen problems were common in both versions I and II of the problem set but ten problems differed. In version II, seven problems were the same as the ones used in the Chi study. The context of the 25 mechanics problems varied, and in version I, the topics included one- and two-dimensional kinematics, dynamics, the work-energy theorem, and the impulse-momentum theorem. Version II, the version that included problems from the Chi study (hereafter referred to as "Chi problems"), also had some questions on rotational kinematics and dynamics (beyond uniform circular motion) and angular momentum. Version I and version II of the problem set may be found in sections A.1 and A.2 of Appendix A, respectively.

Many questions related to work-energy and impulse-momentum concepts were adapted from an earlier study (Singh and Rosengrant 2003) and many questions on kinematics and dynamics were chosen from other earlier studies (Singh 2004, Yerushalmi et al. 2007, Singh et al. 2007) because the development of these questions and their wording had gone through rigorous testing by students and faculty members. Some questions could be solved using one physics principle, for example, conservation of mechanical energy, Newton's second law, or conservation of momentum. The first two columns of Table 2.1 show the question numbers and examples of primary categories in which each question of the problem set version I (not involving Chi problems) can be placed (based upon the physics principle used to solve each problem). Questions 4, 5, 8, 24 and 25 are examples of problems that involve the use of two principles for different parts of the problem. For example, Questions 4, 8, and 24 can be grouped together in one category because they require the use of conservation of mechanical energy and momentum. The first two columns of table 2.2 show the question numbers and examples of the primary categories in which each of the 10 new questions of the problem set version II (involving Chi problems) can be placed.

A good category is based on the physical principles and concepts involved in solving a problem whereas poor or moderate categories are based upon surface features of the problems, e.g., “ramp” for objects on inclined surfaces, “pendulum” for objects tied to string, or “angular speed” if one must solve for angular speed (as opposed to a category based on principles such as rotational kinematics, rotational dynamics, and angular momentum conservation). A moderate

Table 2.1. Examples of the primary and secondary categories and one commonly occurring poor/moderate category for each of the 25 questions for version I of the problem set.

Question	Examples of Primary Categories	Examples of Secondary Categories	Poor/Moderate Categories
1	(a) momentum conservation or (b) completely inelastic collision		speed
2	(a) mechanical energy conservation or (b) 1D kinematics		speed
3	work by conservative force/definition of work		ramp
4	mechanical energy conservation and momentum conservation		only energy or momentum
5	mechanical energy conservation and Newton's second law	centripetal acceleration, circular motion/tension	only tension or only force
6	mechanical energy conservation		only spring
7	work-energy theorem/definition of work or Newton's second law/1D kinematics	relation between kinetic energy and speed	speed
8	(momentum conservation or completely inelastic collision) and mechanical energy conservation		only energy or momentum
9	2D kinematics		cliff
10	Newton's second law	circular motion/friction	only friction
11	linear momentum conservation or completely inelastic collision		speed
12	mechanical energy conservation and work-energy theorem/definition of work	friction	only friction
13	Newton's second law	Newton's third law	force
14	2D kinematics		force/cliff
15	mechanical energy conservation		speed
16	mechanical energy conservation or 2D kinematics		speed
17	Newton's second law	Newton's third law/tension	only tension
18	mechanical energy conservation or 2D kinematics		speed
19	Impulse-momentum theorem		force
20	mechanical energy conservation or 2D kinematics		speed
21	impulse-momentum theorem		force
22	2D kinematics		ramp
23	Newton's second law/1D kinematics or Work-energy theorem/definition of work	kinematic variables	force
24	mechanical energy conservation or momentum conservation or completely inelastic collision		speed
25	mechanical energy conservation and Newton's second law	centripetal acceleration, circular motion/normal force	ramp or only force

Table 2.2. Examples of the primary and secondary categories and one commonly occurring poor/moderate category for each of the 25 questions for version II of the problem set.
This set includes 7 Chi problems. The numbers in the parentheses in column 1 give the problem numbers from version I of the repeated problems and the italicized boldface numbers are the Chi problems.

Question	Examples of Primary Categories	Examples of Secondary Categories	Poor/Moderate Categories
1 (21)	impulse-momentum theorem		force
2	angular momentum conservation		angular speed, moment of inertia
3 (8)	(momentum conservation or completely inelastic collision) and mechanical energy conservation		only energy or momentum
4 (13)	Newton's second law	Newton's third law	force
5 (14)	2D kinematics		force/cliff
6 (15)	mechanical energy conservation		speed
7 (17)	Newton's second law	Newton's third law/tension	only tension
8 (19)	Impulse-momentum theorem		force
9 (24)	mechanical energy conservation and momentum conservation or completely inelastic collision		speed
10	rotational kinematics	rotational dynamics (implicit)	angular speed, only friction
11	angular momentum conservation		angular speed
12 (22)	2D kinematics		ramp
13 (12)	mechanical energy conservation and work-energy theorem/definition of work	friction	only friction
14	(a) mechanical energy conservation or (b) Newton's second law and (kinematics/work energy theorem)		speed
15	Newton's second law		only tension
16	(a) mechanical energy conservation /work-energy theorem/definition of work or (b) Newton’s second law and kinematics	friction, potential energy stored in spring, spring force	only friction, only spring
17	(a) work-energy theorem/definition of work or (b) Newton’s second law and kinematics	friction, kinetic energy, gravitational potential energy	only friction, ramp
18	(a) work-energy theorem/definition of work or (b) Newton’s second law and kinematics	friction, kinetic energy, gravitational potential energy	speed, only friction, ramp
19	angular momentum conservation		angular speed, moment of inertia
20 (2)	(a) mechanical energy conservation or (b) 1D kinematics		speed
21	angular momentum conservation		angular speed
22 (4)	mechanical energy conservation and momentum conservation		only energy or only momentum
23 (5)	mechanical energy conservation and Newton's second law	centripetal acceleration, circular motion/tension	only tension or only force
24 (10)	Newton's second law	circular motion/friction	only friction
25 (25)	mechanical energy conservation and Newton's second law	centripetal acceleration, circular motion/normal force	ramp or only force

category may also include a relevant category whose knowledge is useful for solving the problem, e.g., tension for a problem involving tension in a string, but it is not related to the primary method of solution, e.g., Newton's second law or conservation of mechanical energy.

Although we had assumptions about which categories created by individuals should be considered good or poor, we validated our assumptions with other experts. We randomly selected the categorizations performed by twenty introductory physics students and gave them to three physics faculty who had taught introductory physics recently and asked them to decide whether each of the categories created by individual students should be considered good, moderate, or poor. We asked them to mark each row which had a category name created by a student and a description of why it was the appropriate category for the questions that were placed in that category. If a faculty member rated a category created by an introductory student as good, we asked that he/she cross out the questions that did not belong to that category. The agreement between the ratings of different faculty members was better than 95%. We used their ratings as a guide to rate the categories created by everybody as good, moderate, or poor. Thus, a category was considered "good" only if it was based on the underlying physics principles. We typically rated "conservation of energy" or "conservation of mechanical energy" both as good categories. "Kinetic energy" as a category name was considered a moderate category if students did not explain that the questions placed in that category can be solved using mechanical energy conservation or the work energy theorem. We rated a category such as "energy" as good if students explained the rationale for placing a problem in that category. If a secondary category such as "friction" or "tension" was the only category in which a problem was placed and the description of the category did not explain the primary physics principles involved, it was considered a moderate category. Tables 2.1 and 2.2 show examples of the primary and secondary categories and one commonly occurring poor/moderate category for each question given in the categorization task in the two versions of the problem set.

More than one principle or concept may be useful for solving a problem. The instructions specified that students could place a problem in more than one category. Because a given problem can be solved using more than one approach, categorizations based on different methods of solution that are appropriate were considered "good" (e.g., see Tables 2.1 and 2.2). For some questions, conservation of mechanical energy may be more efficient, but the questions can also be solved using one- or two-dimensional kinematics with constant acceleration. In this chapter, the graphs only show the categories that were rated "good". If a graph shows that a particular group (introductory students in the algebra-based or calculus-based courses, graduate students, or faculty) placed 60% of the questions in a good category, it means that the other 40% of the questions were placed in the moderate or poor categories.

For questions that required the use of two major principles, those who categorized them in good categories either made a category which included both principles such as the conservation of mechanical energy and the conservation of momentum or placed such questions in two categories created by them – one corresponding to the conservation of mechanical energy and the other corresponding to the conservation of momentum. If such questions were placed only in one of the two categories, it was not considered a good categorization.

Two algebra-based introductory physics classes (with respectively 109 and 114 students in the fall semester) and one calculus-based introductory physics class (with 180 students in the spring semester) carried out the categorization task in their first week's recitation classes of their second semester courses. We note that all relevant concepts in the problem sets had been taught in the first semester of the introductory physics courses (whether it was algebra-based or calculus-based). All introductory students were told that they should try their best but they were given the same bonus points for doing the categorization regardless of how expert-like their categorizations were. The 21 physics graduate students who carried out the categorization task were enrolled in a course for Teaching Assistants (TAs) and they performed the categorization in the last class of the semester. The seven physics faculty members who categorized the problems were asked to complete the task when convenient for them and return it to the researchers as soon as possible. We note that one of the two algebra-based classes (with 109 students) was the only class which was given the version II of the problem set (which included Chi problems) to categorize. We also note that, since the introductory students in the Chi study were 8 volunteers who responded to an advertisement, we are unsure whether they were enrolled in the introductory physics course in the same semester when they performed the categorization task or had taken introductory physics earlier.

2.4 RESULTS

Classification of categories created by each individual consisted of placing each category by each person into a matrix, which consisted of problem numbers along one dimension and category along the other dimension. In essence, a “1” was placed in a box if the problem appeared in the given category and a “0” was placed if it did not. For example, for the 109 students in the algebra-based course who categorized version II of the problem set, an average of 7.02 categories per student was created. We recorded 82 proto-categories and then further boiled them down to 59 categories. The latter process was carried out because many categories were interpreted to be equivalent to other categories.

2.4.1 Dependence of Categorization on the Nature of Problems

Table 2.3 shows the list of categories that experts and novices created in the Chi study and includes the percentages of experts and novices in their study and the percentages of introductory physics students in the calculus-based course and two algebra-based courses in our study who chose each of the categories. The “cannot classify/omitted” category in Table 2.3 has the percentage of questions that students noted they could not classify or skipped. For version II of the test given to the algebra-based introductory physics class, that included 7 Chi problems, we have two separate columns in Table 2.3 showing the categorization for only those seven questions and for all questions in version II. A close look at these columns in Table 2.3 shows that the categories students choose is sensitive to the nature of problems asked.

Figure 2.1 shows a histogram of the percentage of questions placed in good categories (not moderate or poor) by introductory students in the algebra-based course that used version II of the problem set (which included the 7 Chi problems). This figure compares the average performance on the categorization task when all problems are taken together to when Chi problems are separated out. Table 2.4 describes the mean percentage of all problems, the Chi problems, and the problems not taken from the Chi study (“other problems”) respectively, as well as the p-value of a t-test performed between the means for the Chi problems and the other problems. Students’ performance on categorization was somewhat poorer ($p < 0.001$) on the Chi problems than on the other problems. For example, approximately 90% of the students placed Chi problems in good categories less than 20% of the time. Conversely, 72% of the students placed the non-Chi problems in good categories less than 20% of the time. Thus, the nature of problems affects how well students categorize those problems.

Individual interviews with four students in which students were asked to outline the

Table 2.3. Students' performance in our study vs. the Chi study.

Novice and expert categories are exactly the same as those described in Chi study. Note that the novices in Chi study were introductory physics students and the experts were graduate students. Categories shaded in gray are those for which the relevant Chi problems were lost and could not be replicated, and therefore none of the students in our study created those categories.

Chi's Categories	% of 1981 novices (8 total)	% of 1981 experts (8 total)	% of algebra-based students version II (109 total)		% of algebra-based students version I (114 total)	% of calculus-based students (180 total)
			All questions (25)	Chi questions (7)		
Novice Categories						
Angular motion (including circular)	87.5	-	72	59	57	42
Inclined planes	50	-	24	19	19	18
Velocity and acceleration	25	-	31	26	51	10.5
Friction	25	-	55	51	52	27
Kinetic energy	50	-	16	15	15	6
Cannot classify/omitted	50	-	44	18	34	39
Vertical motion	25	-	3	3	3	1
Pulleys	37.5	-	16	16	6	2
Free fall	25	-	6	1	4	6
Expert Categories						
Newton's 2nd Law (also Newton's Laws)	-	75	22	18	19	38
Energy principles (conservation of energy, work-energy theorem, energy considerations)	-	75	42	31	35	73
Angular motion (not including circular)	-	75	43	31	39	15
Circular motion	-	62.5	29	28	18	27
Statics	-	50	0	0	0	0
Conservation of Angular Momentum	-	25	7	1	1	1
Linear kinematics/motion (not including projectile motion)	-	25	51	44	42	63
Vectors	-	25	1	1	16	2
Both Novice and Expert Categories						
Momentum principles (conservation of momentum, momentum considerations)	25	75	39	11	33	64
Work	50	25	4	4	41	47
Center of mass	62.5	62.5	2	0	1	0
Springs	75	25	23	23	52	30

Table 2.4. Percentages and p-values of Chi questions and non-Chi questions placed into a good category. See Figure 2.1.

Means (%)			p-value
All questions (25)	Chi questions (7)	Other questions (18)	(Chi vs. Other)
11.6	7.0	13.5	<0.001

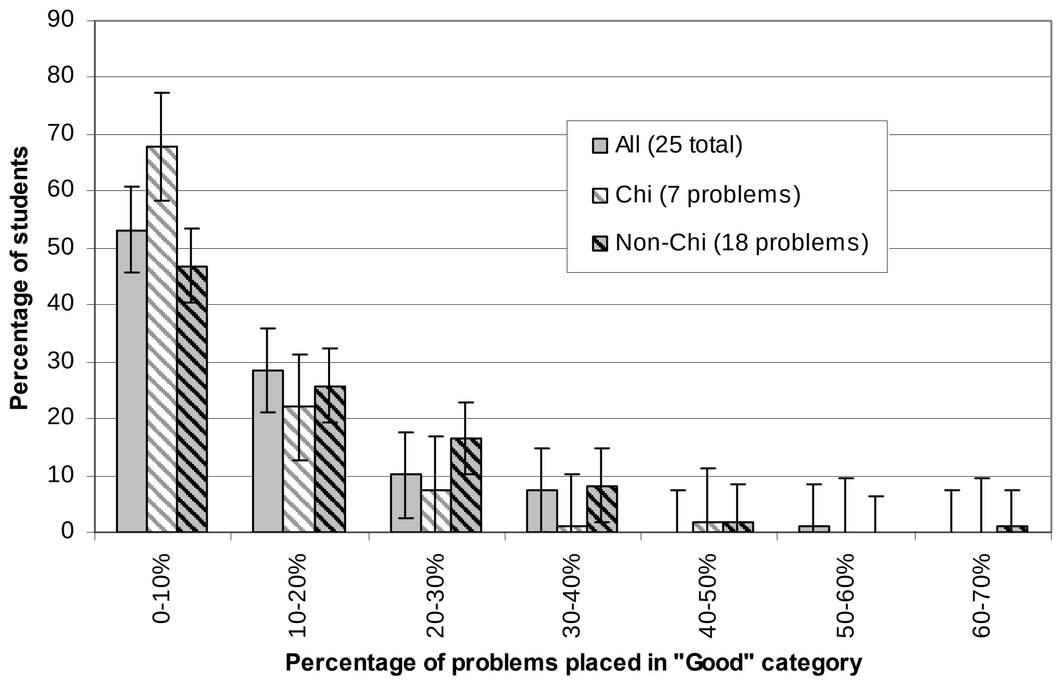


Figure 2.1. Version II histogram – percentage of good categories.
Histogram of algebra-based introductory physics students (109 students) who categorized various percentages of the 25 problems in version II of the problem set in "good" categories when asked to categorize them based on similarity of solution. The 7 Chi problems were categorized worse than the other 18 problems showing that the nature of introductory physics questions is important in students' ability to categorize them. The percentages of students for all 25 problems taken together are also shown.

procedure for solving the problems shows that students spent more time thinking about the 7 Chi problems than the other problems. This difference in the time taken suggests that the Chi problems on average may be more difficult for students compared to the others chosen for this study. This may partly be due to the fact that the 7 Chi problems were either related to rotational motion (which is generally more difficult for students than problems related to linear motion) or they involved non-equilibrium applications of Newton's law or the work-energy theorem.

Students find problems involving conservation of mechanical energy or conservation of momentum somewhat easier to deal with than the principles involved in the Chi problems. Moreover, although one of the Chi problems (problem 14 in version II) can be solved using the principle of conservation of mechanical energy, the problem is much more difficult than the traditional problems involving this principle because both masses given in the problem must be considered as a system together for the mechanical energy of the system to be conserved. However, we note that in Table 2.3, we cannot directly compare the categorizations by introductory students in our study with the categorizations by the students in the Chi study, because at the most only 7 questions are shared between the two studies.

The percentage of introductory students in our study who selected "ramps" or "pulleys" as categories (based mainly upon the surface features of the problem rather than based upon the physics principle required to solve the problem) is significantly lower than in the Chi study. One reason could be the difference in questions that were given to students in the two studies. In our study using version I of the problem set, introductory students sometimes categorized questions 3, 6, 8, 12, 15, 17, 18, 22, 24, and 25 as ramp problems, questions 6 and 21 as spring problems (question 21 was categorized as a spring problem by introductory students who associated the bouncing of the rubber ball with a spring-like behavior) and question 17 as a pulley problem. The lower number of introductory students referring to springs or pulleys as categories in our study could be due to the fact that there are fewer questions than in Chi et al. that involve springs and pulleys. However, "ramp" was also a much less popular category for introductory students in our study than in the Chi study in which 50% of the students created this category and placed at least one problem in that category (although many questions can potentially be categorized as ramp problems even in our study using both versions of the problem set). Again, the main reason for the difference could be the difference in questions because, although we have 7 Chi problems in version II, we cannot compare our data directly with theirs since most questions are different.

2.4.2 Comparison between the Algebra-Based and Calculus-Based Introductory Physics Classes

Table 2.3 also shows that the categorizations by the introductory students in the calculus-based introductory physics course are more expert-like than by the students in the two algebra-

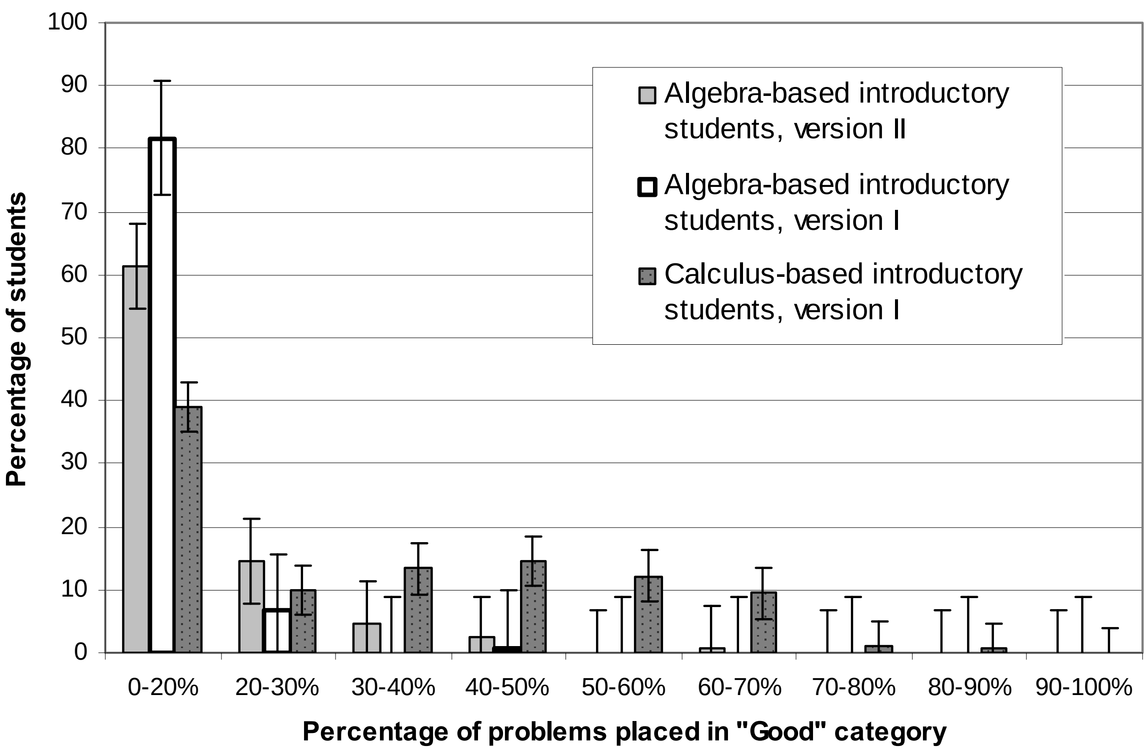


Figure 2.2. Histogram -algebra-based introductory students vs. calculus-based students. Histogram of the calculus-based introductory physics students (version I of the problem set), and two sets of algebra-based introductory students (who were given two different versions of the problem set) who categorized various percentages of the 25 problems in "good" categories when asked to categorize them based on similarity of solution. The calculus-based introductory students categorized the problems better than both groups of algebra-based introductory physics students.

Table 2.5. Percentages and p-values of questions placed into a good category for each group of introductory students. See Figure 2.2.

Group	Means (%)	p-values		
Calculus	34.4		AlgebraI	AlgebraII
Algebra version I	18.7	Calculus	<0.001	< 0.001
Algebra version II	11.6	AlgebraI		<0.001
		AlgebraII		

based courses. The difference between the overall categorization by the calculus-based students compared to the algebra-based students is also evident from Figure 2.2, which is a histogram of the percentage of students in each group vs. the percentage of problems placed in good categories by each group.

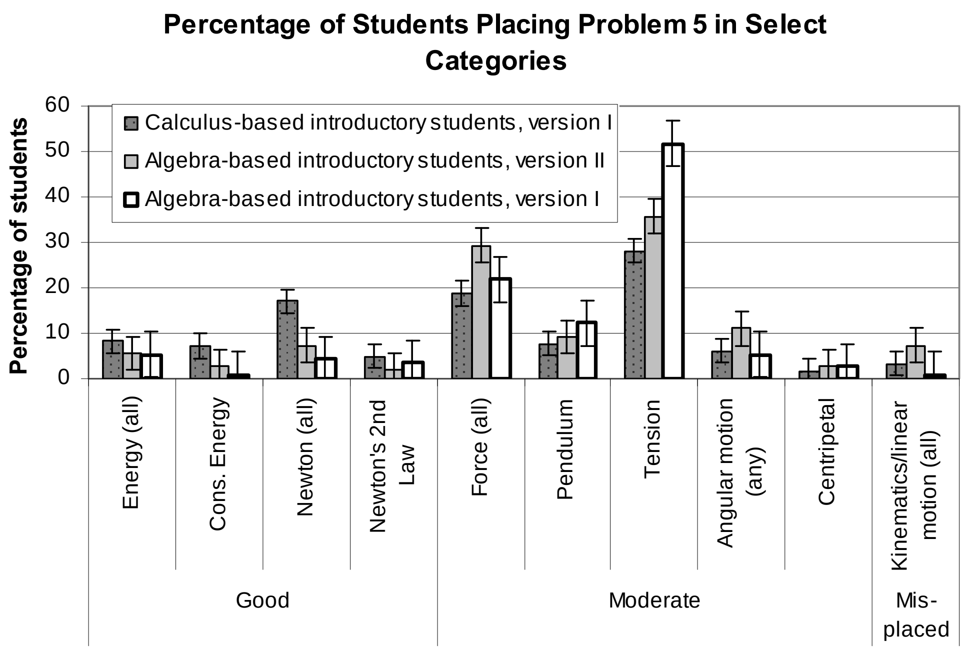


Figure 2.3. Problem 5 histogram – calculus-based intro students vs. algebra-based intro students. Histogram of the calculus-based introductory physics students (version I of the problem set), and two sets of algebra-based introductory students (who were given two different versions of the problem set) who categorized various percentages of problem 5 in different categories when asked to group them based on similarity of solution. Problem 5 requires using both the principles of conservation of mechanical energy and Newton's second law for a non-equilibrium situation. Note that the students in the calculus-based courses were able to classify this problem relatively better than the students in the algebra-based courses. Few students in any group classified this problem in categories involving both principles. Some category names have been abbreviated.

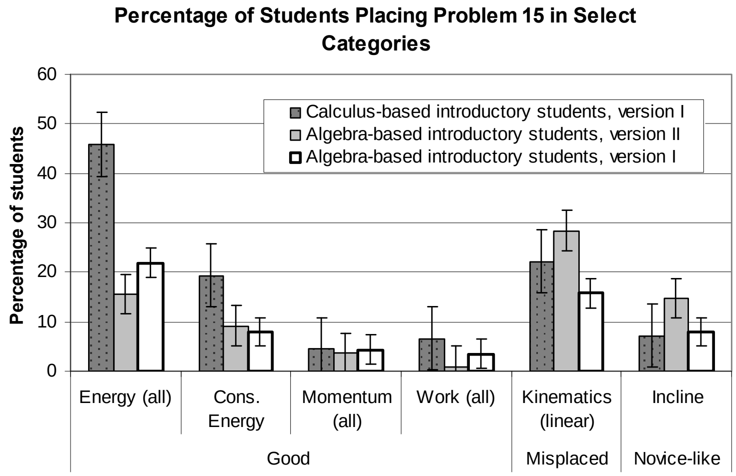


Figure 2.4. Problem 15 histogram – calculus-based intro students vs. algebra-based intro students. Histogram of the calculus-based introductory physics students (version I of the problem set), and two sets of algebra-based introductory students (who were given two different versions of the problem set) who categorized various percentages of problem 15 in different categories when asked to group them based on similarity of solution. Problem 15 requires using the principle of conservation of mechanical energy. Note that the students in the calculus-based courses were able to classify this problem relatively better than the students in the algebra-based courses. Some category names have been abbreviated, e.g., the category "cons. energy" refers to conservation of energy or conservation of mechanical energy. The category "energy (all)" includes all categories involving energy.

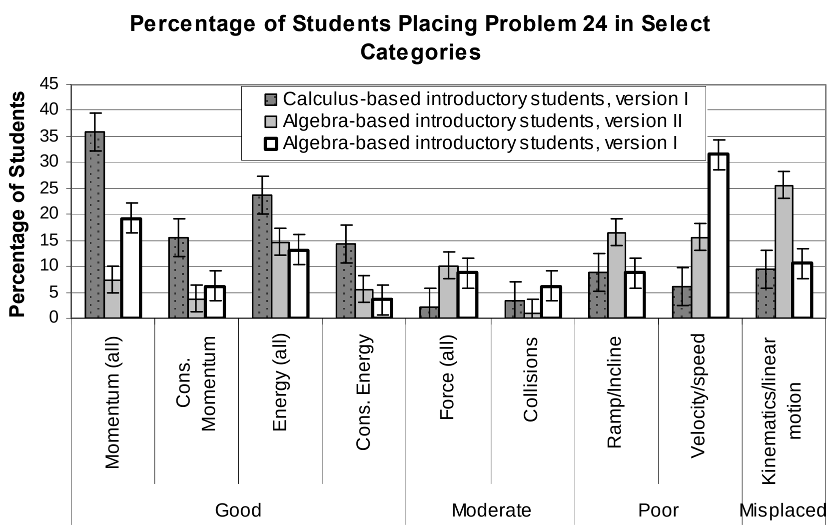


Figure 2.5. Problem 24 histogram – calculus-based intro students vs. algebra-based intro students. Histogram of the calculus-based introductory physics students (version I of the problem set), and two sets of algebra-based introductory students (who were given two different versions of the problem set) who categorized various percentages of problem 24 in different categories when asked to group them based on similarity of solution. Problem 24 required using both the principles of conservation of momentum and conservation of mechanical energy. Note that the students in the calculus-based courses were able to classify this problem relatively better than the students in the algebra-based courses. However, few students in any group classified this problem in categories involving both principles. Some category names have been abbreviated.

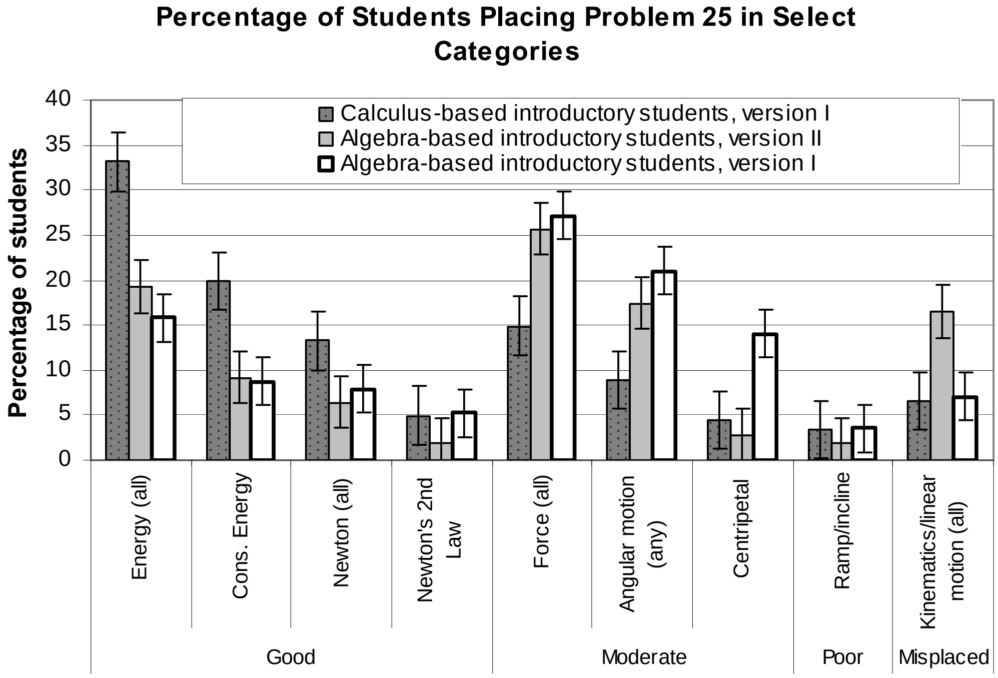


Figure 2.6. Problem 25 histogram – calculus-based intro students vs. algebra-based intro students. Histogram of the calculus-based introductory physics students (version I of the problem set), and two sets of algebra-based introductory students (who were given two different versions of the problem set) who categorized various percentages of problem 25 in different categories when asked to group them based on similarity of solution. Problem 25 requires using both the principles of conservation of mechanical energy and Newton's second law for non-equilibrium situation. Note that the students in the calculus-based courses were able to classify this problem relatively better than the students in the algebra-based courses. However, few students in any group classified this problem in categories involving both principles. Some category names have been abbreviated.

Table 2.5 indicates the mean percentage of categories placed by each group of students into good categories, as well as p-values between the three groups. There is a very significant difference between all three groups: calculus-based students perform better than algebra-based students, and algebra-based students with version I of the the problem set categorized better than those with version II of the problem set.

The calculus-based group has a very flat distribution in Figure 2.2 in the range of 20%-70% of the problems placed in good categories, while both the algebra-based groups perform significantly worse (with algebra-based students using version I of the problem set, which is the

Table 2.6. Percentages and p-values of problems placed into good categories by the introductory student groups. See Figures 2.3 to 2.6.

Figure 2.3 (3a) – Problem 5				
Group	Means (%)	p-values		
Calculus	9.4		AlgebraI	AlgebraII
Algebra version I	4.4	Calculus	0.001	< 0.001
Algebra version II	3.5	AlgebraI		0.73
		AlgebraII		

Figure 2.4 (3b) – Problem 15				
Group	Means (%)	p-values		
Calculus	19.1		AlgebraI	AlgebraII
Algebra version I	7.0	Calculus	<0.001	< 0.001
Algebra version II	9.4	AlgebraI		0.279
		AlgebraII		

Figure 2.5 (3c) – Problem 24				
Group	Means (%)	p-values		
Calculus	22.4		AlgebraI	AlgebraII
Algebra version I	7.8	Calculus	<0.001	< 0.001
Algebra version II	10.5	AlgebraI		0.166
		AlgebraII		

Figure 2.6 (3d) – Problem 25				
Group	Means (%)	p-values		
Calculus	17.8		AlgebraI	AlgebraII
Algebra version I	9.2	Calculus	<0.001	< 0.001
Algebra version II	9.4	AlgebraI		0.909
		AlgebraII		

same as the version used by the calculus-based students, performing worse than the group using version II of the problem set). In fact, if the performance of the algebra-based students on the two versions of the problem set shown in Figure 2.2 is taken as a predictor, the calculus-based students may have performed even better if they were given version II of the problem set.

We also analyzed the kinds of categories students created for each problem. Figures 2.3 through 2.6 compare some of the common categories for selected problems by the three introductory physics groups. Table 2.6 gives the mean percentage of categories placed by each group into good categories and the p-values between the groups respectively for Figures 2.3 through 2.6. To come up with the p-values, we summed the number of counts for all good categories and invoked a chi-squared test using a binary distribution based on whether or not group participants selected good categories for these questions. Calculus-based students again categorize significantly better on individual problems, but the difference between the two algebra-based groups on these problems is not statistically significant.

It should be noted that for all four Figures 2.3-2.6, the "conservation of energy" column is a subset of the "all energy" column. Similarly, the "conservation of momentum" category is a subset of the "all momentum" column. In support of Table 2.3 and Figure 2.2, one can see from Figures 2.3-2.6 that the calculus-based group performed more expert-like categorization overall than the two algebra-based groups. Problems 5, 24 and 25 in version I of the problem set featured in Figures 2.3, 2.5, and 2.6 are challenging problems because they combine two major physics principles. These problems were also examined in a previous study (Yerushalmi et al. 2007, Singh et al. 2007). Very few introductory physics students in any group categorized these problems according to both physics principles required to solve each of these problems. Some students (more in the algebra-based course than in the calculus-based course) created the categories "speed" and "kinetic energy" if the question asked them explicitly to calculate those physical quantities (rather than base their categorization on physical principles, e.g. conservation of mechanical energy or conservation of momentum). The explanations provided by the students as to why a particular category name, for example, "speed," is most suitable for a particular problem were not adequate; they wrote that they created this category because the question asked for the speed.

Figures 2.7-2.9 examine students' categorizations of all 25 problems in version I in select categories ("pendulum", "spring" and "ramp"). Students in the calculus-based courses were less likely to categorize problems in "pendulum" and "spring" categories but categorization in the "ramp" category is quite common amongst students in both the algebra-based and calculus-based courses. It is evident that, for version I of our problem set, students were likely to choose a "spring" category mainly for problem 6 (see Figure 2.7), which was a problem that was removed from version II of the problem set given to the other algebra-based introductory physics course. The "pendulum" category (see Figure 2.8) was also restricted to a few problems, but many problems involved ramps so that the "ramp" category is distributed across many problems (see Figure 2.9).

Some statistical p-value comparisons for Figures 2.7-2.9 (4a-4c) may be found in Table 2.7. We followed a similar method to the one posed in Table 2.6 with the net percentage of categories over all significant problem numbers replacing the net percentage of good and bad categories. We take into account all problems for which the number of students who placed a problem in the given category above a statistical noise level. This noise level is defined by the standard error bar for a null bin (zero respondents for a particular question number). The means

Table 2.7. Select p-values between calculus-based and algebra-based students for select problems placed into surface feature categories.
See Figures 2.7-2.9.

Figure 2.7 (4a) – "pendulum" (4, 5, 6, 15, 25)		
Group	Means (%)	p-value
Calculus-based	3.4	0.001
Algebra-based	7.3	
Figure 2.8 (4b) – "spring" (6)		
Group	Means (%)	p-value
Calculus-based	29.4	<0.001
Algebra-based	51.8	
Figure 2.9 (4c) – "ramp" (3, 6, 8, 12, 15, 17, 18, 22, 24, 25)		
Group	Means (%)	p-value
Calculus-based	9.8	0.416
Algebra-based	10.7	

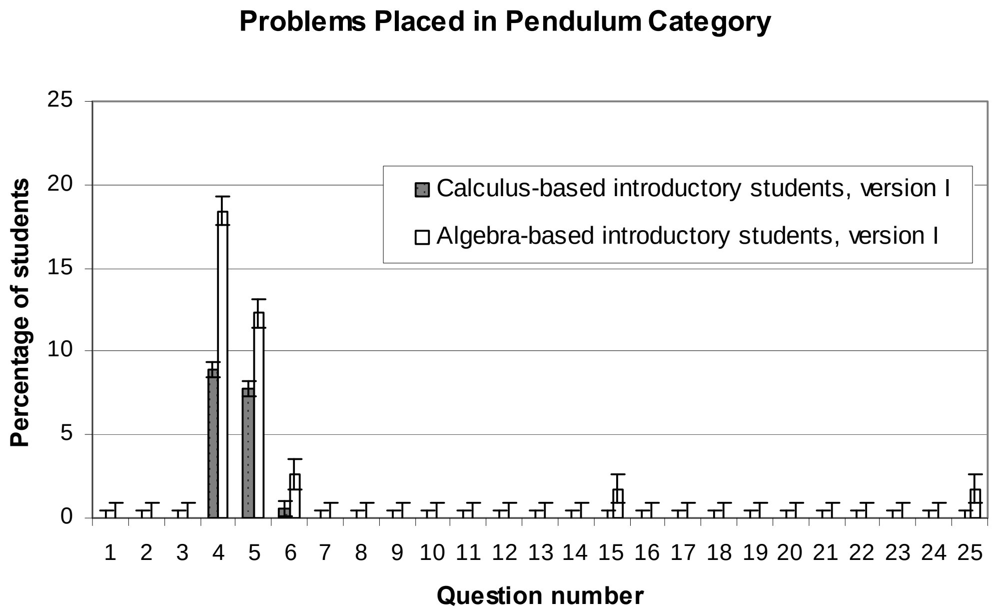


Figure 2.7. Version I introductory students and the "pendulum" category.
Histogram of the calculus-based and algebra-based introductory physics students (version I of the problem set) who categorized various percentages of problems 1-25 in the "pendulum" category when asked to group them based on similarity of solution. Note that the students in the calculus-based courses were less likely to classify these problems in this category than the students in the algebra-based courses.

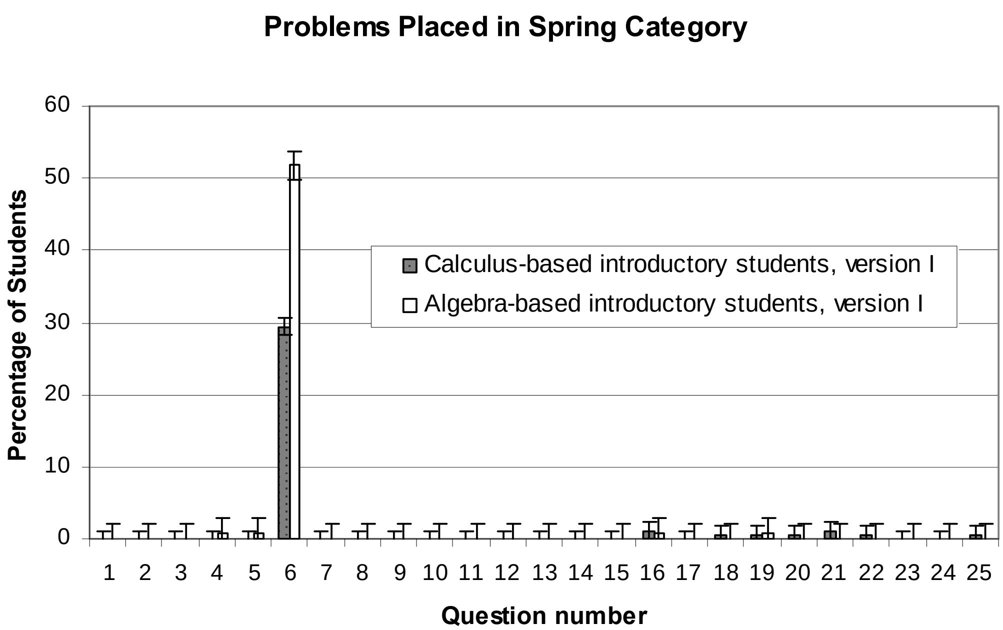


Figure 2.8. Version I introductory students and the "spring" category.
Histogram of the calculus-based and algebra-based introductory physics students (version I of the problem set) who categorized various percentages of problems 1-25 in the "spring" category when asked to group them based on similarity of solution. Note that the students in the calculus-based courses were less likely to classify these problems in this category than the students in the algebra-based courses.

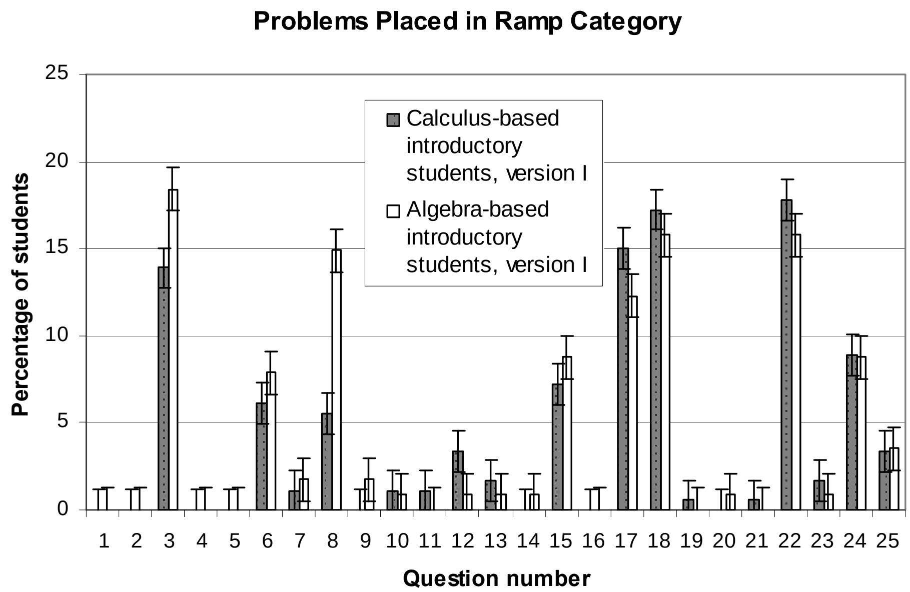


Figure 2.9. Version I introductory students and the "ramp" category.
Histogram of the calculus-based and algebra-based introductory physics students (version I of the problem set) who categorized various percentages of problems 1-25 in the "ramp" category when asked to group them based on similarity of solution. There is no clear difference between the students in the calculus-based courses and those in the algebra-based courses.

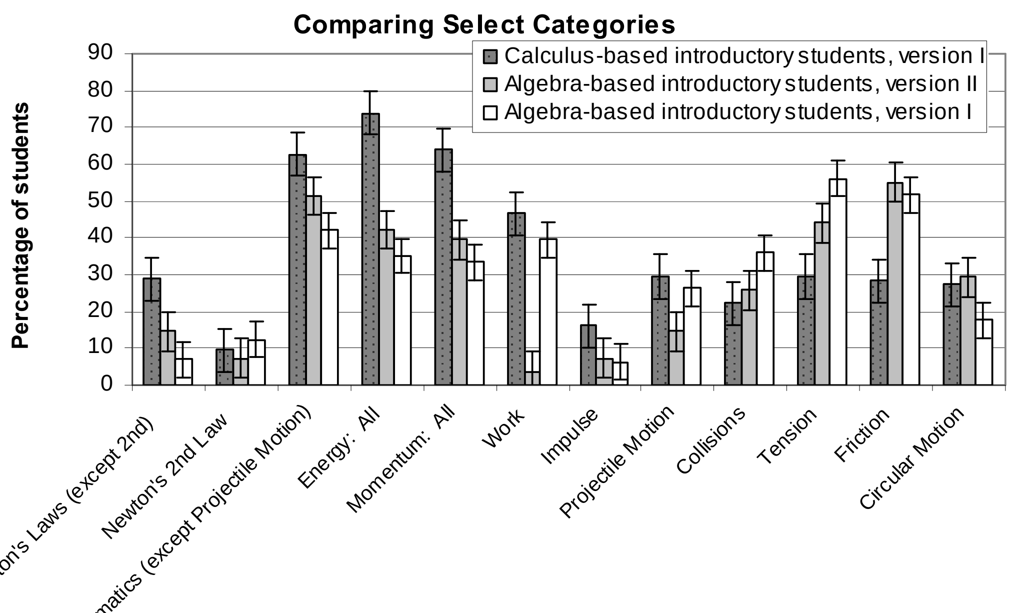


Figure 2.10. Introductory students and selected frequently chosen categories.
Histogram of the percentage of calculus-based and algebra-based introductory physics students (the algebra-based classes used two different versions of the problem set) who chose popular categories when asked to group them based on similarity of solution. Note that the students in the calculus-based courses were more likely to choose categories based upon fundamental physics principles than the students in the algebra-based courses.

are defined for all included problems, which are indicated in parentheses in Table 2.7. We find that the calculus-based students choose the "spring" and "pendulum" categories less frequently than algebra-based students do; however, this is not the case for the "ramp" category, for which the two groups of students are within standard error of each other for most questions with significant responses.

Figure 2.10 shows histograms of the percentages of introductory physics students in the three groups who chose selected categories by placing at least one problem in that category. The figure shows that the calculus-based group is more likely to use more expert-like categories related to kinematics, energy and momentum.

Table 2.8, done in the same style as Table 2.6, shows percentage means and t-tests across "good" categories (from "Newton's Laws" to "Projectile Motion" in Figure 2.10) and sheds further light on this with t-tests. The calculus-based students tended to choose "good" categories more often and "moderate" or worse categories less often than algebra-based students.

Table 2.8. Percentages and p-values between calculus-based and algebra-based students for placing problems into select common categories.
See Figure 2.10.

Figure 2.10 (5) – all "Good" categories				
Group	Means (%)	p-values		
Calculus	41.2		AlgebraI	AlgebraII
Algebra version I	25.0	Calculus	0.001	< 0.001
Algebra version II	22.6	AlgebraI		0.244
		AlgebraII		

The categorization task is primarily conceptual in nature and does not require quantitative manipulations. One may therefore wonder why students in the calculus-based introductory physics courses performed more expert-like categorization than students in the two algebra-based courses. We note that the calculus-based course is predominantly taken by students who major in engineering and are required to have math skills; algebra-based introductory physics is mainly taken by health science and pre-medical students majoring in biology, neuroscience, psychology and other disciplines which do not require mathematical prowess (although scientific reasoning skills are highly desirable even for this group).

One possible explanation of the difference between these two groups is based upon students' scientific reasoning abilities. Even conceptual reasoning of the kind needed in expert-like categorization requires good scientific reasoning skills. Prior research has shown that the students in the calculus-based courses are better at conceptual reasoning and may be better at scientific reasoning skills which pertain to physics (Coletta and Phillips 2005, Coletta et al. 2007, Loverude et al. 2008, Meltzer 2002, Nguyen and Meltzer 2003). The better mathematical preparation and scientific reasoning skills of the calculus-based students may reduce the cognitive load while learning physics and these students may not use all their cognitive resources for processing information that is peripheral to physics itself and may have more opportunity to build a robust knowledge structure. If that is the case, students in calculus-based classes will be able to perform better on conceptual tasks such as categorization than those in the algebra-based courses whose knowledge structure may not be as robust and scientific reasoning not as developed. More research is needed to understand the difference in cognition that led to the differences between the performance of the algebra-based and calculus-based students on the categorization task.

2.4.3 Comparison of Calculus-Based Introductory Students with Physics Graduate Students

Figure 2.11 shows a histogram of the percentage of questions placed in good categories (not moderate or poor) and compares the average performance on the categorization task of 21 graduate students and 7 physics faculty members with the introductory students in the calculus-based introductory physics course. Although the categorization by the calculus-based introductory students is not on par with the categorization by the physics graduate students, there is a large overlap between the two groups (Singh 2009). We note that in the Chi study the experts were graduate students and not physics professors. Figure 2.11 suggests that many calculus-based introductory students are far from being “novices” when categorizing the problem sets as suggested in the Chi study and there is a wide distribution of performance. Average percentages and p-values of problems placed in good categories are given for all three groups in Table 2.9.

The huge overlap between graduate students (experts in the Chi study) and introductory physics students (novices in the Chi study) in Figure 2.11 suggests that it is not appropriate to label all introductory students “novices” and all physics graduate students “experts.” Table 2.10 shows examples of categorization by two graduate students that were considered reasonably good overall and that were not considered reasonably good overall, and Table 2.11 shows similar examples for calculus-based introductory students. Similar to Figure 2.11, Tables 2.10 and 2.11 also suggest that the categorizations made by both the graduate students and calculus-based introductory students have a distribution and labeling graduate students as “experts” and introductory physics students as “novices” is not appropriate.

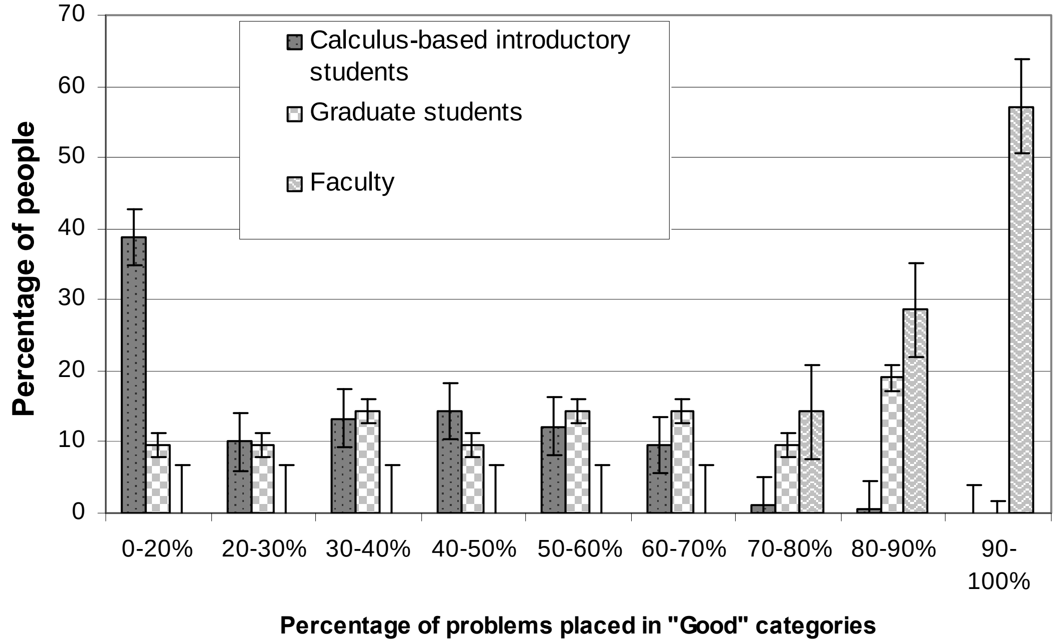


Figure 2.11. Comparison of calculus-based intro students to graduate students and faculty. Histogram of calculus-based introductory physics students, graduate students, and physics faculty who categorized various percentages of the 25 problems in version I in "good" categories when asked to categorize them based on similarity of solution. Physics faculty members performed best in the categorization task followed by graduate students and then introductory physics students, but there is a large overlap between the graduate students and the introductory physics students.

Table 2.9. Means and p-values for percentage of questions placed into a good category for calculus-based students, graduate students, and faculty.
See Figure 2.11.

Group	Means (%)	p-values		
Faculty	88.6		Graduate	Calculus
Graduate	52.4	Faculty	<0.001	< 0.001
Calculus	34.4	Graduate		<0.001
		Calculus		

Figure 2.11 shows that the difference in good categorization performed by the physics professors and physics graduate students is larger than the difference between graduate students and the calculus-based introductory physics students. Physics professors pointed out multiple methods for solving a problem and specified multiple categories for a particular problem more often than the graduate students and introductory students. Professors created secondary categories in which they placed a problem that were more like the introductory students' (and some graduate

Table 2.10. Examples of overall good and poor categorization by four graduate students. The category names in parentheses refer to those that are either moderate or poor categories. The problem numbers in parentheses show problems that are placed in an inappropriate category although the category is itself a good category. A problem number in boldface refers to a problem that should have been placed in two primary categories but was placed only in one category (or the categorization is incomplete). The explanations given by students for choosing a particular category are not reproduced here.

Example of categorizations by a graduate student considered reasonably good **overall**

Category	Problem numbers of problems placed in a category
Conservation of Energy	6, (7), ***8***, ***12***, 15, 18, (22), (23), ***24***, ***25***
Forces	(3), **5**, 10, 13, 17, (19)
1 and 2 Dimensional Motion	2, 9, 14, 16, (19), 20, 22
Conservation of Momentum	1, ***4***, 11, (21), (23)
(Circular Motion)	5, 10, 25

Example of categorizations by a graduate student considered reasonably good **overall**

Category	Problem numbers of problems placed in a category
Conservation of Energy	4, **5**, 6, ***8***, ***12***, 15, 18, (22), ***25***
Impulse	(18), 21
Kinematics	2, 7, 9, 14, 16, 20, 22, 23
Conservation of Momentum	1, 4, 11, ***24***
Force Diagram, Newton's 2nd Law	7, 10, ***12***, 13, 17, 23
Projectile Motion	2, 9, 14, 16, 20
Circular Motion, Rotational Kinematics	***5***, 10, ***25***
Work-Energy	3

Example of categorizations by a graduate student **not** considered good **overall**

Category	Problem numbers of problems placed in a category
(Center of Mass)	1
(Circular Motion)	5, 10
Momentum and Energy Conservation	(6), (11), (21), (22)
Energy Conservation	***3***
Impulse	19
Kinematics	9, 14, (15), 16
Momentum Conservation	***4***, ***24***
Newton's 2nd Law	13, 17
Projectile Motion	2, 9, 14
Relative Velocity	***1***
(Vector Resolution)	18, 20
Work-Energy Equivalence	7, 23
Problems omitted by the student	8, 12, 25

Example of categorizations by a graduate student **not** considered good **overall**

Category	Problems placed in a category
(Comprehensive Problem)	4, 8, 15, 21, 24, 25
Dynamics	(2), **5**, ***7***, (9), 10, 13, (14), (16), 17, (18), (19), (20), (22), ***23***
Energy	2, **3**, 6, 7, 12, 18, (19), 20, (22)
Momentum	1, 11

Table 2.11. Examples of overall good and poor categorization by four introductory students. Examples of categorization by four calculus-based introductory students, some of which were considered reasonably good overall and some that were not considered good overall. See Table 2.9 for explanation of numerical notation.

Example of categorizations by a calculus-based introductory student considered reasonably good **overall**	
Category	Problems placed in a category
Momentum	1, ***4***, 11, 19, ***24***
Kinematics	2, 9, 14, 16, 20
Work	3, 7, 23
Free body diagram (Newton's Laws)	***5***, 10, 13, 17, (21), (22)
Conservation of energy	6, ***8***, 12, 15, 18, ***25***
Example of categorizations by a calculus-based introductory student considered reasonably good overall	
Category	Problems placed in a category
Collisions	***4***, ***8***, ***19***, ***21***
Conservation of Energy	2, (9), 15, 16, 18, 20
Linear Motion	14, 22, ***23***
Conservation of Momentum	1, 11, ***24***
Newton's Laws	***5***, 10, (12), 13, 17, ***25***
Work	3, 7
Example of categorizations by a calculus-based introductory student not considered good overall	
Category	Problems placed in a category
(Tension)	3, 4, 5, 17
(Things being shot/dropped)	2, 9, 14, 16, 20, 21
2-D Movement	(1), (7), (8), (13), (19), (23)
(Spring)	6
(Mass changing)	1, 11
(Slopes/Inclined Planes)	12, 15, 17, 18, 22, 24, 25
Rotation	10
Example of categorizations by a calculus-based introductory student not considered good overall	
Category	Problems placed in a category
Momentum	1, 11
(Velocity X & Y components)	2, 9, 14, 16, 18, 19, 20, 22
(Weight)	1, 3, 7, 8, 11, 13, 23, 24
(Pendulum)	4, 5
(Spring)	6
(Centripetal Motion)	10
Energy	12, 15, 18, (22)
(Tension)	17
(Contact Forces)	21

students') primary categories. For example, in the questions involving tension in a rope or frictional force (see Appendix A), many faculty members created secondary categories called "tension" or "friction," but also placed those questions in a primary category, based on a fundamental principle of physics. Introductory physics students and even graduate students were much more likely to place questions in inappropriate categories than the faculty, e.g. placing a problem that was based on the impulse-momentum theorem or placing conservation of momentum in the conservation of energy category. Some questions involved two major physics principles, e.g. question 4 related to the ballistic pendulum in version I of the problem set. Most faculty members categorized it in both the "conservation of mechanical energy" and "conservation of momentum" categories in contrast to most introductory students (and many graduate students) who either categorized it as an energy problem or as a momentum problem. The fact that most introductory students (and many graduate students) only focused on one of the principles involved to solve question 4 in version I is consistent with an earlier study in which students either noted that this problem can be solved using conservation of mechanical energy or conservation of momentum but not both (Singh and Rosengrant 2003).

Many of the categories generated by the faculty, graduate students and all three groups of introductory physics students were the same, but there was a difference in the fraction of questions that were placed in good categories by each group. What introductory students, especially those in the algebra-based courses, chose as their primary categories were often secondary categories created by the faculty members. Rarely were there secondary categories made by the faculty members, for example, a secondary category called "apparent weight," that were not created by students. There were some categories, such as "ramps" and "pulleys," that were made by introductory physics students but not by physics faculty or graduate students. Even if a problem did not explicitly ask for the "work done" by a force on an object, faculty were more likely to create and place such questions which could be solved using the work-energy theorem or conservation of mechanical energy in categories related to these principles. This task was much more challenging for the introductory physics students who had learned these concepts recently, especially those in the algebra-based courses (and even for some graduate students). For example, in version I, it was easy to place question 3 in a category related to work because the question asked students to find the work done on an object, but placing question 7 in the "work-energy" category was more difficult because students were asked to find the speed.

2.4.4 Correlation between Categorization and Course Grade

Figures 2.12 and 2.13, respectively, examine how the percentage of good categories created by students and the percentage of problems placed in good categories are correlated with the cumulative score in the course for the students in the algebra-based course who categorized version I of the problem set. Figures 2.14 and 2.15, respectively, examine how the percentage of good categories created by students and the percentage of problems placed in good categories are correlated with the final exam score in the course for the same class. Although there is a positive correlation in all of these graphs, it is weak, which implies that we cannot easily infer students' ability to categorize problems based upon their course performance in a traditional course. This may partly be due to the fact that the assessment in a traditional course does not necessarily emphasize the kinds of skills that are assessed by the categorization task.

2.4.5 Interviews with Individual Students

To get an in-depth understanding of students' thought processes while categorizing the problems and to analyze the extent to which students could outline the procedure for solving the problems

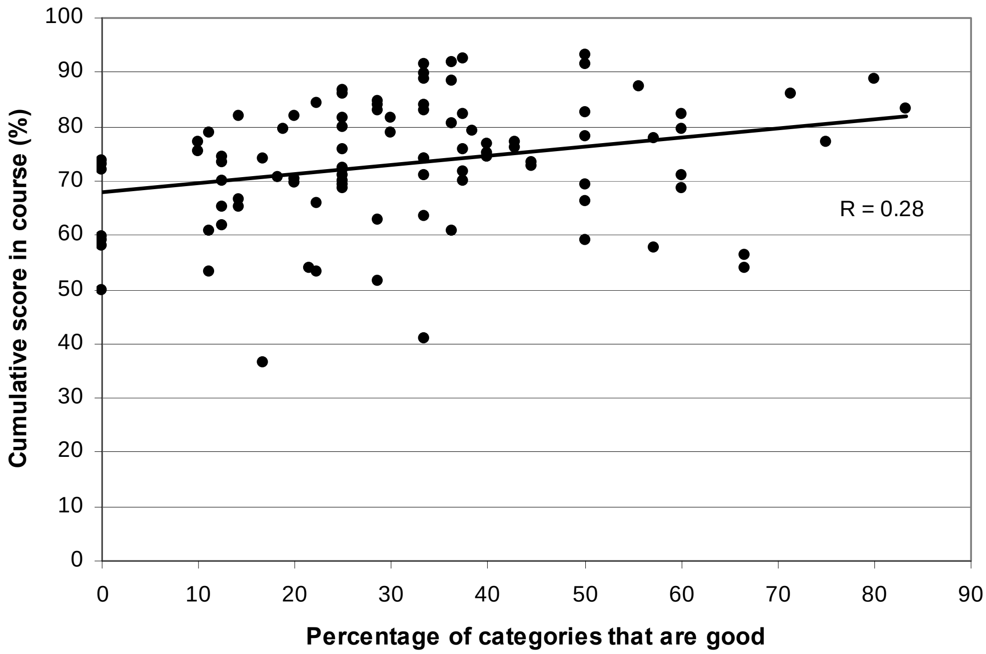


Figure 2.12. Cumulative score in the course vs. the percentage of categories that were considered good for the algebra-based introductory students who were given version I of the problem-set to categorize.

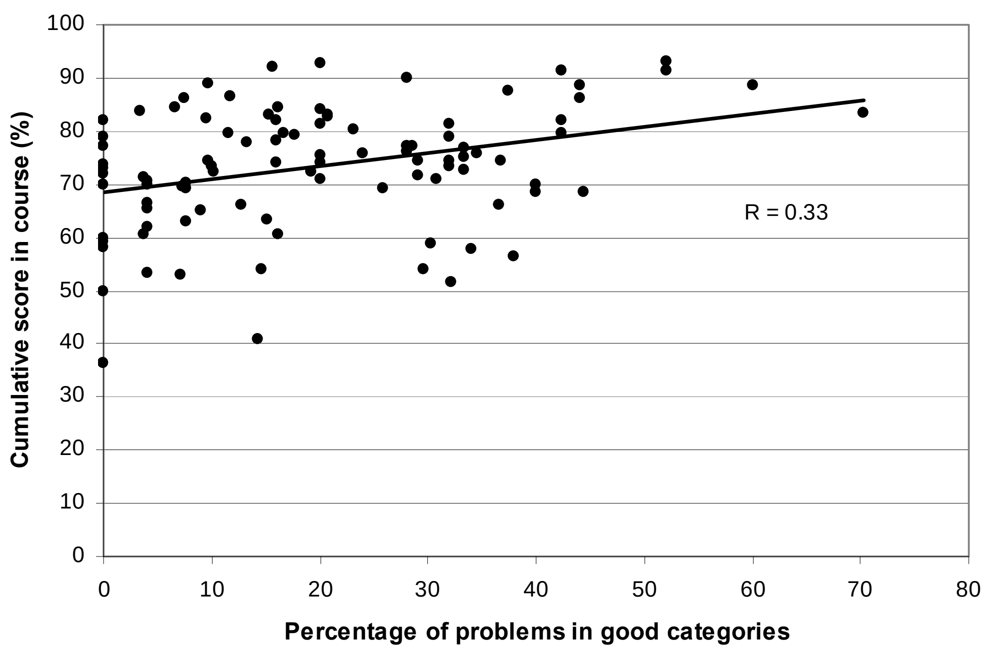


Figure 2.13. Cumulative score in the course vs. the percentage of problems that were placed in good categories for the algebra-based introductory students who were given version I of the problem-set to categorize.

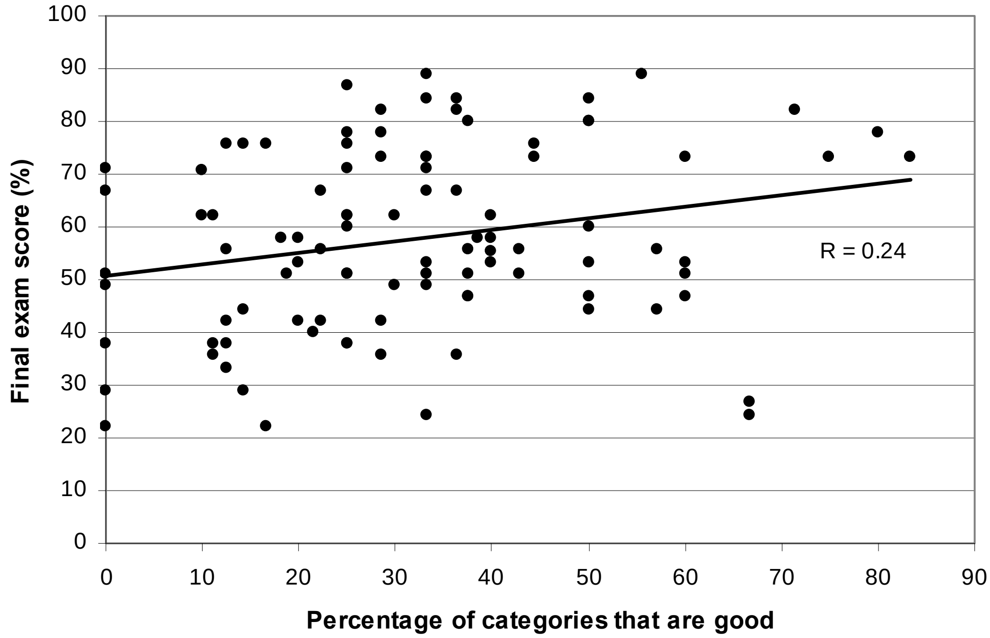


Figure 2.14. Final exam score vs. the percentage of categories that were considered good for the algebra-based introductory students who were given version I of the problem-set to categorize.

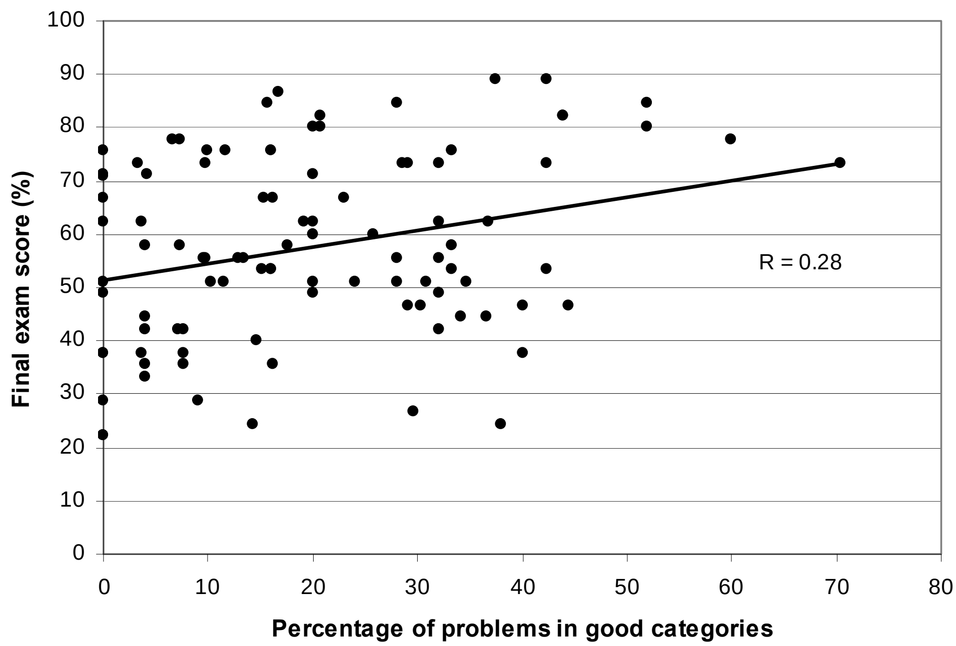


Figure 2.15. Final exam score in the course vs. the percentage of problems that were placed in good categories for the algebra-based introductory students who were given version I of the problem-set to categorize.

they categorized, we interviewed four student volunteers (referred to as students 00, 41, 60, and 62 from the ID numbers used during analysis) individually from the algebra-based introductory physics II class that was given version II of the problem set. All but one student had taken the introductory mechanics course within a year. The attitude of all the four students towards the class was generally favorable or indifferent and all of the students were majoring in a biology-related field.

The procedure of the interviews was as follows. First, the students stated aloud their thought process while categorizing all of the 25 problems. Next, the students were asked to outline a solution procedure for some problems selected from the Chi problems and others that were not from the Chi study. The Chi problems and non-Chi problems were chosen to possess similarity in surface features and/or deep-features as much as possible.

The four algebra-based introductory students who were given version II of the categorization problem set during the interview had a mixed set of categories. They could recall some concepts and create some categories that were based upon physics principles, but they also chose some categories that focused on the surface features of the problems. There was some evidence (e.g., with the "friction" category) that some of these moderate categories were not entirely due to students' failure to realize that these are problems involving other forces that can be solved, e.g., using Newton's second law. But students sometimes deliberately chose more novice-like categories instead of more expert-like categories based upon physics principles because they felt the need to address specific details as opposed to the general physical principles. Upon asking for clarification, two students (62 and 41), who categorized problems in categories involving specific forces, e.g., friction or tension, mentioned that they preferred these categories to the more general category of force because they found the term "force" to be vague and the more specific description of different forces removed ambiguity. There is also evidence that all expert categories are not chosen because they may be seen as superfluous in light of other categories that are already created by a student (e.g., student 60 decided not to create a kinematics category because he already had an energy category in which he had placed the problems that he would have placed in the kinematics category).

A physics professor was also specifically asked why he had placed a problem only in a "Newton's second law" category or a "work-energy theorem" category as opposed to additionally including the specific forces or definition of work in his categorization. He responded that he thought that the task was about categorizing problems based upon the laws of physics and procedures for solving problems, and while the essential knowledge of forces and definitions of work were important to solve the problem, they were not the most fundamental issues that made the solution to the problems similar.

The seven problems taken from the Chi study were generally poorly classified on average in comparison to the other 18 problems in version II. In addition, one student (00) took a significantly longer time to classify the Chi problems than the other 18 problems. Problems 10 and 11 were generally placed together in an "angular speed and angular momentum" category, although problem 10 can be solved using rotational kinematics and problem 11 using the conservation of angular momentum. Another Chi problem, Problem 16, was placed by itself in a spring category by three students; the other student (41) placed this problem in a "frictional force" category. The other Chi problems were generally placed in categories that were either based upon a surface feature or a secondary category such as "frictional" or "tension force." The greater difficulty in categorizing Chi problems is consistent with the data for 109 students shown in Figure 2.1. It is possible that the Chi problems were overall more difficult for students than the other problems. For example, several Chi problems are solved most directly by using the work-energy theorem (or a combination of Newton's laws and kinematics). As can be seen from Table

2.1, very few students overall from this group selected “work” as a category in comparison to the groups which categorized version I of the problem set, and the closest the four interviewed students came to recognizing related principles was a reference to converting potential energy into kinetic energy by student 60.

After the students had categorized the problems, they were asked to outline the procedure towards a solution for five Chi problems and five non-Chi problems. For outlining the procedure, the five Chi problems chosen from version II of the problem set were problems 11, 14, 15, 16, and 18. These Chi problems were taken verbatim from the textbook by Halliday and Resnick. Problem 11 can be solved using the principle of angular momentum conservation. Problems 14 and 15 were two problems with identical surface features that asked for a solution of different quantities. Problem 18 had a block on an inclined plane and can be solved using the work-energy theorem (or Newton’s second law and kinematics) in the presence of friction. Problem 16 describes a block forced against a horizontal spring on a table with friction and can also be solved using conservation of energy and the work-energy theorem. Incidentally, physics faculty members who were given these problems pointed out that problem 14 will be clearer if the man “started from rest” and problem 18 did not mention the coefficient of static friction which was relevant for determining whether the block will come down from the highest point on the inclined plane where it is momentarily at rest. None of the students noticed these deficiencies in the problems because noticing them requires higher level of expertise.

Two of the non-Chi problems for which the students were supposed to outline the procedures were problem 4 (problem 4 can be solved using Newton’s second law similar to problem 15) and problem 7 (a surface feature of problem 7 is an inclined plane similar to problem 18 but problem 7 can also be solved using Newton’s second law). Another non-Chi problem selected was problem 25, which required multiple principles and is procedurally complex similar to problem 18. Another non-Chi problem selected was problem 19 for its unique astronomy-related surface features to see how students worked around them. This problem can be solved using the conservation of angular momentum of the system similar to the Chi problem 11. Problem 23 was another non-Chi problem selected which involved a person swinging and required two principles to solve (conservation of mechanical energy and Newton's second law).

The interview lasted for over an hour and students 00 and 41 could not outline the procedures for all 10 problems due to other obligations. Student 00 seemed to focus heavily upon drawing inferences directly from equations (an equation sheet was provided). When asked why he immediately looked for equations without thinking conceptually, he stated that he believed the course to be "95% math" and not much conceptual work. Student 00 also seemed to have trouble with problems that required consideration of Newton's second law (particularly problems 4, 7, 15) and the conservation of mechanical energy (problem 14). He stated that he disliked problems with pulleys and inclines, as he found the solutions to be messy. As in the categorization, student 62 started out moderately well – she recognized that she should not use rotational kinematics for problem 11 and was able to make some connections respectively between problems 4 and 7 (free body diagram and Newton's second law). However, later she reverted to more of a "plug-and-chug" approach based upon finding the right equation and couldn't address problems 16, 18, 19, 23 and 25 well. Student 41 demonstrated a very methodical approach in outlining the procedures as he did for the first section. While he seemed to make some conceptual connections with most of the problems, he did not remember the material well. Student 60 was able to correctly approach problem 4 and was able to figure out elements of the solution for some problems by using Newton's second law or energy conservation (e.g. problem 25), although once in a while (e.g. problem 18) he started searching for an equation to plug and chug without planning what he was doing. Interestingly, he seemed fairly concerned about whether he had gotten correct answers for several of the questions and categories and asked for an explanation for the questions later.

Overall, all four students struggled with the categorization process and with the outlining of the procedure for the selected problems. They would often cite approaches that would not work well for the problem, or state an approach that was only partially correct (e.g. using either conservation of energy or Newton's second law, but not both, for problems 23 and 25). Sometimes, students could state that two problems should be placed in the same category because they should both use the same solution method but they had difficulty with correctly outlining the procedure. For example, in Newton's second law problems, they had difficulty with forces, e.g. they often made mistakes in drawing the free-body diagrams with all relevant forces. Also, students had a tendency to search for formulas that would solve the problems even without performing a conceptual analysis of the problem.

2.5 SUMMARY AND CONCLUSIONS

We revisited the Chi categorization study three decades later at the same institution (University of Pittsburgh) in the "ecological" classroom environment with several hundred introductory physics students. We asked introductory students in the calculus-based and algebra-based physics courses to categorize 25 introductory mechanics problems based upon similarity of solution. Two versions of the problem sets were used in the study, with one version developed later including the seven problems that can be inferred from the Chi study. The later version was developed because version I (that did not include Chi problems) showed significant deviation from the dichotomy of introductory physics students as “novices” and physics graduate students as “experts”. We find that students’ ability to categorize problems depends upon the nature of the questions asked but the overall qualitative trends were not strongly dependent on the version of the problem set given to the students. It is not possible to compare our data directly with that in the Chi study because most of their questions are no longer available. Although no direct comparison is possible, in both versions of the problem sets we used, the percentage of introductory physics students who chose categories such as “ramp” and “pulley” was significantly lower than the percentages reported in the Chi study.

Moreover, the categorization performed by the calculus-based introductory physics students was significantly better than that performed by the algebra-based introductory physics students. Also, there was a large overlap between the categorization performed by calculus-based introductory physics students and the graduate students in physics that were assessed “good”. This large overlap in the performance of the two groups suggests that, unlike the characterization in the Chi study, there is indeed a wide distribution of expertise as assessed by the categorization task in both groups and it is not appropriate to classify all introductory physics students as “novices” and all physics graduate students as “experts”. We also find that the categorization performed by physics faculty members was significantly better than that performed by the graduate students. We note that, in the Chi study, the comparison was between the categorization performed by 8 introductory physics students (labeled novices) and 8 physics graduate students (labeled experts).

While there is a large overlap between the calculus-based introductory students and graduate students in terms of the percentage of categorization that was assessed as “good”, there are different reasons for each group for the overlap. Unlike the introductory students, the category names of the physics graduate students never included “ramps” or “pulleys.” In this sense, graduate students were more sophisticated than the introductory students in choosing their categories. The categorization by the graduate students was often not assessed “good” for other reasons. Some examples of this are if they only placed a problem in a secondary category (e.g. “centripetal force”), placed a problem that involved two physics principles in only one principle category, misplaced a problem (e.g. placed a problem related to “impulse and momentum” concepts in an energy-related category), or placed problems in a category such as “velocity” instead of grouping problems based upon the physics principles. However, we note that, on an average, the graduate students were more likely than an introductory physics student to place a problem involving two physics principles correctly in both categories.

In the future, it will be useful to investigate how the introductory students’ and graduate students’ categorization will differ if they were given the names of the categories they could choose from (which would include both poor categories such as ramps and pulleys and good categories based upon the laws of physics) but were told that they need not use all of them and could even come up with their own categories. Future investigation will also explore the similarities and differences in introductory students’ and graduate students’ responses if they were asked to solve the problems or at least outline the solution procedure rather than only being asked to do the categorization. While our interviews with the introductory physics students asked them to outline the solution procedure, it will be revealing to carry out similar interviews with graduate students. We hypothesize that the graduate students will perform better at outlining the solution procedure for the problems the than the calculus-based introductory students, but the amount of overlap between the two groups will be useful to analyze with regard to student development within a curriculum.

Finally, we postulate that inclusion of categorization tasks in instruction can enhance learning. The categorization task focuses on higher-order thinking skills and asks students to evaluate the similarity of solution despite the fact that the problems placed in the same category may have very different surface features (Schoenfeld 1985, Schoenfeld 1989, Schoenfeld 1992, Schoenfeld and Herman 1982, Bransford and Schwarz 1999, Newell 1990, National Research Council 1999a and 1999b, Simon and Kaplan 1989, Anderson 1999). Such activity rewards conceptual analysis and planning stages of quantitative problem solving and discourages the plug-and-chug approach. Without guidance, students often jump into implementing a problem solution without thinking whether a particular principle is applicable. A categorization task can be used as a tool to help students learn effective problem solving and organize their knowledge hierarchically because such a task can guide students to focus on the similarity of problems based upon the underlying principles rather than focusing on the specific contexts (Hardiman et al. 1989, Dufresne et al. 1992). One instructional strategy for incorporating a categorization task is to give such a task to small groups of introductory physics students with different levels of expertise. One could ask students to categorize problems based upon similarity of solution, and then discuss and debate why different problems should be placed in the same group without asking them to solve the problems explicitly. Then, there can be a class discussion about why some categorizations are better than others and students can be given a follow-up categorization task individually to ensure individual accountability.

2.6 ACKNOWLEDGMENTS

We are very grateful to Jared Brascher for help in analyzing one set of introductory physics data. We thank F. Reif, R. Glaser, J. Levy and R. P. Devaty for useful discussions and to the faculty members who administered the categorization task in their classes for their help. We thank the National Science Foundation for financial support.

2.7 REFERENCES

Anderson, J. (1999). *Learning and memory: An integrative approach, 2nd Ed.* NY: Wiley.

Bransford, J., & D. Schwartz, D. (1999). "Rethinking transfer: A simple proposal with multiple implications." *Review of Research in Education* 24, 61-100.

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

Chi, M. T. H. (1997). In *Thinking Aloud* Chapter 1, Eds. Van Someren, Barnard, and Sandberg, 1, Erlbaum, 1994; M. T. H. Chi, "Quantifying qualitative analysis of verbal data: A practical guide." *J. Learning Sci.* 6(3), 271.

Chi, M. (2000). "Self-explaining expository texts: the dual processes of generating inferences and repairing mental models." In *Advances in Instructional Psychology.* Edited by R. Glaser. Mahwah, NJ: Erlbaum Associates, pp. 161-238.

Coletta, V., Phillips, J., & Steinert, J. (2007). "Interpreting force concept inventory scores: normalized gain and SAT scores." *Phys. Rev. ST-PER* 3, 010106 (1-5).

Coletta, V. & Phillips, J. (2005). "Interpreting FCI Scores: Normalized Gain, Pre-instruction Scores, and Scientific Reasoning Ability." *Am. J. Phys.*, 1172-1182.

Dufrense, R., Gerace, W., Hardiman, P., & Mestre, J. (1992). "Constraining novices to perform expert-like problem analyses: Effects on schema acquisition." *J. Learning Sci.* 2, 307–331.

Dufresne, R., Mestre, J., Thaden-Koch, T., Gerace, W. & Leonard, W. (2005). "Knowledge representation and coordination in the transfer process." *In Transfer of learning from a* *modern multidisciplinary perspective.* Edited by J. P. Mestre. 393 pages, Greenwich, CT: Information Age Publishing, 155-215.

Hardiman, P., Dufresne, R. & Mestre, J. (1989). "The relation between problem categorization and problem solving among novices and experts." *Memory and Cognition*, 17, 627-638.

Halliday, D., & Resnick, R. (1974). *Fundamentals of physics, 2nd edition.* NY: Wiley.

Larkin, J. (1981). "Cognition of learning physics." *Am. J. Phys.* 49 (6), 534-541.

Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980). "Expert and novice performance in solving physics problems." *Science* 208, 1335-1342.

Loverude, M., Kanim, S. & Gomez, L. (2008). "Curriculum design for the algebra-based course: Just change the d's to deltas?" In *2008 Physics Education Research Conference.* Edited by C. Henderson, M. Sabella, and L. Hsu. *AIP Conf. Proc.* 1064, 34-37.

Maloney, D. (1994). *Research in Problem Solving: Physics, in Handbook of Research on the Teaching and Learning of Science.* Edited by D. Gabel. MacMillan.

McDermott, L. (2001). "Oersted Medal Lecture 2001: Physics Education Research: The key to student learning." *Am. J. Phys.* 69(11), 1127-1137.

Meltzer, D. (2005). "Relation between students' problem-solving performance and representational format." *American Journal of Physics*, 73, 463-478.

Meltzer, D. (2002). "The relationship between mathematics preparation and conceptual learning gains in physics: A possible 'hidden variable'". *Am. J. Phys.* 70(12), 1259-1268.

National Research Council (1999a), *How people learn: Brain, mind, experience and school*. Committee on developments in science on learning. Edited by J. Bransford, A. Brown and R. Cocking. Washington, DC: National Academy Press.

National Research Council (1999b). *How people learn: Bridging research and practice.* Committee on learning research and educational practices. Edited by M. Donovan, J. Bransford, and J. Pellegrino. Washington, DC: National Academy Press.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Nguyen, N. L. & Meltzer, D. (2003). "Initial understanding of vector concepts among students in introductory physics courses." *Am. J. Phys.* 71(6), 630-638.

Ozimek, D., Engelhardt, P., Bennett, A., & Rebello, N. (2004). "Retention and transfer from trigonometry to physics." Edited by J. Marx, P. Heron, and S. Franklin. *AIP Conf. Proc.* 790, 173-176.

Schoenfeld, A. (1985). *Mathematical problem solving.* New York: Academic Press.

Schoenfeld, A. (1989). “Teaching mathematical thinking and problem solving.” In *Toward the thinking curriculum: Current cognitive research.* Edited by L. B. Resnick and B. L. Klopfer. Washington, DC: ASCD, pp. 83-103.

Schoenfeld, A. (1992). “Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics.” In *Handbook for research on mathematics teaching and learning.* Edited by D. Grouws. New York: McMillan, Chapter 15, pp. 334-370.

Schoenfeld, A., & Herrmann, D. (1982). “Problem perception and knowledge structure in expert novice mathematical problem solvers.” *J. Exp. Psych., Learning, Memory, and Cognition* 8, 484-494.

Simon, H., & Kaplan, C. (1989). *Foundations of cognitive science*. Edited by M. I. Posner. Cambridge, MA: MIT press.

Singh, C. (2002). “When physical intuition fails.” *Am. J. Phys*, 70(11), 1103-1109.

Singh, C. & Rosengrant, D. (2003). “Multiple-choice test of energy and momentum concepts.” *Am. J. Phys.* 71 (6), 607—617.

Singh, C. (2004). “Interactive video tutorials for enhancing problem solving, reasoning, and meta-cognitive skills of introductory physics students.” In *Proceedings of the Phys. Educ. Res. Conference, Madison, WI.* Edited by S. Franklin, K. Cummings and J. Marx. *AIP Conf. Proc.* 720, 177—180.

Singh, C., Yerushalmi, E. & Eylon, B. (2007). “Physics Learning in the context of scaffolded diagnostic tasks (II): Preliminary results.” *Proceedings of the Phys. Ed. Res. Conference, Greensboro, NC, AIP Conf. Proc.* 951, 31-34.

Singh, C. (2009). “Categorization of problems to assess and improve proficiency as teachers and learners.” *Am. J. Phys.* 77(1), 73-80.

Yerushalmi, E., Singh, C. & Eylon, B. (2007). “Physics learning in the context of scaffolded diagnostic tasks (I): The experimental setup.” *Proceedings of the Phys. Ed. Res. Conference, Greensboro, NC, AIP Conf. Proc.* 951, 27-30.

3.0 SELF-DIAGNOSIS IN INTRODUCTORY PHYSICS

3.1 ABSTRACT

The study discussed in this chapter has two parts. In both, cognitive apprenticeship was applied to a self-diagnosis exercise for introductory physics students in recitation sections. The first part of the study sought to understand how well students in an introductory physics course diagnosed their mistakes in a quiz problem with different levels of scaffolding support when the problem they self-diagnosed was unusually difficult. We examined issues related to transfer, in particular the fact that the transfer problem in the midterm that corresponded to the self-diagnosed problem was a far transfer problem. An assessment rubric was developed to assess students' work and will also be discussed. In the second part of the study we used a related intervention in which we repeated the study methodology with the same students in the same groups using a new quiz problem that was more typical for these students and a near transfer problem. We discuss how these changes affected students' ability to self-diagnose and transfer from the self-diagnosed quiz problem to a transfer problem on the midterm exam. Finally, we compare and contrast the results of the two parts of the study.

3.2 STUDY 1: SELF-DIAGNOSIS AND TRANSFER FOR AN ATYPICAL QUIZ

3.2.1 Introduction

The goal of this study is to shed light on the learning processes and outcomes associated with "self-diagnosis tasks", a type of *formative assessment* (Black and Wiliam 1998) tasks. Students are required to present a diagnosis of their mistakes in which they identify and explain the nature of their mistakes beyond simply correcting their quiz solutions. Self-diagnosis tasks aim to foster *intentional learning* (Bereiter and Scardamalia 1989) in the context of problem solving; students will not only focus their attention on the task (arriving at a solution to a specific problem), but also on more general learning goals such as elaboration of the solvers' conceptual understanding. The aim is to foster *self-explanations*, an activity shown to lead to significant learning gains in the context of problem solving (Chi et al. 1989, Chi et al. 1994). As with other formative assessment tasks, the aim is to promote an activity that emphasizes learning in the context of the exam through process and feedback.

The activity of self-diagnosis requires interpreting the possibly mistaken solution formerly composed by the student as a textual artifact from a new perspective. This artifact represents to some extent the mental model of the student when she/he first approached the problem on the quiz.

Students might gain a new perspective from their prior work merely by asking themselves reflective questions, trying to clarify what they did and why they did it. Indeed, it has been shown that expert problem solvers are characterized by continuous evaluation of their progress (Larkin et al. 1980, Maloney 1994), asking themselves implicit reflective questions such as: "What am I doing?", "Why am I doing it?", and "How does this help me?" (Schoenfeld

1992) Yet novice students might need external support in learning to ask such questions. Examples of such support include a solved example from the instructor or a diagnosis rubric focusing their attention on possible mistakes.

In this study, we focused on three self-diagnosis tasks that varied in the external support students received via instruction and resources. In one task, students received only minimal guidance, and were allowed to use their notes and textbooks to diagnose their errors. In the second task, students diagnosed the mistakes in their solution after the teaching assistant had presented an outline of the solutions and against a rubric reflecting general problem solving steps (including steps such as "problem description", "plan", "evaluate") common to several problem solving strategies described in the research literature (Reif 1994, Heller et al. 1992, Heller and Hollabaugh 1992). In the third task, students compared their own solution with a fully written worked out example of the same problem.

A worked out example reflects the mental model of the instructor. Thus, when students self-diagnose their solution assisted by a worked out example, they need to relate their own mental model, as represented by their former solution, to the mental model of the instructor as they interpret it when reading the instructor's worked out example and construct a new understanding. In Vygotsky's terms, following negotiation with the artifacts provided by the instructor, the students had to change their situation definition (Wertsch 1985) for the problem. A situation definition for the problem might include understanding what the target variable is that one has to find, which physics concepts and principles might be invoked and applied to solve the problem, or how the reasoning in the solution should be presented. Clearly, the cognitive load in such a process is high.

One should consider whether a situation definition is within the zone of proximal development of the students, which Vygotsky defined as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or collaboration with more capable peers” (Vygotsky 1978). The tasks that students perform should accordingly engage their cognitive activity to meet this potential development level to allow a productive process of learning to take place. One might also expect that students' ability to self-diagnose would depend on their prior knowledge related to the problem.

In order to study the extent to which the self-diagnosis task is within the zone of proximal development of the students we focused on two contexts: one has a non-typical problem situation while the other has a typical one. In these contexts we study the following questions:

- How well do students self-diagnose and correct the mistakes in their solutions in each different self-diagnosis task?
- How do students' self-diagnosis attempts depend on prior problem-solving achievements?
- What is the effect of students' self-diagnosis in different tasks on subsequent problem solving?

In this section we focus on the non-typical problem situation. Section 3.3 will focus on comparing the non-typical problem situation with a typical one. We will first review research findings that constitute the rationale for developing self-diagnosis tasks. We will follow by describing the experimental setup: The different self-diagnosis tasks, and the timeline for the different treatment groups. Before presenting our findings we will describe special analysis tools we developed to assess students' self-diagnosis. A discussion of both experiments may be found in section 3.4.

3.2.2 Rationale for Self-Diagnosis Tasks

The idea that learning is promoted by reviewing one's own solutions is abundant among physics instructors (Yerushalmi et al. 2007). It is common practice to ask students to correct the mistakes in their solutions or compare them to a worked out example. This idea is supported by research comparing expert and novice problem-solvers in two contexts: the process of self-monitoring while solving a problem, and the process of learning through self-explanations while interacting with physics problem solutions (Chi et al. 1989).

As mentioned in section 3.2.1, one feature that characterizes expert problem solvers is continuous evaluation of their progress (Larkin et al. 1980, Maloney 1994) with implicit reflective questions (Schoenfeld 1992). Experts are also recognized as using a strategic approach: first carrying out a qualitative analysis of the situation and then developing a plan to solve the problem on the basis of this analysis (Polya 1945, Reif et. al. 1976, Larkin 1979, Larkin and Reif 1979, Larkin 1980, Heller and Reif 1984, Maloney 1994). In contrast, Schoenfeld (1992) found that novices often start solving a problem by quickly choosing an approach and then sticking with that approach even if it turns out to be unfruitful. They are also not likely to evaluate their final answer (Larkin 1980, Maloney 1994, Reif 1995). While self-monitoring is directed mainly towards arriving at a solution, it might also involve self-diagnosis directed towards more general learning goals. One would expect such diagnosis to result in elaboration of the solver's conceptual understanding, knowledge organization and strategic approach.

Chi et al. (Chi et al. 1989, Chi and VanLehn 1991) analyzed the explanations students provide when trying to make sense of the reasoning underlying the derivation of worked out examples in standard textbooks. Chi referred to these explanations as self-explanations (Chi 2000). It was found that students who self-explain more while studying from a textbook learn more, even though their self-explanations are fragmented and sometimes incorrect. Chi proposes two central mechanisms for constructing self-explanations: generation of inferences and self-repair (Chi 2000). These mechanisms allow both to fill in the gaps in the incomplete text as well as allowing individuals to repair their own mental representations. We hypothesize that, as in the study of worked out examples, in the act of reviewing and correcting one's own solutions, students will generate self-explanations. Moreover, one would expect that in this context, self-repair and inference generation processes may lead to changes in mental models, and that a student who self-explains more will learn more.

While there is a similarity between these two contexts, there is also a difference that may change the characteristics of the learning processes and outcomes. In self-explaining a worked out example, somebody else has constructed the artifact that the learner reads. Thus, the text acts as a mediator between the mental model of the expert and that of the learner. As a result, there is a need to negotiate between the mental model of the learner and the intention of the writer that may impose additional cognitive load. The learners have to relate their own knowledge to the product of a process that has been carried out by somebody else. In the context of self-explaining an artifact which the learners produce themselves, they interact with the outcomes of processes that they themselves have carried out, and thus they can interpret them more directly. Another factor is the correctness and coherence of the artifact. One would expect a solution produced by an expert to be more correct and coherent than a solution produced by the student.

It has been shown that the activity of self-explanation can be enhanced through interventions that require students to present their explanations (Chi et al. 1994, Bielaczyc et al. 1995). In physics education, a common instructional approach to enhance the generation and quality of self-explanations is that of cognitive apprenticeship (Collins et. al. 1989). Students work collaboratively (Cummings et. al. 1999, Heller and Hollabaugh 1992, Heller et. al. 1992, van Heuvelen 1991, Mestre et. al. 1993, Leonard et al. 1996) or with a computer (Reif and Scott 1999, VanLehn et al. 2002) where they must externalize and explain their thinking while they solve a problem. In many of these interventions, students receive modeling of a problem-solving strategy that externalizes the implicit problem solving strategies used by experts, and students are required to use it while getting appropriate feedback as needed. Similar heuristic steps can be found in different strategies: first describe the problem, then plan and construct a solution, and finally check and evaluate the solution. This approach is often used with motivating realistic problems that need an expert-like approach mimicking the culture of expert practice (Yerushalmi et al. 2007). These strategies have been shown to improve students' problem solving skills (planning and evaluating rather than searching for the appropriate equation and never evaluating) as well as their understanding of physics concepts (Cummings et. al. 1999, Foster 2000, Heller and Hollabaugh 1992, Heller et. al. 1992, Mestre et. al. 1993, Reif and Scott 1999, van Heuvelen 1991).

One challenge these approaches often face is that traditional assessment is often focused on product rather than the process, thus undermining the intended outcomes. Traditional assessment may have a negative impact through over-emphasis on grades and under-emphasis on feedback to promote learning. Thus, traditional assessment approaches are lacking in formative assessment. Black and Wiliam (1998) suggest that, for formative assessment to be productive, students should be trained in self-assessment so that they can understand the main purposes of their learning and thereby grasp what they must do to succeed. Thus, tests and homework can be invaluable guides to learning, but they must be clear and relevant to learning goals. The feedback should give guidance toward improvement and students must be given opportunity and help to work on the improvement.

Thus, an alternative instructional strategy that provides an opportunity for students to diagnose their errors involves *self-correction tasks* that are carried out after completing the solution (post factum) and require students to submit a corrected solution. However, self-diagnosis is not guaranteed to occur from self-correction. Students may or may not self-diagnose the solution through identifying where they went wrong and explaining the nature of the mistakes. An instructional strategy that bypasses this difficulty is to require students to present a diagnosis as part of the activity of reviewing their own solutions. Tasks involved in such instructional strategy, which we shall call "*self-diagnosis tasks*", involve an *explicit requirement* to carry out *self-diagnosis activities* when given some feedback on the solution.

Perkins & Swartz (1992) define a scale for thinking processes on the basis of the individual's awareness to these processes. An *implicit process* is one in which the individual is not aware of the thinking process. In a *partially explicit process,* the individual identifies explicitly the activities (e.g. "I make a decision now" or "I look for evidence"). For a *strategic process,* the individual plans and carries out an organized sequence of activities using tools for thinking and for decision-making ("this problem is complicated, so I will think of a simpler one"). Finally, a *reflective process* is one in which individuals use critical thinking tools to improve their thinking. While self-correction tasks or cooperative work promote mainly the *implicit* self-diagnosis activities, self-diagnosis tasks promote the higher-level diagnostic processes.

A variety of self-diagnosis tasks are reported in the literature on self-assessment (Black and Wiliam 1998, Black and Harrison 2001) and in particular in physics education (Yerushalmi et al. 2007, Singh et al. 2007, Yerushalmi and Eylon 2000). There are several forms of self-diagnosis tasks. *Instructions* on how to carry out the diagnosis (e.g., spoken guidelines, self-diagnosis rubrics, structure of sample solution, etc.) provide a level of detail describing the possible deficiencies in students' approaches towards the solution and in its implementation *Resources* (e.g., information provided about the correct problem solution, notebooks and textbook) are often available to the students as they diagnose their solutions. Customized *feedback* may be provided for diagnostic information about the solution.

The combination of instructions, resources and feedback can be used to calibrate different tasks for different levels of students. For example, consider the case where the resource for diagnosis is a worked out example. Research in this context focuses on making analogies (Gick and Holyoak 1983, Neuman and Schwarz 1998, Eylon and Helfman 1984) and shows that many students don't know how to use a worked out example to solve a transfer problem. The problems are similar in required general procedure, but different in detailed procedures. Students' representations that are organized around surface features (Chi et al. 1981, Eylon & Reif 1984) may prevent them from retrieving and implementing procedures from the worked out example. Eylon & Helfman (1984) found that medium- and high-achieving students benefited most from instruction which explicitly presented them with both the procedures and worked out examples, as opposed to either one of those supports by itself. This finding suggests how to write worked-out examples, which can be useful for a student in self-diagnosis.

3.2.3 Experimental Setup

The purpose of the experiment was to obtain data to compare the learning processes and outcomes in different self-diagnosis tasks. In each task, students carried out self-correction or self-diagnosis on their solutions after they solve an atypical physics problem. The actual task structure was inspired by teachers' work in a professional development workshop in which they customized instructional innovations to promote self-monitoring in physics problem solving (Yerushalmi and Eylon 2000). Tasks differed in the instructions and resources students received.

3.2.3.1 Sample

The study involved an introductory algebra-based course for pre-meds (N~180), one instructor and two teaching assistants. Recitation classrooms were distributed into a control group (Group A) and three self-diagnosis intervention groups (B, C & D). Each of these groups carried out a different self-diagnosis task. The students in different recitation classes were assigned to the different intervention groups, listed in table 3.1.

Table 3.1. Distribution of groups into self-diagnosis (SD) tasks.

Group Resources		Control (A)	Intervention (B) Instructor outlines correct solution; students fill in a self-diagnosis rubric	Intervention (C) Instructor provides a worked out example	Intervention (D) Minimal guidance: Students can use their notes + text books
Quiz 6	TA 1	~100 students	31 students		
	TA 2			28 students	25 students
SD	TA 1	~100 students	31 students		
	TA 2			25 students	24 students

3.2.3.2 Interventions and Control

In all groups, students first attempted a challenging, complex problem on a quiz. This was the sixth quiz of the semester and will hereafter be referred to as “quiz 6” for convenience. The quiz problem used for the first self diagnosis task is shown in figure 3.1. This problem is a non-typical problem in the course, as it involves non-equilibrium application of Newton's second law in the context of a non-uniform circular motion, an uncommon example of a quiz problem in an algebra-based introductory physics course.

The problem is a "context-rich problem" (Heller and Hollabaugh 1992). It has a context and motivation connected to reality, no explicit cues (e.g. “what is the apparent weight?”), requires more than one step to solve, and contains more information than is needed (e.g., the car's mass). Context-rich problems require students to analyze the problem statement, determine which principles of physics are useful and what approximations are needed (e.g., smooth track), plan and reflect upon the sub-problems constructed to solve the problem.

Quiz 6 problem

A friend told a girl that he had heard that if you sit on a scale while riding a roller coaster, the dial on the scale changes all the time. The girl decides to check the story and takes a bathroom scale to the amusement park. There she receives an illustration (see {right}), depicting the riding track of a roller coaster car along with information on the track (the illustration scale is not accurate). The operator of the ride informs her that the rail track is smooth, the mass of the car is 120 kg, and that the car is set in motion from a rest position at the height of 15 m. He adds that point B is at 5m height and that close to point B the track is part of a circle with a radius of 30 m. Before leaving the house, the girl stepped on the scale which indicated 55kg. In the rollercoaster car the girl sits on the scale. Do you think that the story she had heard about the reading of the scale changing on the roller coaster is true? According to your calculation, what will the scale show at point B?

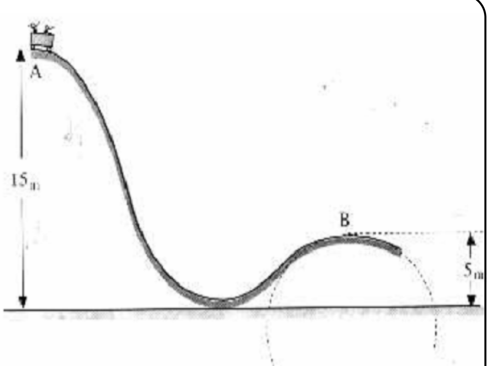


Figure 3.1. First problem used in a self diagnosis task.

The research literature suggests that students can be supported in solving such problems by guiding them to use an explicit problem-solving strategy. Previously in the semester, instructors provided students guidelines for how they should present their problem solutions, in a manner that echoes the steps of a problem-solving strategy. These guidelines are presented in figure 3.2.

Problem description: Represent the problem in Physics terms: Draw a sketch, list known and unknown quantities, target variable
Solution construction: Present the solution as a set of sub-problems, in each sub-problem write down:
- The target unknown quantity you are looking for
- The intermediate variables needed to find the target unknown quantity
- The physics principles you'll use to find it
- The process to extract the unknown

Check answer: Write down how you checked whether your final answer is reasonable

Figure 3.2. Guidelines for presenting a problem solution according to a problem solving strategy.

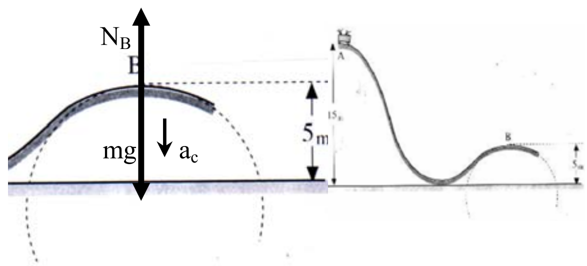
For the control group recitation class (also referred to as "group A"), the instructor discussed the solution for the problem with the students in the recitation following the quiz, but the students were not required to engage in a self–diagnosis task. In the three intervention groups (respectively referred to as "group B," "group C" and "group D"), in the recitation following the quiz, the instructor gave his students a photocopy of their solutions, and asked them to diagnose mistakes in their last week's quiz solution. Students were credited 50% of the TA's quiz grade for attempting the diagnosis. The instructor also motivated them by saying that self diagnosis will help them learn.

General evaluation	Performance level	Explain what is missing?	Circle and number mistakes you find in the solution. Fill in the following rubric.				
Problem description	Full Partial Missing	• In sketch • Known /unknowns	Mistake #	***Mark x if mistake is in:***			***Explain mistake***
Solution construction	Full Partial Missing	• Sub-problems' unknown values • Principles used		Physics	Math	Other	
			1				
Check answer	Full Partial Missing	• Possible checks for reasonability of answer	…				
			…				
			…				

Figure 3.3. The self-diagnosis rubric used by group B.

1. Description of the problem

Knowns:
The height of release: h_A=15m
Speed of the car at point A: v_A=0
The height at point B: h_B=5m
The radius at point B: R_B=30m
The mass of the car: M=120 kg
The mass of the girl: m= 55kg
Target quantity: N_B = Normal force at point B.
Assumptions: the friction with the track is negligible



2. Constructing the solution

Plan: During the motion of the girl along the curved track, the magnitude of her velocity as well as the radial acceleration changes from point to point. If the radial acceleration changes, we infer that the net force acting in the radial direction on the girl changes as well. The net force on the girl is the sum of 2 forces acting on her: the force of gravity and the normal force. To calculate the normal force at point B we can use Newton's 2nd Law $\Sigma\vec{F} = m \cdot \vec{a}$; however we will need to know the acceleration at this point. To calculate the centripetal acceleration at B, $a_R = v_B^2/R$ we need to know the speed at point B.

We will calculate the velocity of the girl at point B using the law of conservation of mechanical energy between point of departure A and point B (the mechanical energy is conserved since the only force that does work is the force of gravity which is a conservative force. The normal force does no work because it is perpendicular to the velocity at every point on the curve).

Sub-problem 1: calculating the velocity at point B

We'll set ground as the reference for potential energy and compare the total mechanical energies of the girl and car at points A and B:

$E_A = E_B$

Since the speed is zero at point A the kinetic energy is zero at that point. Therefore we get: $PE_A = KE_B + PE_B$

$\Rightarrow (m+M)gh_A = (m+M)gh + \frac{1}{2}(m+M)v_B^2$

$\Rightarrow gh_A = gh_B + \frac{v_B^2}{2}$

We can calculate the speed at B in the following way:

$v_B^2 = 2gh_A - 2gh_B = 2(10m/s^2)(15m - 5m) = 200m^2/s^2$

Sub problem 2: calculating the normal force at B

Using Newton 2nd law

$\Sigma\vec{F} = m\vec{a}_R$

$N_B - mg = -ma_R$

$N_B - mg = -mv_B^2/R_B$

$N_B = mg - mv_B^2/R_B$

$\Rightarrow N_B = (55kg)(10\,m/s^2) - (55kg)(200(m/s)^2/(30m)$

$\Rightarrow N_B = 183.33N$

Final result: when the car crosses Point B at the track, the scale indicates 183.3 N

3. Reasonability check of the final result:

- limiting cases of the parametric solution $N_B = mg - mv_B^2/R_B$: At rest (v=0): N = mg,
On horizontal surface (R → infinity): N = mg

Figure 3.4. Sample solution aligned with guidelines. This solution was provided to group C.

For group B, the instructor discussed an outline of the correct solution with the students and they were required to fill in a self-diagnosis rubric (see figure 3.3). The rubric was designed to direct students' attention at two possible types of deficiencies: deficiencies in approaching the problem in a systematic manner ("general evaluation" part) and deficiencies in the physics applied. The students in group C were provided with a worked out example that the instructor handed out during the self-diagnosis activity (see figure 3.4). This solution followed the guidelines for presenting a problem solution (see figure 3.2). Students in group D received minimal guidance; they were asked to circle and describe their errors on their photocopied solutions, aided by their notes and books only, without being provided the solution.

Prior to the intervention, the students got acquainted with the type of problems used: context rich problems, as more than half of the weekly quizzes required them to solve context rich problems. The intervention took place in the 6th and 7th quiz. Initial training took place in quiz 5. Each intervention group followed a modified version of the intervention sequence: Students got a sample incorrect solution of a “training problem” and diagnosed it according to their treatment group. Then, the instructor modeled how s/he would diagnose the sample incorrect solution for the “training problem”.

Table 3.2. Experimental sequence.

		Control group (A)	Intervention groups (B,C,D)
Initial Training:	*quiz 5*	The training followed a modified version of the intervention sequence: - Students got a sample incorrect solution of “training problem” and diagnosed it according to their intervention group - The TA modeled how she/he would diagnose the sample incorrect solution for the “training problem”	
Pretest	*quiz 6*	Students take quiz	
Treatment	*Reflection on quiz 6*	TA discuss solution to quiz 6	Groups participate in their respective self-diagnosis exercise
Posttest	*2nd Midterm*	Students attempt one far transfer problem paired to the quiz problem	

The sequence described in table 3.2 took place in all intervention groups. Aside from the initial training, this sequence was repeated for quiz 7, which had a paired problem on the third midterm as opposed to the second midterm.

3.2.4 Analysis Tools

Analysis of data requires a robust method of grading that would allow assessing students' solutions as well as students' self-diagnosis. We developed a scoring rubric in a way that only minor changes would be needed to adapt the rubric to different kinds of problems.

3.2.4.1 Structure of the rubric

In constructing the rubric, we integrated top-down and bottom-up approaches (Chi et al. 1997). In the top-down approach we constructed an a priori representation of the "ideal knowledge" underlying an expert approach to the problem, and looked for the extent to which each student's approach included certain elements of the "ideal knowledge". We developed a scoring rubric that allows identifying the gap between the student's knowledge reflected in his/her solution and the "ideal knowledge" needed to solve the problem appropriately.

In representing the expert "ideal knowledge" in the rubric, there are generic versus specific elements. Generic elements include the presence of physical principles within the problem as well as a presentation of the solution, i.e. how well it is communicated and justified. As students in the course were asked to follow a strategic problem-solving approach, we took the presentation of the solution according to such approach as representing a solution "ideally" communicated and justified. Thus, we constructed *physics* and *presentation* knowledge categories in the rubric. The physics category is divided into the subcategories of *invoking* a physical principle and *applying* that principle. Each row in each subcategory therefore represents every physical principle that a student will have to invoke and apply to correctly solve the specific problem. The presentation category includes three subcategories. Problem *description* considers whether the student presented a helpful description of the problem's situation in terms of physics concepts and principles, e.g., if a diagram is drawn to help visualize the problem. *Planning and solution construction* evaluates whether the student constructed a good plan for solving the problem with regard to the target quantity and intermediate problem steps needed to obtain this quantity. *Evaluation* considers whether the student checked the reasonability of his or her answer once it is obtained so as to make sure he or she did the problem correctly (e.g., units, extreme cases, etc.). The physics and presentation subcategories are essentially general and do not have to be changed from problem to problem; only the specific criteria based upon the subcategories need to be changed. Specific elements have to do with the physical principles and presentation steps that the student must invoke and apply in order to solve the specific problem.

However, to uncover how the students are actually thinking and the possibly incorrect mental models that they use to solve the problem, one has to take in parallel a bottom-up approach. To that end, we went over students' work and identified common mistakes in approaches to solve the problem. We represented these common approaches in the rubric under novice "incorrect ideas". The rubric has additional rows in the physics subsection that tracks if a student invoked an inappropriate principle that doesn't apply to the problem or applied inappropriately the principles she/he correctly invoked. Such analysis allows us to note a student who realized he/she made some specific mistake, even though he/she can't correct it.

The work of each student was evaluated in three ways. The researcher diagnosis of the student's quiz solution (RDS) is simply an evaluation of the students' initial quiz performance according to the rubric (i.e. not based on the TA's score). The student's self-diagnosis of his/her solution (SDS) is where we interpreted the self-diagnosis the student provided in terms of the rubric. The researcher's judgment of this student's self-diagnosis (RSD) compares the researcher's diagnosis and the student's diagnosis of the student solution. To represent these three grading evaluations, we constructed three columns in the rubric.

The method for scoring is as follows: In the RDS and RSD columns, "+" is given if a student correctly performs/identifies a mistake defined by some subcategory. A "-" is given if the student incorrectly performs, fails to identify a mistake, or identifies a mistake incorrectly. If a student is judged to have gotten something partially correct, then the grader may assign "++/-," "+/-," or "+/--." The term "n/a" is assigned if the student could not reasonably address a subcategory given the prior work done. In the SDS column, we report how the student would grade oneself with this rubric. For example, if a student correctly diagnoses a mistake, a "-" is given since this is the grade the student gives himself or herself in the solution. If a student does not refer to a mistake he has made, an "x" is assigned.

The validity of the rubric is determined by the extent to which it indeed allows us to map the student's solution to the expert "ideal knowledge" as well as to novice-like incorrect ideas. The validity was determined by four experts in physics education who perceive it as measuring an appropriate performance of the solution and self-diagnosis. Two researchers performed analysis on the students' work and any disagreements were discussed and resolved. The inter-rater reliability achieved was 80% of the aggregate of all subcategory items for a sampling of 20% of the students graded by both researchers.

3.2.4.2 Scoring rubric for quiz 6

In the following section, we will explain how the structure of the rubric explained above is adapted to quiz 6, and demonstrate how we used this rubric to assess a sample student solution and self-diagnosis.

The problem in figure 3.1 features a girl of mass m riding a rollercoaster that consists of a steep hill of height h_A followed by a circular bump with height h_B and radius R_B, and asks, given that the girl is sitting on a scale on the rollercoaster cart, how much will the scale read at the top of the circular bump?

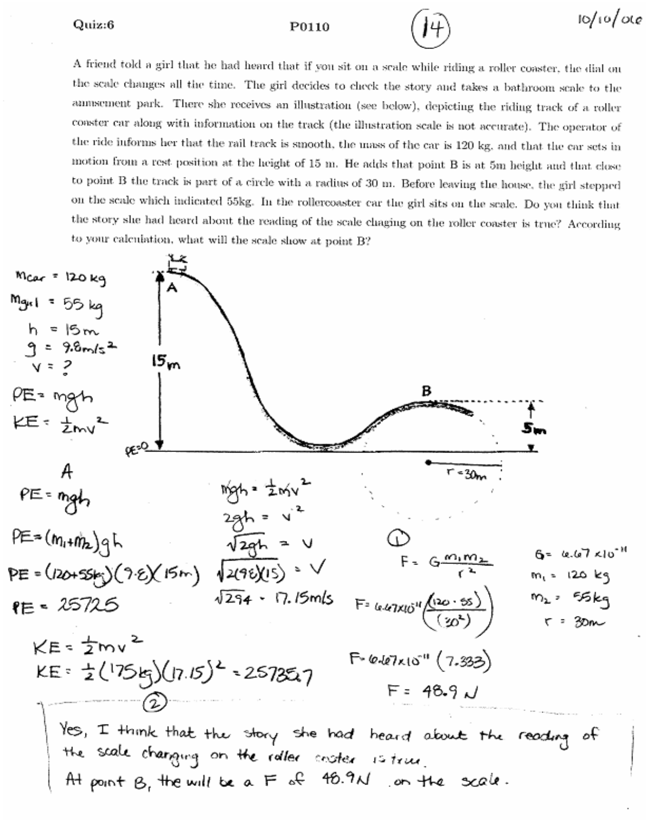
The rubric addresses both the "ideal knowledge" and the "novice knowledge per se" (incorrect ideas). In the physics part, the ideal knowledge requires invoking and applying the mechanical energy conservation law and non-equilibrium application of Newton's second law in circular motion. The student will need to understand that the target variable is the normal force the scale exerts on the girl. To calculate the normal force at point B, the student will have to invoke Newton's 2nd Law: $\Sigma\vec{F} = m \cdot \vec{a}$. He/she will also need to find intermediate variables: the net force (sum of the force of gravity and the normal force) and the acceleration at point B. To calculate the acceleration, he will have to invoke the expression for centripetal acceleration: $a_c = v_B^2/R$. The intermediate variable, the speed of the cart at the top of the circular bump, can be found using the law of conservation of mechanical energy, $PE_i + KE_i = PE_f + KE_f$, between the point of departure and point B at the top of the bump (which is justified because all forces doing work are conservative forces). Therefore, there will be two rows in the invoking subcategory to evaluate if the student referred to the required principles (Conservation of mechanical energy and Newton's second law), and two corresponding rows in the applying subcategory to evaluate if the student applied them correctly.

Also, the system needs to be properly defined, as the mass that should be taken is the mass of the girl, not that of the girl and the car (the student must realize this on his/her own even though the problem gives the values of both masses). We found several common incorrect ideas in the physics part across the student sample: defining the system inappropriately and/or inconsistently, calculating kinetic and/or potential energy without applying energy conservation, incorrectly applying energy conservation, calculating Newton's second law without applying it to circular motion, or using the centripetal force as the normal force.

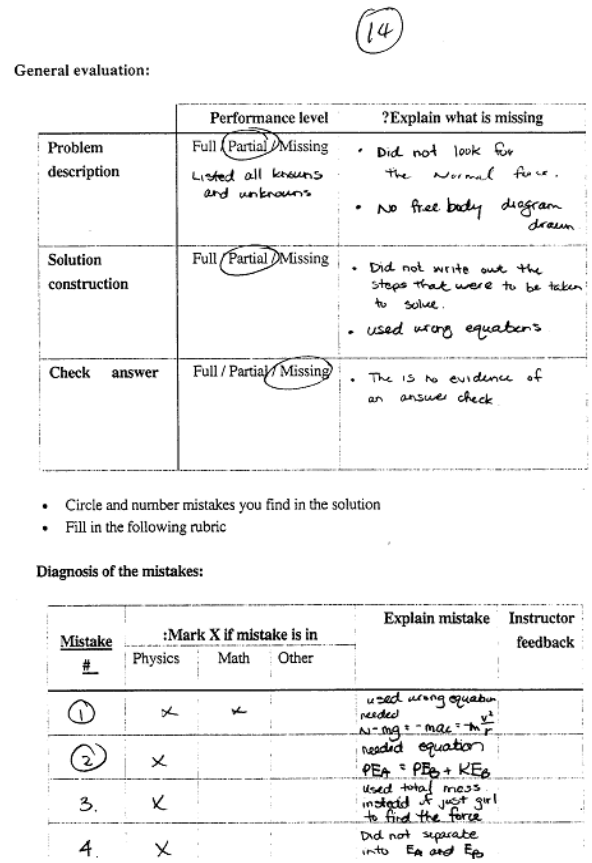
We added to the rubric an algebra category to reflect mathematical mistakes made during the problem-solving process, for example forgetting a coefficient when rewriting an equation. This category is meant to differentiate from physics errors and was not analyzed as the other categories were, instead serving as a means of tracking algebraic errors for the other grader during reliability checks.

In the presentation part, the description should include a visual representation of the forces acting on the girl at point B (where her "new" weight needs to be calculated), the velocities and heights or E_k and E_p at points A and B, the vector corresponding to the acceleration at point B, the axes, the radius of the circle and writing down clear and appropriate known quantities. An appropriate target quantity (the normal force) and intermediate variables (speed and acceleration at point B) must be chosen in the planning part. In the checking part the student needs to check reasonability of the answer, and the units.

Figures 3.5 and 3.6, respectively, represent a student solution for the problem presented earlier (see figure 3.1) and a self-diagnosis attempt of a student who was assigned to intervention group B (Outline + Self-Diagnosis Rubric). The other groups did not use the rubric


Quiz:6 P0110 (14) 10/10/06

A friend told a girl that he had heard that if you sit on a scale while riding a roller coaster, the dial on the scale changes all the time. The girl decides to check the story and takes a bathroom scale to the amusement park. There she receives an illustration (see below), depicting the riding track of a roller coaster car along with information on the track (the illustration scale is not accurate). The operator of the ride informs her that the rail track is smooth, the mass of the car is 120 kg, and that the car sets in motion from a rest position at the height of 15 m. He adds that point B is at 5m height and that close to point B the track is part of a circle with a radius of 30 m. Before leaving the house, the girl stepped on the scale which indicated 55kg. In the rollercoaster car the girl sits on the scale. Do you think that the story she had heard about the reading of the scale chaging on the roller coaster is true? According to your calculation, what will the scale show at point B?


(14)

General evaluation:

	Performance level	?Explain what is missing
Problem description	Full / Partial / Missing Listed all knowns and unknowns	• Did not look for the Normal force. • No free body diagram drawn
Solution construction	Full / Partial / Missing	• Did not write out the steps that were to be taken to solve. • used wrong equations
Check answer	Full / Partial / Missing	• The is no evidence of an answer check

- Circle and number mistakes you find in the solution
- Fill in the following rubric

Diagnosis of the mistakes:

Mistake #	:Mark X if mistake is in Physics	Math	Other	Explain mistake	Instructor feedback
1	X	X		used wrong equation needed $N-mg=-ma_c=-m\frac{v^2}{r}$	
2	X			needed equation $PE_A = PE_B + KE_B$	
3.	X			used total mass instead of just girl to find the force	
4.	X			Did not separate into E_A and E_B	

Figures 3.5 and 3.6. Sample student's solution and self diagnosis (quiz 6). The circled numbers 1 and 2 are references labeled in figure 3.6 - student's self-diagnosis. The circled number 14 is the code number for the student. The student belongs to group B that was required to fill in a diagnosis rubric.

presented in figure 3.6. In table 3.3, we present the rubric adapted to quiz 6 along with the scoring of this student.

Table 3.3 shows that this specific student correctly invoked conservation of energy on the original quiz and therefore did not address it during self-diagnosis. In the researcher's judgment column (RSD), the scorer then states "n/a" and does not consider this invoked law in assigning a score for the researcher's judgment. The student did not invoke Newton's second law, and therefore in the RDS column he received a "-". However, the student diagnosed this problem, and therefore received a "-" in the SDS column. This results in a "+" grade in the RSD column for diagnosing this problem.

Table 3.3. Rubric developed for self-diagnosis study for quiz 6.
The student featured in Figures 3.5 and 3.6 is graded to serve as an example.

General Task	Specific Criteria		RDS	SDS	RSD
PHYSICS PRINCIPLES (Ph.)					
Invoking physical principles	**Ideal knowledge**				
	1. Conservation of mechanical energy (student's quiz is correct)		+	x	n/a
	2. Non-equilibrium applications of Newton's second law in centripetal motion		-	-	+
	Incorrect ideas				
	3. inappropriate principle: "-" marked if inappropriate principle is used in student's solution or diagnosis (student used gravitational law here)		-	-	+
	Ideal knowledge	**Incorrect ideas**			
	defining the system appropriately and consistently	defining the system inappropriately and inconsistently	-	-	+
Applying physical principles	1. conservation of mechanical energy	e.g. calculation of KE/PE without energy conservation	-	-	+
		(etc.)			
	2. Non-equilibrium applications of Newton's second law	referring to centripetal force as a physical force			
		forgetting normal force	(-)	-	+
		(etc.)			
ALGEBRA (Alg.)					
Algebra	Algebraic manipulation		+	x	n/a
PRESENTATION (Pre.)					
Description (Des.):	1. Invokes a visual representation	Free body diagram, acceleration vector, axis , defining PE = 0, radius of the circle	-, -, -, +, +	-. x, x, x, x	+, -, -, n/a, n/a
	Clear/appropriate knowns (the student listed most but missed one)		++/-	+	++/-
Plan/Solution Construction (Plan): representing the problem as a set of sub-problems	1. Appropriate target quantity chosen (first chose F_g, fixed in diagnosis)		-	-	+
	2. did not write down surplus equations or intermediate variables		-	x	-
	3. appropriate intermediate variables explicitly stated (v_b but not a_c)		+/-	+/-	+
	4. explicitly stating in words or a generic form the principles used to solve for this intermediate variables (not done by student)		-	x	-
Evaluation (Che.)	1. writing down the units (student sometimes did so)		+/-	x	+/-
	2. checking the answer		-	-	+

Table 3.4. Rubric scoring for the sample student.

Category	Grading
RDS	**Ph: 0.40; Pre: 0.36**
SDS	**Ph: 0.80; Pre: 0.72**
RSD	**Ph: 0.50; Pre: 0.64**

After the categories were coded, an overall scoring took place for each of RDS, SDS and RSD. We weighed each "+" as worth 1 point and a "-" as worth 0. A mixed value (e.g. "++/-") is worth an average of the "+" and "-" symbols (e.g. 0.67). Thus, we weighed each subcategory as worth 1, 2/3, 1/2, 1/3 or 0 point given marks of "+," "++/-," "+/-," "+/--," or "-," respectively. The grade also took into account when a student did not invoke a principle and thus would necessarily not apply it by different weights we assigned in order to avoid penalizing the same error twice, e.g. weighting the score on the "invoking" category twice if the student did not invoke a principle on the quiz attempt (since he/she cannot apply what he did not invoke). The weights also ensured that a student who did not invoke a principle on the quiz was penalized more than a student who did invoke the correct principle but did not apply it correctly.

Table 3.4 displays how the scores given in table 3.3 are interpreted as an overall score. We differentiated the researcher's judgment of the students' self-diagnosis into physics and presentation grades. The overall score for a physics score can be interpreted as an average of all possible criteria that the student correctly addressed. The number scale is 0 (0%) to 1 (100%).

3.2.5 Results

To compare the groups we performed analysis of covariance (ANCOVA) taking the quiz physics grade as a covariate (after checking the homogeneity of slopes).

3.2.5.1 Self-diagnosis - Physics

The physics self-diagnosis rubric scoring reflected both the *expert ideal knowledge* and the *novice knowledge per se* (what ideas the student believes are needed to solve the problem are reflected in his/her solution and diagnosis) by describing in detail which mistakes students could diagnose. This approach allowed us to better differentiate between students in subcategory analysis (see tables 3.7 and 3.10).

Tables 3.5 and 3.6 present a comparison of students' performance of self-diagnosis of the physics in the three self-diagnosis groups (i.e. the RSD column of the rubric). Note that all students made mistakes on the quiz – nobody could answer the problem completely correctly – and so all of them are included in the analysis.

Table 3.5. Grades for students' self-diagnosis in physics.

	Group B	Group C	Group D
Mean	0.73	0.57	0.24
Std. Err.	0.049	0.051	0.055

Table 3.6. P Values for physics self-diagnosis grades.

	Group B	Group C	Group D
Group B		0.02	<0.0001
Group C			<0.0001

The ANCOVA analysis shows that all groups differ from each other (p value < 0.05). Group B clearly performed better than group C, who in turn performed better than group D. The findings make sense with respect to the level of scaffolding support each group received. The TA's outline provided a problem solution for group B while still requiring thought to examine details. The students' rubric (see figure 3.6) provided structure with which to understand what items to evaluate. Group D's activity, which provided the least support, simulates the most common diagnostic context: students referring to the back of the book answer. The finding might suggest that we can't expect much from students in the most common diagnostic context for a sufficiently difficult problem.

Table 3.7 presents students' performance of self-diagnosis in the physics subcategories (i.e. the percentage of students who diagnosed their mistakes out of those who made mistakes in each sub category). In the invoking subcategory, 35% of the students in group B did not invoke one or both correct physics principles, and 55% of that 35% correctly diagnosed this mistake completely. In group D, only 40% diagnosed their mistakes completely, and in group C only

Table 3.7. Self-diagnosis evaluation of the physics subcategories.
The scoring is as follows. "+" = a correct diagnosis, "+/-" = a partially correct diagnosis, and "-" = no diagnosis or an incorrect diagnosis. "Total" refers to percentages of students who had mistakes in their quiz regarding some subcategory, and the percentages listed in the scoring are of that total.

	Group B			Group C			Group D		
	+	+/-	-	+	+/-	-	+	+/-	-
Invoking	55%	40%	5%	35%	53%	12%	40%	45%	15%
	Total: 35%			Total: 55%			Total: 24%		
Applying	15%	80%	5%	15%	58%	28%	10%	62%	29%
	Total: 66%			Total: 53%			Total: 62%		

35% did so. In the applying subcategory, the diagnosis was harder than the diagnosis of missing invoked principles, as no more than 15% of those who had application mistakes diagnosed them completely in any group. We conclude that it is much easier for students to identify problems in invoking principles for a difficult, atypical physics problem than to identify problems in how these principles are applied. We note that the difference in between the groups regarding subcategories is aligned with the difference in the overall self-diagnosis performance and can be explained along the same lines. Group B received maximal guidance, while group D received minimal guidance.

3.2.5.2 Self-diagnosis - Presentation

The presentation self-diagnosis grade is divided into the subcategories of description, planning, and checking. Thus, the presentation self-diagnosis grade reflects whether or not the students realized correctly if in their solution they fully described the problem (e.g. sketched a free body diagram, represented the knowns correctly, etc.), clearly described an appropriate solution plan (e.g. defined the target and intermediate variables), and checked the solution (e.g. checking units, limiting cases, etc.). Tables 3.8 and 3.9 present a comparison of students' performance of self-diagnosis of the presentation in the self-diagnosis groups.

Table 3.8. Grades for students' self-diagnosis in presentation.

	Group B	Group C	Group D
Mean	0.42	0.10	0.12
Std. Err.	0.022	0.023	0.025

Table 3.9. P Values, ANCOVA analysis for presentation self-diagnosis.

	Group B	Group C	Group D
Group B		<0.0001	<0.0001
Group C			0.57

The ANCOVA analysis (again, taking the quiz physics grade as covariate) shows that groups D and C have significantly lower presentation-self-diagnosis grades than group B (p value < 0.05). The findings make sense as groups C and D received the least support for the presentation part. The rubric that group B received (see figure 3.7) explicitly required students to explain what is missing in the problem description (a sketch, known and unknown quantities, target variable), solution construction (set of sub-problems defined by intermediate variables looked for and the physics principles used to find them) and checking of the answer. The findings may suggest that, in order to let the students better diagnose their mistakes, a clear instruction requiring the students to provide specific details is needed. Even a thoroughly detailed solution that emphasized these problem-solving steps was not helpful to group C.

Table 3.10 presents a between-group comparison of the percentage of students who diagnosed their mistakes regarding the different sub-categories of presentation-self-diagnosis.

There is a difference in the diagnosis of description mistakes from group B to groups C and D. While 20% of the students in group B who made description mistakes fully diagnosed

Table 3.10. Self-diagnosis – presentation subcategories.
See table 3.7 for scoring details.

	Group B			Group C			Group D		
	+	+/-	-	+	+/-	-	+	+/-	-
Description	20%	41%	39%	7%	32%	62%	0%	43%	57%
	Total: 67%			Total: 69%			Total: 67%		
Plan	34%	46%	20%	16%	53%	31%	6%	55%	39%
	Total: 85%			Total: 75%			Total: 86%		
Check	39%	59%	3%	0%	82%	18%	0%	88%	12%
	Total: 100%			Total: 100%			Total: 96%		

these mistakes, 7% in group C and no one in group D did so. One may also note that a lower percentage of students in group B failed to diagnose any description errors than groups C and D did.

Although the overall scores are quite low for all groups in the planning section, there is a noticeable difference in the diagnosis of mistakes between group B and groups C and D. While 34% of those who had problems in planning in group B fully diagnosed these mistakes, 16% in group C and 6% in group D did so. Fewer students in group B failed to diagnose these issues (20%) than in groups C (31%) and D (39%).

An interesting aspect of planning the solution is identification of the target variable. In this problem the students were supposed to find the normal force acting on a girl moving in a circular motion at a certain point. This normal force indicates the weight of the girl at this point. Many of the students failed to recognize this, and a very common mistake was to find a new mass for the girl at this point. 64% of the students who did not correctly identify the target variable in group B fully recognized this problem in self-diagnosis. Only 33% in group C and 11% in group D did so.

In the checking subcategory, there is a noticeable difference in the diagnosis of mistakes in checking the solution between group B and groups C and D. 39% of those who had mistakes in checking the solution in group B fully diagnosed these mistakes, while no one in groups C and D did so.

We note that the difference between the groups regarding sub categories is aligned with the difference in the overall self-diagnosis performance and can be explained along the same lines: group B received maximal guidance, while groups C and D received less guidance in the presentation part.

3.2.5.3 Self-diagnosis and Prior Knowledge

Table 3.11 presents correlations between the quiz grades and the self-diagnosis grades for both physics and presentation. All groups were averaged together for this comparison as their results were roughly equivalent. For physics, one notes a marginal negative correlation between the grade of the physics part of the quiz, and the grade of the diagnosis of this part. In other words, students who make more physics mistakes for this problem are better at diagnosing them. A possible explanation to a negative correlation may be that students who did well on the original

Table 3.11. Correlations between quiz and self-diagnosis (SD) performance for all groups for quiz 6.

Physics: Quiz vs. SD		Presentation: Quiz vs. SD	
Correlation	P value	Correlation	P value
-0.172	0.12	0.277	0.01

quiz had less of an opportunity to trace mistakes because they had fewer mistakes to begin with. However, this means that there is some set value for the number of mistakes in a given problem solution below which the students are unlikely to be able to diagnose their errors. It seems that if such a bar indeed exists it would involve diagnosing the application of the principles that few students were able to perform. These results would suggest that students who did poorly on the quiz would reduce the gap with respect to students who did well in the quiz. Such a hypothesis is further considered in section 3.4.

For presentation, one notes a positive correlation between the grade of the presentation part of the quiz, and the grade of the diagnosis of this part. Students whose presentation was deficient were worse also at diagnosing it. This means that the initial gap between students who did well and students who did not remained in the presentation part.

3.2.6 The Midterm Problem

The problem on the second midterm, presented in figure 3.7, functioned as a posttest problem. The aim was to find out students' abilities to transfer the understanding they gained when diagnosing the problem in quiz 6 to solving a problem in a somewhat similar context (i.e. motion along a vertical circular path in a non-equilibrium situation). The midterm took place within a week of the recitation in which students self-diagnosed their solutions to quiz 6. Table 3.12 compares the paired midterm and quiz problems for quiz 6.

Both problems employ the same physical principles, i.e. Newton's second law applied in a non-equilibrium situation involving centripetal acceleration and conservation of mechanical energy. Furthermore, both questions also require recognition of similar target variables (in the form of either a normal force or tension force) and intermediate variables involving centripetal acceleration and velocity at the maximum/minimum point on a circular trajectory. However, the problems differ in context. In addition, for the midterm problem, students had to first realize that they should focus on the lowest point on the circular trajectory. However, we found in the categorization study in chapter 2 that students at the introductory level did not associate these

A family decides to create a tire swing in their back yard for their son Ryan. They tie a nylon rope to a branch that is located 16 m above the earth, and adjust it so that the tire swings 1 meter above the ground. To make the ride more exciting, they construct a launch point that is 13 m above the ground, so that they don't have to push Ryan all the time. You are their neighbor, and you are concerned that the ride might not be safe, so you calculate the maximum tension in the rope to see if it will hold. (a) Where is the tension greatest? (b) Calculate the maximum tension in the rope, assuming that Ryan (mass 30 kg) starts from rest from his launch pad. Is it greater than the rated value of 750 N? (c) Name two factors that may have been ignored in the above analysis, and describe whether they make the ride less safe or more safe.

Figure 3.7. Midterm problem corresponding to quiz 6.

Table 3.12. Comparison of quiz 6 and midterm problems.

	Principles	Variables	FBD	Context	details
Quiz 6	EC 2^{nd} law	$\vec{v}$ $\vec{a}_c$ $\vec{N}/\vec{T}$	$\uparrow \vec{N}/\vec{T}$ $\downarrow m\vec{g}$	Roller coaster	$\vec{a}_c$ in opposite direction to $\vec{N}$
Midterm				Tire swing	$\vec{a}_c$ in the same direction as $\vec{T}$

two problems as belonging to the same category. Thus, the midterm problem is considered a far transfer problem with respect to the quiz.

3.2.6.1 Midterm Physics Scores

Table 3.13 gives the mean physics scores for all self-diagnosis groups on the midterm problem. In consideration of the effect of the TAs on the inter-group comparison, we present analysis of each TA's groups separately. Table 3.14 shows ANCOVA p-value comparisons between the groups which each TA presided over. They show that group B, which received the maximum support, did somewhat better than the control group A. One might conclude that that the rubric and solution outline provided students with a clear picture of what they did wrong. However, despite actually seeing the complete solution, group C fared worse on the midterm than group D, who had to try and figure out the solution on their own from textbook and notes.

Table 3.13. Means and standard deviations of each group's midterm physics scores.

	First TA		Second TA	
Group	A	B	C	D
Mean	0.424	0.526	0.329	0.473
Std. Dev.	0.042	0.053	0.048	0.063

Table 3.14. P-value comparison between midterm physics scores of each TA's groups.

First TA	Group B	Second TA	Group D
Group A	0.112	Group C	0.071

There are at least two possible interpretations for this result, based upon two traits of innovation and efficiency proposed by Schwartz et al. (2005). See Figure 3.8 for a visual description.

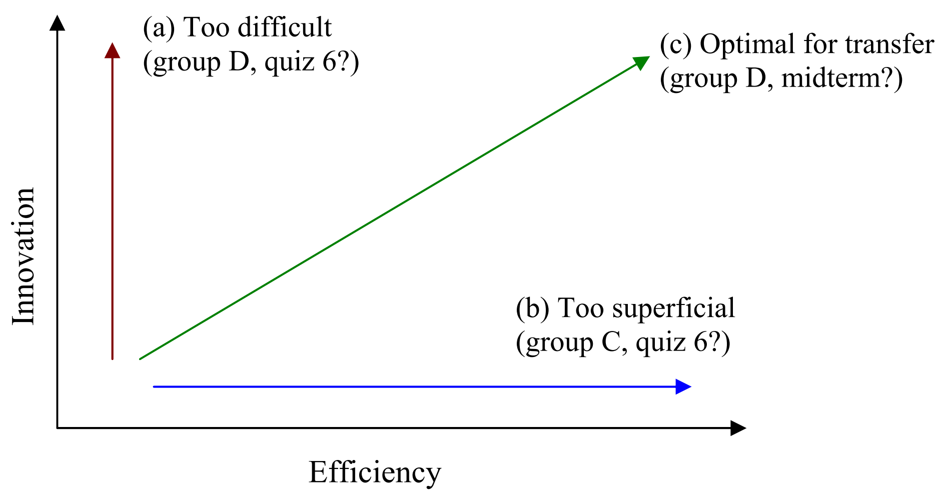


Figure 3.8. Innovation vs. efficiency graph.

First, innovation may best be described as the novelty of the problem, which translates to the cognitive engagement the problem requires to be solved. Innovation is a therefore a good trait for a problem to have. If a problem is too innovative, however, the amount of cognitive engagement transcends the zone of proximal development (Vygotsky 1978), i.e. the problem is too difficult given the learner's current state of knowledge. The struggle that group D exhibited on the self-diagnosis of quiz 6 indicates that the available knowledge content in the text and notes was insufficient for students to diagnose their errors, leading to a poor performance on physical principles (see arrow (a) in Figure 3.8).

Second, efficiency deals with the ability to "rapidly retrieve and accurately apply appropriate knowledge and skills to solve a problem or understand an explanation" (Schwartz et al. 2005). Efficient instruction, with regard to the self-diagnosis study, pertains to a full explanation of a problem as provided to group C. The potential pitfall, however, is that if instruction is too efficient, it may result in superficial learning. It is apparent from Table 3.13 that group C's self-diagnosis was not meaningful in the sense that an elaborated solution allowed self-diagnosis to occur on a more superficial level (see arrow (b) in Figure 3.8). Students could simply compare and contrast their answer with the detailed correct solution without necessarily thinking deeply about their errors.

Schwartz et al. therefore propose a combination of innovation and efficiency. Group D's midterm performance relative to that of group C can be understood if there was a self-diagnosis stage that occurred after the formal self-diagnosis exercise. That is, after the students struggled twice to solve the problem without the explanation provided to the other two groups, group D may have tried to make sense of the solution provided to all students after the exercise concluded. This additional cognitive engagement may have led to a more thorough and efficient understanding of the problem providing them with deeper insight into the correct solution. In essence, group D may have actually received a more valuable self-diagnosis treatment than group C (see arrow (c) of Figure 3.8). While both groups had access to the solution at some point, group D students also had prior confrontation with their inability to solve and diagnose the problem, but group C's experience was largely self-contained to the solution.

3.2.6.2 Physics Self-Diagnosis vs. Midterm Performance

The intra-group comparison is shown in table 3.15 in the form of a correlation between the self-diagnosis physics scores of the treatment groups and their midterm scores. Each different group is reported for physics scores in both studies to highlight differences between groups in the physics category. A borderline positive correlation for group B's self-diagnosis score is shown, indicating a tendency for the self-diagnosis to have helped students somewhat on the midterm. All other intra-group comparisons yield no correlations. This might also be explained in that between the self-diagnosis task and the midterm, a separate learning process takes place.

Table 3.15. Correlation of physics scores: self-diagnosis vs. midterm.

Group	Correlation	p value
B	0.35	0.16
C	0.13	0.55
D	0.07	0.80

3.2.6.3 Midterm Presentation Scores

Table 3.16 gives the mean presentation scores for all groups on the midterm problem, and an inter-group analysis of p-values between the treatment groups are shown in Table 3.17. There is no significant difference or correlation between any of the groups. In addition, there was no effect of the treatment on presentation performance, even though group B did better on the self-diagnosis exercise. An intra-group analysis, presented in table 3.18, shows no correlation of the midterm scores with self-diagnosis scores.

Since group B fared as poorly as the other groups with regard to presentation score on the midterm despite a better performance on presentation self-diagnosis, we must consider the "one-time" nature of the intervention. The students only performed the self- diagnosis exercise once, which would probably not be enough to effectively develop presentation skills, even though it is possible to understand the physical principles necessary to solve the problem. It would be of interest to examine whether applying the intervention consistently throughout the semester would help develop presentation skills.

The results in this section will be discussed with the results of section 3.3 in section 3.4.

Table 3.16. Means and standard deviations of each group's midterm presentation scores.

	First TA		Second TA	
Group	A	B	C	D
Mean	0.426	0.437	0.410	0.462
Std. Dev.	0.020	0.025	0.022	0.030

Table 3.17. P-value comparison between midterm presentation scores of each TA's groups.

First TA	Group B	Second TA	Group D
Group A	0.580	Group C	0.154

Table 3.18. Correlation of presentation scores: self-diagnosis vs. midterm.

Group	Correlation	p value
B	-0.208	0.409
C	-0.162	0.482
D	0.081	0.791

3.3 PART 2 OF THE STUDY: A MORE CONVENTIONAL QUIZ PROBLEM

3.3.1 Introduction

The previous section of this chapter describes the rationale for our study on a self-monitoring assessment task in which students identify and explain their own errors with different levels of scaffolding (Mason et al. 2008, Cohen et al. 2008, Yerushalmi et al. 2008). This task involved asking students to solve a difficult, context-rich problem in a quiz and then asking them to diagnose their mistakes. We considered three different levels of scaffolding during the intervention to determine what level of support helped students self-diagnose the best. We then asked a far transfer problem in the midterm exam to explore whether self-diagnosis helped students in different experimental groups compared to the control group. Our primary findings were that students with a higher level of scaffolding performed better on self-diagnosis, but that transfer to the midterm problem was limited overall for all intervention groups. Considerations of performance on the quiz problem and midterm problem suggested that the problems were too difficult for students; not a single student got full credit on the quiz problem and few were able to answer the corresponding midterm problem correctly.

Furthermore, there was evidence that transfer from the quiz problem to the midterm problem was too "far." Specifically, a high percentage of students did not consider the presence of centripetal acceleration, a vital consideration for both problems, even in the midterm problem. A common trend among the students who missed the problem was to only apply a simple equilibrium application of Newton's second law at the point of interest. In addition, there was an initial part to the midterm problem that asked students to determine the point of interest before solving the rest of the problem (i.e., at what point in a pendulum's swinging motion is the tension in the rope the greatest?). Approximately 15% of students incorrectly believed that tension force was at maximum in the rope at the highest point and therefore neither invoked the conservation of mechanical energy nor centripetal acceleration which is zero at the highest point where the pendulum is momentarily at rest. For these students, it was not possible to judge whether they would have realized the isomorphic nature of the self-diagnosed problem and the quiz problem if they were asked to find the tension at the lowest point of a swinging pendulum.

We decided to conduct another experiment using the same guidelines as described in section 3.2. In this study, another set of problems was chosen for self-diagnosis of a quiz and transfer in a different midterm with the following criteria: a) while the quiz problem that students self-diagnosed was also a context-rich problem with two parts, the two principles used were both more familiar to introductory students and b) the paired midterm problem is not as superficially different from the self-diagnosed quiz problem and thus does not require as far of a transfer. The goal of this section is to examine students' ability to self-diagnose this easier problem with different scaffolds and their ability to transfer to the near transfer problem. We conclude by reflecting on what we learned from self-diagnostic activities on both these problems. We again focus on how well students review their own solutions in order to improve them or learn from them. Similar to the approach in section 3.2, we will consider both the students' knowledge content and the students' approach to problem-solving strategy, as self-monitoring may pertain to arriving at a solution as well as more general learning goals (Larkin et al. 1980).

The process of the alternative-assessment "self-diagnosis" task is the same as in the companion study. Students are required to present a diagnosis (namely, identifying where they went wrong, and explaining the nature of the mistakes) as part of the activity of reviewing their quiz solutions. We consider the same research questions posed in section 3.2 but in the context of a more typical quiz problem and a closer transfer task than previously. Finally, we discuss the similarities and differences between students' self-diagnosis on the two problems and their subsequent performance on transfer problems.

We adapt the scoring rubric developed for the previous study and described in table 3.4 to conform to the different problems discussed in this study. This rubric remained aligned with two approaches (Chi 1997) which map the student statements to an expert "ideal" knowledge representation and to the student's novice knowledge per se.

3.3.2 Sample Problem: Quiz Problem Used for Self-Diagnosis

The problem used for quiz 7 which students self-diagnosed features a friend of mass m = 60kg, running on a horizontal plane with a known speed and jumping on a stationary skate board with a considerable mass M = 5kg. The friend wishes to reach a minimum height of 3m on a slope, while riding the skate board. The student needs to decide whether the friend's plan would work. The reader is specifically instructed not to ignore the mass of the skateboard.

In order to solve this problem, a student will need to understand that the target variable is the maximum height that the skate board can reach, or the speed of the skate board at the height of 3m. To calculate maximum height, the student will have to invoke energy conservation before and after the climb: $PE_i + KE_i = PE_f + KE_f$ (which is justified because all forces doing work are conservative forces, ignoring friction). He/she will need to find the intermediate variable, which is the speed v of the friend plus the skate board after the jump. To calculate this speed, he will have to invoke the expression for momentum conservation: $mv_i = (m+M)v_f$.

Figures 3.9, 3.10 and 3.11, respectively, represent the problem for quiz 7, the quiz attempt and self-diagnosis attempt of a sample student. This student's work also features examples of what a student might do correctly, such as correctly invoking the energy conservation. This work also features some examples of what he might have done incorrectly, such as not invoking the momentum conservation at all.

3.3.3 The Rubric Adapted

3.3.3.1 Specific Knowledge: Categories and Subcategories

Table 3.19 shows the corresponding rubric which was adapted from the rubric used in the first study (see Table 3.3). We again consider "ideal" expert novice representation as well as novice representation, as explained in the accompanying study in section 3.2. One may note that generic considerations of the rubric may be retained without change and specific considerations

> You are helping a friend prepare for the next skate board exhibition. Your friend who weighs 60 kg will take a running start and then jump with a speed of 1.5 m/s onto a heavy duty 5 kg stationary skateboard. Your friend and the skateboard will then glide together in a straight line along a short, level section of track, then up a sloped concrete incline plane. Your friend wants to reach a minimum height of 3 m above the starting level before he comes to rest and starts to come back down the slope. Knowing that you have taken physics, your friend wants you to determine if the plan can be carried out or whether he will stop before reaching a 3 m height. Do not ignore the mass of the skateboard.

Figure 3.9. Quiz 7 problem.

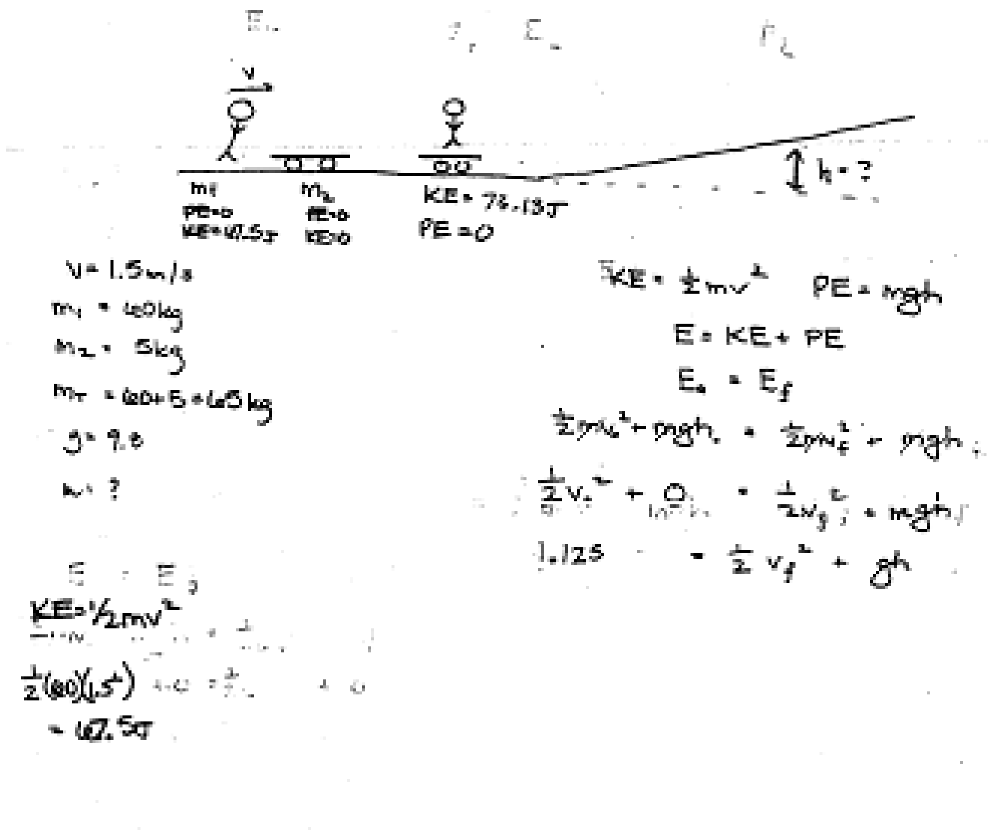
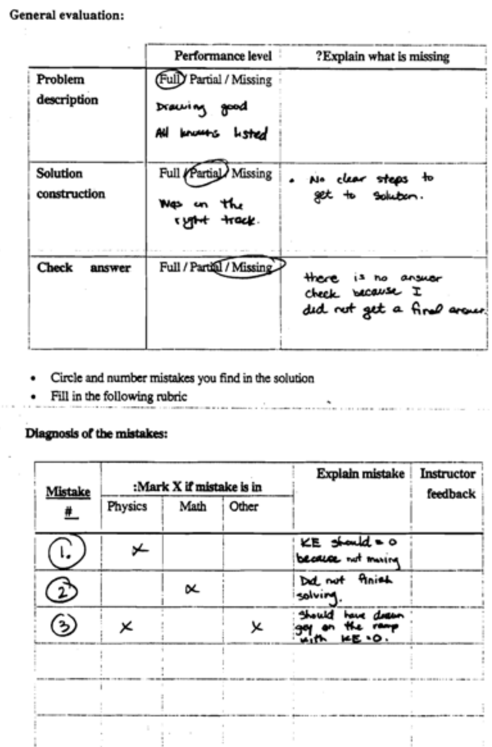


Figure 3.10. Sample student's quiz 7 solution from the self-diagnosis study.
Here, the student invokes energy conservation well but does not invoke momentum conservation.


General evaluation:

	Performance level	?Explain what is missing
Problem description	Full / Partial / Missing Drawing good All knowns listed	
Solution construction	Full / Partial / Missing Was on the right track.	• No clear steps to get to solution.
Check answer	Full / Partial / Missing	there is no answer check because I did not get a final answer.

- Circle and number mistakes you find in the solution
- Fill in the following rubric

Diagnosis of the mistakes:

Mistake #	:Mark X if mistake is in			Explain mistake	Instructor feedback
	Physics	Math	Other		
1.	x			KE should = 0 because not moving	
2		x		Did not finish solving.	
3	x		x	Should have drawn guy on the ramp with KE = 0.	

Figure 3.11. Sample student's quiz 7 self-diagnosis attempt from the self-diagnosis study. The other groups did not receive the worksheet depicted in figure 3.10 but instead wrote their diagnosis elsewhere. See Figure 3.9 for this student's initial quiz attempt.

related to the problem are adapted relatively easily.

The two physical principles involved in the problem used for quiz 7 are conservation of linear momentum and conservation of mechanical energy. The quiz problem was easier for students than the problem used in quiz 6 because the more complex Newton's 2^{nd} law in a non-equilibrium application involving centripetal acceleration is replaced by a one-dimensional application of conservation of linear momentum involving a completely inelastic collision. The plan/solution category again has three different subcategories similar to those in the rubric featured in Table 3.3. This time the visual representation is the respective initial and final states of the skateboarder and the skateboard for each step of the problem. In particular, we consider the masses and velocities of the friend and the skateboard before and after they collide. We also examine the beginning and end points of the skateboarder traveling up the hill. Here, the student may either describe the potential and kinetic energy at the bottom of the hill and at the top of the skateboard's motion or the initial and final velocities and heights of the motion after the collision. The other criteria of the presentation remain unchanged.

3.3.3.2 Scoring

Table 3.20 displays how the overall scores given in table 3.19 are interpreted. An explanation of the scoring system is discussed in section 3.2. The overall score for physics can be interpreted as an average of all possible criteria that the student correctly addressed.

3.3.3.3 Reliability

To assure objectivity in scoring using the rubric, two researchers used the process described in the first study. They again independently scored ~10 sample students using the rubric for the problem diagrammed in Figures 3.9 and 3.10. They then discussed how criteria should be applied to the students' work objectively. We found that the scorers could agree to within at least 80% in scoring the rubric. This established a reasonable inter-rater reliability that was consistent in a more thorough analysis of about 200 students.

3.3.4 Experimental Setup & Sample

Once again, it was our intention to achieve ecological validity, namely, to simulate conditions that are feasible in actual implementation in a classroom given the time constraints of teachers and students. As in the first part of the study in the previous section, we performed the same

Table 3.19. Rubric scoring for self-diagnosis study for quiz 7.
The student featured in figures 3.9 and 3.10 is graded to serve as an example for quiz 7. Please see table 3.3 for an explanation on notation.

General Task	Specific Criteria		RDS	SDS	RSD
PHYSICS PRINCIPLES (Ph.)					
Invoking physical principles	**Ideal knowledge**	**incorrect ideas**			
	1. Conservation of mechanical energy (student correct)	calculation of energy/energies without energy conservation	+	x	n/a
	2. Momentum conservation	calculation of momentum without momentum conservation	-	x	-
	Incorrect ideas				
	3. inappropriate principle: "-" marked if inappropriate principle is used in student's solution or diagnosis		n/a	x	n/a
	Ideal knowledge	**incorrect ideas**			
	defining appropriately and consistently the system	defining inappropriately and inconsistently the system	+	x	n/a
Applying physical principles	1. conservation of mechanical energy	The initial speed is incorrect, or not completing calculation	-	-	+
	2. Momentum conservation	The speed of skate board before collision is not zero	n/a	n/a	n/a
		(etc.)			
ALGEBRA (Alg.)					
Algebra	Algebraic manipulation		-	-	+
PRESENTATION (Pre.)					
Description (Des.):	1. Invokes a visual representation	masses & velocities before and after collision, velocities and heights or potential and kinetic energies before and after climbing	+, +/-	+,+	+,+/-
	2. Clear/appropriate knowns (here: all knowns are listed)		+	+	+
Plan/Solution Construction (Plan): representing the problem as a set of sub-problems	1. Appropriate target quantity chosen (here the student does not finish work, not fixed in diagnosis)		-	x	-
	2. did not write down surplus equations or intermediate variables		n/a	n/a	n/a
	3. appropriate intermediate variables explicitly stated (here student does not find v)		-	x	-
	4. explicitly stating in words or a generic form the principles used to solve for this intermediate variables (not done by student)		-	-	+
Evaluation (Che.)	1. writing down the units (student sometimes did so)		+/-	x	+/-
	2. checking the answer		-	-	+

Table 3.20. Sample scoring of the student analyzed in Table 3.19.

Category	Grading
RDS	**Ph: 0.40; Pre: 0.36**
SDS	**Ph: 0.80; Pre: 0.72**
RSD	**Ph: 0.50; Pre: 0.64**

procedures in actual classrooms and accepted the modifications introduced by the instructor who participated in the experiment.

As before, the study involved the same introductory algebra based course for students interested in health professions (N~200), the same instructor and the same two teaching assistants teaching their respective groups (one TA taught groups A and B, and the other taught groups C and D). Including both TAs, the recitations were distributed into control groups and three self-diagnosis treatments groups each of which carried out the self-diagnosis task with different scaffolds. In all treatment group recitations, students first solved a quiz problem, and in the next recitation session they were asked to circle mistakes in their photocopied solutions and explain what they did wrong (see table 3.21 and compare to table 3.1). Group B is split into two subgroups, B and B', that are defined by recitation sections and which received an instructor outline and diagnosis rubric. We report numbers for these groups separately.

The treatments for each group were the same as in the first study. Groups A and A', which are two separate recitations that combined are group A from the companion study, again form the control group and received no intervention. In addition to group B in the first study, another group (group B') received an outline of the problem and the diagnosis rubric demonstrated in figure 3.3. Group C, as in the first study, received a complete written expert solution, and group D was asked to self-diagnose with their text and notebooks.

Table 3.21. Distribution of groups in quiz 7 by self-diagnosis tasks.

Control	Self-diagnosis tasks		
Groups A and A'	Groups B and B'	Group C	Group D
control	Instructor outline, diagnosis rubric	Worked out example	Minimal guidance: notes + text books
~100 students *3 sections*	2 groups: (B)- 31 students, (B')-29 students	1 group (C), 26 students	1 group (D), 23 students

Table 3.22. Physics quiz scores for self-diagnosis groups on quiz 7.

Quiz 7		Control		Intervention B+B' Outline + Rubric		Intervention C Sample solution	Intervention D Minimal guidance
Group		A	A'	B	B'	C	D
Solution - physics	Mean	0.47	0.53	0.45	0.50	0.40	0.50
	Std. Dev.	0.18	0.25	0.15	0.21	0.22	0.18
Solution - presentation	Mean	0.40	0.40	0.49	0.43	0.36	0.38
	Std. Dev	0.12	0.1	0.12	0.13	0.07	0.13

Table 3.22 shows students' scores on the quiz for both physics and presentation. Interestingly, despite the hypothesis that this problem would be a relatively easy quiz problem, students still struggled on the quiz problem, which involves two fundamental physics principles. One common tendency was to fail to invoke the momentum conservation principle altogether and implement mechanical energy conservation assuming that the friend's running velocity was the same as the velocity of the friend on the skateboard. The relatively poor performance on the quiz across all groups was not only true for the "physics" part but also for the "presentation" part where students were given scores on their problem-solving strategy. The difficulty with this quiz problem suggests that even substituting a difficult concept such as Newton's 2nd Law in a non-equilibrium application involving centripetal acceleration with a one-dimensional conservation of linear momentum does not make the quiz problem significantly easier for the students, since students tend to focus only on one part in a two part problem.

3.3.5 Self-Diagnosis Findings

To compare the groups on their self-diagnosis activities, we again performed ANCOVA analysis taking the quiz physics grade as a covariate (after checking homogeneity of slopes). We were interested in finding out what students were able to diagnose in each group.

3.3.5.1 Self-diagnosis - Physics

The physics self-diagnosis grade reflected both the *expert ideal knowledge,* as well as the *novice knowledge per se* (the ideas the student believed were necessary to solve the problem)**,** by reflecting in detail which mistakes students could diagnose. This approach allowed us to better differentiate between students.

Tables 3.23 and 3.24 present a comparison of students' performance on self-diagnosis (SD) of physics aspects in the self-diagnosis groups. Note that, even in this quiz problem, all students made mistakes, thus all of them are included in the analysis. Grades were tabulated according to the scoring method outlined in section 3.2.

Tables 3.23 and 3.24 suggest that an easier problem perhaps allowed students to make effective use of whatever resources and tools they were provided even when the scaffolding support merely allowed students to use their notes and textbook. This can be seen from the fact that even group D students did fairly reasonable self-diagnosis unlike the diagnosis performed by group D students in the first study when the quiz problem was extremely difficult.

Table 3.23. Physics grades for students' self-diagnosis on quiz 7.

	Group B	Group B'	Group C	Group D
Mean	0.56	0.70	0.62	0.61
Std. Err.	0.056	0.064	0.06	0.065

Table 3.24. P Values and ANCOVA analysis for students' self-diagnosis of physics on quiz 7.

	Group B	Group B'	Group C	Group D
Group B		0.14	0.45	0.62
Group B'			0.44	0.35
Group C				0.84

Table 3.25. Self-diagnosis results for physics subcategories.

Explanation of symbols is as follows. "+" represents a correct diagnosis; "+/-" represents a partially correct diagnosis; and "–" represents an incorrect diagnosis or no diagnosis performed. "Total" refers to the total percentages of students who had mistakes in their quiz regarding some subcategory; the students who got a subcategory correct were not included.

Subcategory	Group B			Group B'			Group C			Group D		
	+	+/-	-	+	+/-	-	+	+/-	-	+	+/-	-
Invoked	29%	23%	48%	10%	22%	68%	29%	17%	53%	30%	36%	34%
	Total: 39%			Total: 29%			Total : 48%			Total: 46%		
Applied	57%	43%	0%	40%	40%	20%	100%	0%	0%	50%	50%	0%
	Total: 11%			Total: 8%			Total : 8%			Total: 4%		

Table 3.25 presents students' performance of self-diagnosis in invoking the correct physics principles and applying them (i.e., the percentage of students who diagnosed their mistakes out of those who made mistakes in each sub category).

In the first study, for the self-diagnosis of quiz 6, it was easier to identify mistakes in invoking principles than it was to find mistakes in applying those principles, and also it was easier to invoke a correct principle than it was to apply the principle correctly. Here, the opposite is true. For all three groups, there were more students unable to invoke all the correct principles than there were students who invoked principles but failed to apply them correctly. In particular, it appears that many students simply overlooked conservation of momentum and only addressed conservation of energy. This finding is consistent with earlier studies on the ballistic pendulum (Singh and Rosengrant 2003) in which a majority of students either invoked the conservation of mechanical energy principle or the conservation of momentum principle but not both because they were completely focused on one principle only. Another possible reason for this is that the students felt the mass of the skateboard was negligible compared to the mass of the friend and therefore saw no need to invoke momentum conservation; however, the problem specifically asked the reader to not neglect the mass of the skateboard.

At least 50% of the students who applied physics principles incorrectly were able to address the error in all cases. Everyone who missed the physics principles in group C, which received the complete expert solution, was able to self-diagnose correctly. The more robust self-diagnosis for quiz 7 compared to quiz 6 may be due to quiz 7's relative ease as compared to the problem used for quiz 6. Once students were aware of their errors in applying correct principles they tended to at least partially address these errors even with minimal scaffolding tools.

3.3.5.2 Self-diagnosis - Presentation

The presentation self-diagnosis grade is again divided into the subcategories of description, planning, and checking similar to the subcategories in the first study. Tables 3.26 and 3.27 present a comparison of students' performance on self-diagnosis of the presentation in the self-diagnosis groups. The ANCOVA analysis (again, taking the quiz physics grade as covariate) shows that the presentation-self-diagnosis grades of groups D and C once again differ from those of either group B or B' (p value < 0.05). However, this time there is a slightly higher grade for groups C and D. This may be a result of students having already performed self-diagnosis in quiz 6. However, this is still not a very large improvement, which combined with the similar relevant findings in the previous study offer the argument that one-time diagnosis is

Table 3.26. Grades for the self-diagnosis for presentation for quiz 7.

	Group B	Group B'	Group C	Group D
Mean	0.50	0.35	0.26	0.25
Std. Err.	0.028	0.032	0.031	0.03

Table 3.27. P values for ANCOVA analysis of self-diagnosis for presentation.

	Group B	Group B'	Group C	Group D
Group B		0.001	<0.0001	<0.001
Group B'			0.03	0.029
Group C				0.94

Table 3.28. Self-diagnosis results for presentation subcategories. The scoring notation is the same as in table 3.25.

Subcategory	Group B			Group B'			Group C			Group D		
	+	+/-	-	+	+/-	-	+	+/-	-	+	+/-	-
Description	7%	91%	2%	8%	76%	16%	2%	77%	21%	0%	61%	39%
	Total: 59%			Total 74%			Total: 77%			Total: 84%		
Plan	26%	27%	48%	23%	0%	77%	27%	5%	69%	32%	6%	61%
	Total : 62%			Total: 61%			Total: 68%			Total: 63%		
Check	23%	70%	7%	32%	36%	32%	0%	42%	58%	0%	73%	27%
	Total : 97%			Total: 97%			Total: 100%			Total: 100%		

not enough to get students to consider their problem-solving strategy. Clear instruction requiring students to provide specific details may be necessary.

Table 3.28 presents a between-group comparison of the percentage of students who diagnosed their mistakes regarding the different sub categories in presentation-self-diagnosis. Students struggled in describing the problem, particularly in groups C and D. Almost everyone performed a partial diagnosis of missing problem description of the problem, and very few if any completely diagnosed problem description errors correctly.

Students did not usually check their answers. This led to most students only performing a partial diagnosis or no diagnosis at all in that portion of the self-diagnosis. However, groups C and D performed more poorly than groups B and B' did.

In the planning section, the overall scores were quite low regardless of group, and most students did not diagnose their errors. With the exception of group B, which had a higher proportion of partially correct diagnoses, the majority of students in each group did not address errors in planning. It is possible that they considered the plan for the problem solution sufficiently straightforward that it did not need to be addressed.

In general, students in all groups again performed poorly on diagnosing errors in problem presentation. Groups B and B' performed slightly better than groups C and D on problem description and checking. Again, however, this may be attributed to the additional scaffolding they had in order to perform the self-diagnosis.

3.3.6 Self-diagnosis vs. Prior Knowledge

Table 3.29 presents correlations (for all groups and between groups) between the physics scores in the quiz and the physics self-diagnosis scores, and between the presentation scores and the presentation self-diagnosis scores. Here we were interested to see how students' performance in self-diagnosis depended upon their prior quiz performance. One notes a small-to-moderate positive correlation between the quiz grade and self-diagnosis grade for both physics and presentation. In both cases, students who make fewer mistakes are better at diagnosing them. This means that the initial gap between students who did well and those who did not remained. Therefore the students who performed poorly on the quiz were not helped as much by the self-diagnosis task on this problem as those who performed better on the quiz. Since group D students (with only textbook and notes) were able to perform as well on self-diagnosis on the physics part as the other groups, a high level of scaffolding was not as much of a factor in the physics part of the self-diagnosis task for this quiz problem that involved two simpler physics principles.

Table 3.29. Correlations between quiz scores and self-diagnosis scores for all students over physics and presentation categories.

Physics: Quiz vs. SD		Presentation: Quiz vs. SD	
Correlation	p value	Correlation	p value
0.23	0.01	0.39	<0.0001

3.3.7 Effect on Midterm Performance

Before we discuss the effect of self-diagnosis on the performance on the midterm exam, we note that all treatment groups performed reasonably good self-diagnosis on the physics part (no statistical difference between group B and the other treatment groups). Research focused on making analogies (Gick and Holyoak 1983, Eylon and Helfman 1984) shows that many students don't know how to use a worked out example to solve a transfer problem (similar in required general procedure (principles/intermediate variables), different in detailed procedures or surface features). Students' representation, organized around surface features (Chi et a1. 1981, Eylon and Reif 1984) prevents students from retrieving and implementing procedures from the worked out example. Medium and high achieving students benefited most from instruction explicitly presenting them with the procedures and worked out examples rather than worked out examples without the procedures (Eylon and Helfman 1984). Based upon these studies, we hypothesized that diagnosing one's own mistakes using a solution outline and a rubric that focuses student's attention on the procedure will enhance transfer to isomorphic problems in midterm exams.

As stated earlier, the midterm problem paired with the quiz 6 problem (discussed in section 3.2) was a far transfer problem on which students did not perform well. Therefore, we investigated a midterm problem which was a nearer transfer problem for the quiz 7 problem discussed in the earlier section. Table 3.30 compares both the quiz problems and the corresponding midterm problems that were used to investigate transfer from the self-diagnostic activities on the quizzes. Students are less likely to identify the quiz 6 problem as isomorphic to the corresponding midterm problem than they are to identify the isomorphism between the quiz 7 problem and the corresponding midterm problem.

3.3.7.1 Implication of Implementation in Actual Classrooms

We will now analyze the effect of self-diagnostic activities on quiz 7 by investigating transfer to an isomorphic midterm problem both in terms of retaining the corrected ideas (i.e., invoking and applying appropriately the physics principles and concepts required), and justifying the solution in a manner reflecting a strategic problem solving approach in retaining the proper

Table 3.30. Comparison of quizzes 6 and 7 with their isomorphic midterm problems.

Paired problem set	Activity	Principles	Variables	FBD	Context	Details
Quiz 6 and midterm	Quiz	Energy conservation, Newton's 2nd law in non-equilibrium setting	v a_c Force: Normal (N) /Tension (T)	↑ Normal, tension force ↓ gravity	Roller coaster	Normal force in opposite direction as centripetal acceleration
	Midterm				Tire swing	Tension force in same direction as centripetal acceleration
Quiz 7 and midterm	Quiz	Energy conservation, momentum conservation	v_i h		Skateboard	v_i : horizontal component only
	Midterm				Car	v_i: horizontal and vertical components

characteristics of the presentation of reasoning. But before that, we note that the study was performed in the context of a semester long physics course with lectures, recitations and exams. One must therefore consider additional self-diagnosis activity which may occur after the strictly individual in-class self-diagnosis activity. This additional self-diagnosis activity may be individual or cooperative.

For the midterm problem corresponding to quiz 7, there was almost a month's time between the self-diagnosis task in which students reflected on the quiz 7 problem and the midterm exam. During this period of time, the solution to the quiz 7 problem was posted on the course website and was accessible to all students. As in the companion study, students could look at the solution to the quiz problem on their own before the midterm exam and continue analyzing the quiz 7 problem alone or with others. These considerations must be addressed when dealing with the impact of self-diagnosis on the midterm exams. It is therefore possible that the midterm performance on the transfer problem includes students' effort beyond the scope of the self-diagnosis in-class exercise.

Moreover, as in the companion study, an outcome of the decision to perform the experiment in actual classrooms is that we must consider the effect of the differences in TAs and the effect of interactions within the groups. It is possible that the teaching assistants' different styles as well as the interaction within each group have introduced differences in performances of different groups. Both inter-group (between-subjects) and intra-group (within-subjects) effects will be examined. We will further compare each TA's groups separately so that the difference between TA styles is not relevant.

3.3.7.2 The Midterm Problem

The problem used in the midterm exam, which was a month after the self-diagnosis exercise is described in figure 3.12. In the midterm problem, instead of jumping horizontally onto a skateboard, Fred Flintstone is shown jumping into his car at an angle of 45 degrees with the downward vertical. The students are again asked to find the initial velocity of the car with Fred in it and the maximum height that the car will roll up a hill afterwards. This problem is

Fred Flintstone just got off work, and exits in his usual way, sliding down the tail of his dinosaur and landing in his car (see Figure). Given the height of the dinosaur (h=10 m), it's not hard to calculate his speed v as he enters his vehicle.
Conservation of energy yields the following equation: mgh=1/2 mv2, where m=100 kg is Fred's mass and v is his speed. Algebraic manipulation yields $v = \sqrt{2gh}$ =14 m/s. Judging from the picture taken in Figure 1E, the angle at which Fred enters the car is approximately 450. (a) If the mass of the car is M=200 kg, find the speed with which Fred is driving in the last frame (Figure 1F), assuming he hasn't used his feet to pedal.
(Remember also that there are no fossil fuels since there are no fossils yet.)
(b) Assuming that there is no friction or air resistance, determine the maximum height H that Fred and his car can travel without extra pushing.

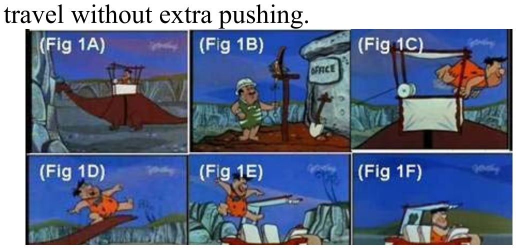


Figure 3.12. Midterm problem paired with quiz 7.

similar to the quiz 7 problem in that it employs the same physical principles, i.e., the conservation of linear momentum and the conservation of mechanical energy. Thus, the two problems are isomorphic in the physics principles involved.

3.3.7.3 Midterm Results: Physics

The various treatment groups are the same as shown in Table 3.15. As with the self-diagnosis exercise, we separated the researcher's judgment of the students' self-diagnosis and solution into physics and presentation grades.

Table 3.31 shows the mean physics score for all groups on the midterm problem. To be able to consider the effect of the TAs on the inter-group comparison, we present analysis of each TA’s groups separately. Table 3.32 shows ANCOVA p-value comparisons between group B and the control group and between group C and group D, respectively. They show that group B,

Table 3.31. Means and standard deviations of each group's midterm physics scores.

	First TA			Second TA	
Group	A	B	B'	C	D
Mean	0.64	0.66	0.68	0.72	0.76
Std. Err.	0.025	0.04	0.048	0.058	0.043

Table 3.32. P-value comparison between midterm physics scores of each TA's groups.

First TA	Group B	Group B'
Group A	0.73	0.4
Group B		0.67

Second TA	Group D
Group C	0.94

Table 3.33. Correlation between self-diagnosis and midterm physics scores for each group.

Group	Correlation	p value
B	-0.26	0.18
B'	0.26	0.24
C	0.32	0.14
D	0.53	0.01

which was provided a rubric and solution outline, did about as well as the control group A. This finding is interesting since the scaffolding provided by self-diagnosis made no difference on midterm performance. As noted earlier, since this midterm exam was one month later (instead of a few days later in the previous study) and students were provided the written solution for the relatively easier Quiz 7 problem, learning outside of the in-class self-diagnosis may be responsible for the midterm performance.

In the previous study, despite actually seeing the complete expert solution, group C had fared worse on the midterm than group D, which was asked to self-diagnose using textbook and notes. In this midterm transfer problem, however, groups C and D both perform equally well.

Additionally, all groups performed better on the midterm than the recitation quiz problem and performed roughly equivalently. This finding suggests that the transfer was sufficiently close that students were able to take advantage of whatever scaffolding they received as well as the posted solution on the course website. The extra time that the students had between the self-diagnosis task and the midterm is also a possible factor.

The intra-group comparison is shown in Table 3.33 in the form of a correlation between the self-diagnosis physics scores of the treatment groups and their midterm scores. There is a moderately strong positive correlation for group D, and correlations for the other groups are not statistically significant. This indicates that those who performed better on self-diagnosis of physics errors in group D performed better on the midterm. The positive correlation for group D makes sense because the textbook and notes were more useful for self-diagnosis of the more typical quiz 7 problem than for the quiz 6 problem, and those who performed self-diagnosis well were more likely to do better on the midterm problem. By comparison, in the companion study there was no correlation for group D. All other intra-group comparisons yield no significant correlations, which may suggest that, between the self diagnosis task and the midterm which were one month apart, a separate learning process may be taking place.

While group D students were able to self-diagnose at roughly the same level as the other groups and were able to transfer to the midterm problem, it is possible that a higher level of scaffolding would be useful for students who might still struggle with the self-diagnosis task with only the textbook and notes. While it is true that group D did about as well as group C on both the self-diagnosis task and the midterm problem, our data suggest that the better the student was at self-diagnosis, the better he/she was likely to perform on the midterm. If a student struggled with the self-diagnosis task, it was more likely that the student would also struggle on the midterm problem.

3.3.7.4 Midterm Results: Presentation

Table 3.34 gives the mean presentation scores for all groups on the midterm problem, and an inter-group analysis of p-values between the self-diagnosis groups is shown in table 3.35. While there is a significant difference between groups A and B', the raw gain is about 5 percentage points. There are no other statistical differences between any of the groups. Similar to the midterm presentation performance after Quiz 6 self-diagnosis, there was no effect of the treatment on presentation performance for group B on the midterm exam. However, groups C and D improved from their self-diagnosis performances on quiz 7. An intra-group analysis, presented in table 3.36, shows no correlation between the midterm scores and self-diagnosis scores. Again, the overall scores are somewhat poor; however, they are still somewhat higher than the scores for quiz 6.

The data report shows that group B fared about the same as the other groups with regard to presentation score on the midterm despite a better performance on presentation self-diagnosis. This finding and the finding of the first part of the study with an atypical problem in section 3.2

Table 3.34. Means and standard deviations of each group's midterm presentation scores.

	First TA			Second TA	
Group	A	B	B'	C	D
Mean	0.47	0.48	0.53	0.45	0.50
Std. Err.	0.013	0.016	0.022	0.027	0.024

Table 3.35. P-value comparison between midterm presentation scores of each TA's groups.

First TA	Group B	Group B'
Group A	0.57	0.014
Group B		0.11

Second TA	Group D
Group C	0.09

Table 3.36. Correlation between self-diagnosis and midterm presentation scores for each group.

Group	Correlation	p value
B	0.20	0.30
B'	0.24	0.24
C	-0.024	0.91
D	0.22	0.32

support the assertion that the “one-time” nature of the intervention was not enough to help students develop effective problem-solving strategies. The self-diagnosis task discussed in this chapter is only the second attempt by students at the self-diagnosis exercise, and students need a more sustained intervention throughout the semester to develop presentation skills, which are a habit of mind. It would be of interest to examine whether consistent intervention in the recitations throughout the semester would help students develop presentation skills.

3.4 SUMMARY AND CONCLUSIONS

3.4.1 Physics: Quiz and Self Diagnosis

In sections 3.2 and 3.3, we examined introductory physics students' performance on self-diagnostic tasks that involved different levels of scaffolding in different recitation groups. The first quiz problem chosen for self-diagnosis was a non-traditional context-rich problem discussed in section 3.2. The second quiz problem chosen for self-diagnosis, described in section 3.3, was also context-rich involving two major principles of physics. However, the application of each of the two principles was not as difficult as the application of Newton’s second law in the non-equilibrium situation involving centripetal acceleration in the first quiz problem

In both quiz problems, students’ initial quiz performance was relatively poor, even though they showed marginal improvement on the second quiz problem discussed in the second study (mean physics score over all groups ~ 0.45) as opposed to the more difficult problem in the first study (mean physics score over all groups ~ 0.35). In fact, a few students explicitly told the instructor that they do not expect a quiz problem with two combined physical principles because they find such problems very difficult. A majority of students also showed a poor presentation performance because introductory students have not yet learned an organized problem-solving approach.

However the self-diagnostic performance differed between experiments in several ways, especially in diagnosing physical concepts. While for the first study (involving quiz 6 and the related midterm problem) there was a definite difference between groups on self-diagnosing physical errors, this difference did not appear in the second study (involving quiz 7 and the related midterm problem). Group D particularly stood out as able to recover with only textbook and/or notebook support in the second study, but was unable to do the same in the first study.

Reflecting back on students' performance on both self-diagnosis tasks, we infer that the level of scaffolding that is sufficient to self-diagnose mistakes depends upon the difficulties in applying physics principles. For example, group D, which was only provided textbook and notes as scaffolding tools, was unable to use these inadequate tools effectively for self-diagnosis of mistakes on the first problem. On the other hand, the same students were able to use their textbook and notes to diagnose their mistakes effectively in quiz 7, which involved conservation of mechanical energy and conservation of momentum principles.

In fact, examples of solved problems that involve either momentum conservation with completely inelastic collision or the conservation of mechanical energy exist in the textbook students used. The instructor also worked out example problems in class related to each of these principles separately. Moreover, one solved example in the textbook was about the ballistic pendulum in the chapter on momentum, which is after the chapter on energy. Although the surface features of the quiz problem which involves a person jumping on a skateboard and climbing up a hill are different from those of the ballistic pendulum problem, both solutions involve both momentum and mechanical energy conservation. The fact that students who used their notes and textbook did well on self-diagnosis suggests that, for this problem, the usual diagnostic tools (textbooks and notes) helped students diagnose their mistakes.

Prior studies related to learning from solved examples (e.g. Pirolli and Recker 1994, VanLehn 1996, Singh 2008a, b) suggest that students can take advantage of solved examples and solve other problems when the problems they are asked to solve are sufficiently similar to the example problems. However, when students have to make analogies and solve problems, which do not have the same surface features, students have more difficulty in solving transfer problems. For the second quiz problem discussed in this chapter, the average physics self-diagnosis score for group D was 61%, which is comparable (with no statistically significant difference) to the self-diagnosis scores of groups B and C that got higher levels of scaffolding.

One possibility is that the students in group D took advantage of the ballistic pendulum problem which had different surface features from the quiz problem but which also involved both the conservation of momentum and conservation of mechanical energy principles. In this case, the students must be able to see the underlying similarity of the quiz problem and the ballistic pendulum problem. Our earlier study (Singh and Rosengrant 2003) asking students to categorize introductory mechanics problems based upon similarity of solution shows that a majority of introductory students do not group together the problem given in quiz 7 and the ballistic pendulum problem. Thus, if browsing over the example of the ballistic pendulum in the textbook (Cutnell and Johnson 2007) helped students transfer the applications of the conservation of mechanical energy and momentum principles to the problem in quiz 7, students have indeed done very good self-diagnosis.

Another possibility is that some students in group D who were self-diagnosing quiz 7 browsed over one solved example on conservation of momentum in the textbook or notes and another solved example on conservation of mechanical energy from another chapter of the textbook or notes and then combined this learning to diagnose the mistakes in the quiz problem. These students' ability to combine these principles together from two different solved examples is commendable especially since the textbook and notes had these two principles discussed in different chapters and no example was explicitly done in the class that combined both principles. In this sense, students were able to synthesize knowledge from different solved examples to diagnose their mistake.

Comparison of the self-diagnosis on quiz 6 and quiz 7 shows that on quiz 6, a majority of students who had difficulty in invoking both physics principles (conservation of mechanical energy and Newton's second law) were able to self-diagnose their mistakes in invoking but in quiz 7, a large number of students were not able to self-diagnose their mistakes in invoking one of the two principles. The difficulty in self-diagnosing the mistake in invoking the momentum conservation principle was common for quiz 7 among students of all groups despite the fact that the problem statement explicitly asked students not to ignore the mass of the skateboard. In fact, group D, which was provided the least amount of scaffolding tools (textbook and notes) performed the best in self-diagnosing mistakes in invoking the physics principles involved in quiz 7. The better self-diagnosis in invoking physics principles by group D compared to other groups may be due to the fact that group D was most cognitively engaged in the self-diagnosis process. Group C was given an elaborate expert solution which many students did not use effectively for self-diagnosis. Group B, which obtained the TA outline before self-diagnosis, also had many students who did not carefully consider their mistakes in invoking physics principles in quiz 7. More research is needed to understand why students in groups B and C were not as effective as group D in self-diagnosing their mistakes in invoking physics principles despite being given more elaborate scaffolding tools. Two thirds of the students in group D were able to do it whereas a majority of students in groups B and C were unable to self-diagnose even with the tools provided.

Thus self-diagnostic performance on physics improved for all groups in the second study as opposed to the first study, but students who did not invoke a principle struggled with identifying the missing principle in the second study. However, very few students made application errors in quiz 7 and the main source of error in quiz 7 was the failure to recognize one of the two principles. In particular, students omitted conservation of linear momentum and neglected the mass of the skateboard despite being explicitly told not to ignore it. Comparison of students' mistakes in applying physics principles in quiz 6 and quiz 7 shows that approximately 60% of the students (including all groups) were unable to apply the physics principles correctly even if they invoked them in quiz 6, whereas for quiz 7, more than 90% of the students who invoked a physics principle were able to apply it correctly. Moreover, every student who could not apply the physics principle correctly on his/her own was able to self-diagnose the mistakes in quiz 7.

The reason students had so much difficulty in applying physics principles in quiz 6 has to do with the fact that one of the principles involved Newton's second law in a non-equilibrium situation involving centripetal acceleration. Newton's second law in a non-equilibrium situation involving centripetal acceleration is difficult for students, for example, due to mental models that students may have prior to learning the material (McCloskey 1982, Itza-Ortiz et al. 2004). Arguably, Newton's second law in a non-equilibrium application in quiz 6 is more difficult to approach for introductory students than conservation of linear momentum in a one dimensional problem in which the vector nature of momentum was not explicit in quiz 7. Moreover, the conservation of mechanical energy principle was common to both problems and some students may recognize that this principle was missing more easily on the second attempt.

In fact, students in algebra-based physics courses are not used to such an atypical non-equilibrium problem and they particularly have difficulty with the concept of centripetal acceleration. Also, many students in quiz 6, which asked for the reading of the scale on which the girl was sitting, were unable to determine the target variable correctly. They did not realize that they had to find the normal force that the scale exerts on the girl. It should be noted that, while a majority of the students in all groups were able to self-diagnose their mistakes at least partially in applying physics principles in quiz 6, the percentage of students who were able to self-diagnose all their mistakes in applying physics principles was less than 15%. Many students in groups B and C still treated the problem as an equilibrium problem or treated centripetal force as a physical force despite browsing over the expert solution or TA's outline. An analysis of students' self-diagnosis shows that students considered these types of considerations to be minor issues. For example, students in groups B and C often wrote that they made a "math" mistake and got the sign wrong when in reality their mistake was a "physics" mistake and the sign was wrong because the centripetal acceleration multiplied with mass was treated as a physical force. Future research will focus on how to help students focus on these issues more deeply. One strategy we will investigate is the effect of students carrying out the self-diagnosis with peers.

Comparing group D's performance on the two self-diagnosis task, we find that group D performed significantly better on diagnosing errors in physical concepts in the second quiz than in the first quiz. The material in the notes and textbooks was more accessible for one-dimensional momentum conservation and mechanical energy conservation than it was for a non-equilibrium application of Newton's second law involving centripetal acceleration. The scaffolding tools (textbook and notes) that students in group D were provided were adequate for many students in group D for the level of difficulty of quiz 7 but not for the level of difficulty for quiz 6. The non-equilibrium application of Newton's second law in quiz 6 was so difficult that many students were unable to self-diagnose mistakes in identifying the target variable, dealing with the vector nature of forces, the non-equilibrium nature of the problem and deciphering the centripetal acceleration. Both studies together point to the importance of knowing students' prior knowledge and providing tools commensurate with their zone of proximal development to scaffold their learning.

3.4.2 Physics: Midterm-exam Performance

The comparison of the midterm exam scores in the two parts of the study shows that the midterm physics scores improved significantly in the second study compared to the first study for all groups (0.64 < physics average < 0.76 in the second midterm transfer problem). We hypothesize that several factors may have contributed to this increase in the second midterm score on the transfer problem compared to the first midterm score. One reason for the better performance could be that the second midterm problem was a near transfer problem to the quiz 7 problem while the first midterm transfer problem discussed in the companion problem was a far transfer problem to the quiz 6 problem. Moreover, for the first midterm transfer problem, students had to realize that the tension is largest at the lowest point of the pendulum's trajectory (not at the highest point). If students did not realize this fact then the first midterm problem cannot be viewed as a transfer problem to the quiz 6 problem. Such initial considerations were not required before the students could make the transfer from quiz 7 to the second midterm problem, which may have made the transfer easier. In addition, it appears that applying conservation of momentum in the second midterm problem was easier than applying Newton's second law in a non-equilibrium context in the first midterm problem, which may have made the performance worse on the transfer problem on the first midterm than the second midterm. In fact, as discussed earlier, very few students could completely correctly diagnose their mistakes in "applying physics principles" in quiz 6 involving non-equilibrium application of Newton's second law with centripetal acceleration.

Moreover, while the self-diagnosis activities for quiz 6 took place earlier in the same week in which the first midterm was given, there was almost one month between the self-diagnosis activities for quiz 7 and the second midterm exam in which the transfer problem to quiz 7 was given. Within one month students had much more opportunity to learn about relevant concepts in the transfer problem from other sources including the correct solution that was posted on the course website. The opportunity to learn from other sources over a one month period may be why group A (control) performed quite reasonably as well on the transfer problem. This is perhaps because all students ultimately had access to the correct solution. While one can argue that within a one month period students can also forget what they learned from self-diagnosis activities, we hypothesize that this is less likely to be an issue because students' difficulties with quiz 7 were mostly related to invoking physics principles but they did not have difficulty in applying conservation laws unlike the difficulty in applying Newton's second law for a non-equilibrium situation. In fact, the second midterm performance of all groups on the transfer problem is significantly better than their performance on quiz 7.

Considering the midterm performance of group D, which was the group with the least amount of scaffolding tools in both parts of the study, we find that group D performed comparably to the other groups on the midterm attempts in both studies. Considering the fact that group D's self-diagnosis was significantly poorer on quiz 6 than on quiz 7, one reason for the improvement of group D is that having to struggle with the quiz 6 problem twice (once to solve it and another time to self-diagnose mistakes using textbook and notes) may have spurred a more determined effort outside of the recitation to learn how the problem is done. In the first part of the study, group D struggled with the non-typical quiz 6 problem both on the quiz and on the self-diagnosis. However, in the second part of the study, group D was able to self-diagnose mistakes on quiz 7 with the textbook and notes, and their self-diagnosis scores are comparable to the other experimental groups (groups B and C). In fact, the performance on the transfer problem in the second midterm ranged from 64% for the control group to 76% (highest) for group D. While we cannot directly compare group B and group D on midterm exams since they had different TAs, it is possible that group D experienced more cognitive involvement in their self-diagnostic activity using textbook and notes than groups B and C did regardless of resulting midterm performance, which would be another example of efficiency combined with innovation (Schwartz et al. 2005) making for more effective learning. This may be supported by the moderately strong correlation between physics self-diagnosis and corresponding midterm performance of group D in the second study. Table 3.27 shows that students in group D who performed the self-diagnosis task well were more likely to perform better on the second midterm. By contrast, students in group C had an elaborate expert solution as the scaffold for self-diagnosis but their cognitive involvement in self-diagnosis may have been less deep.

3.4.3 Presentation

Both parts of the study indicate that group B performed significantly better at self-diagnosis of the presentation part than groups C and D, and that the latter two groups were very poor at self-diagnosing the deficiencies in their presentation. This implies that introductory students need a deliberate scaffold, in the form of a rubric, in order to improve their presentation skills. However, groups C and D were able to perform slightly better in the second part of the study than they did in the first part (with an atypical problem). It is possible that they may slowly be able to better diagnose their presentation approach with repeated study.

However, in both parts of the study, students (including those in group B) did not improve significantly in their average midterm presentation score compared to their average quiz presentation score (midterm performance for all groups 0.45 < presentation average < 0.53). This makes sense in the light that there were only two attempts at self-diagnosis. Learning effective problem presentation skill is a habit of mind and may require a sustained intervention as opposed to one or two individual attempts. Thus, while group B students were significantly better at self-diagnosis of the presentation part than the other groups when they were provided with a rubric, without the rubric on the midterm their performance on presentation was quite poor and comparable to others. It would be useful to attempt this sort of exercise regularly over the course of a semester and see if it may benefit students' problem-solving approach.

3.4.4 Future Considerations

These findings suggest that, as problems become sufficiently novel or difficult for a student, more scaffolding is needed to correctly self-diagnose the mistakes and learn from them. If the material is more familiar and simpler for a student, then less scaffolding may be sufficient to help students self-diagnose their mistakes. However, there wasn't a strong correlation between the level of self-diagnosis and students' performance on the midterm (e.g., in the first study, group D performed poorly on self-diagnosis activities but performed well on the midterm exams). We hypothesize that this may be due to additional diagnostic activities that may have taken place outside the classroom, individually or with others, because group D students were particularly frustrated at not being able to solve the quiz 6 problem or diagnose their mistakes with the textbook or notes. It may be useful to carry out control studies outside of the classroom where the transfer problem is presented right after the self-diagnostic activities to students in different intervention groups. Such a control study, while not authentic, will help separate the outside of the classroom learning from what was learned from self-diagnosis activities.

Moreover, in the first part of the study, we coupled self-diagnostic activities on a very difficult quiz 6 problem with a far transfer problem which had very different surface features and found that the transfer was poor for all groups. Similarly, in the second study, we coupled self-diagnostic activities on a context-rich problem involving two parts which was relatively easier with a near transfer problem on the midterm. In the second case, we found that the students in all groups performed significantly better on the second midterm transfer problem than on the quiz 7 problem. To further understand the correlation between problem difficulty, self-diagnosis and transfer, another experiment may be devised in which a higher difficulty problem (such as the quiz 6 problem) is coupled with a near transfer problem and a lower difficulty problem (such as the quiz 7 problem) is coupled with a far transfer problem. For example, the self-diagnosis on the roller-coaster problem in quiz 6 can be coupled with a transfer problem which is very much like the roller-coaster problem and the self-diagnosis on the quiz 7 problem can be coupled with a ballistic pendulum problem which requires conservations of mechanical energy and momentum but has different surface features than quiz 7. These studies will help us better understand the effect of self-diagnosis on later performance.

In addition, it will be useful to investigate the effect of sustained intervention over the duration of the course. Students may improve more significantly, both in diagnosing errors in physical principles and problem-solving strategy and in applying that learned knowledge to a future similar problem. It will also be useful to consider scaffolding mechanisms and implement strategies to investigate improvement in students' self-diagnostic skills over the course of a semester.

3.5 ACKNOWLEDGMENTS

We thank J. Levy for allowing us to conduct this study in his class. The research for both studies was supported by ISF 1283/05 and NSF DUE-0442087.

3.6 REFERENCES

Bereiter, C., & Scardamalia, M. (1989). "Intentional learning as a goal of instruction." In *Knowing, Learning, and Instruction: Essays in honor of Robert Glaser,* Hillsdale, NJ: Lawrence Erlbaum Associates, 361-392.

Bielaczyc, K., Pirolli, P., & Brown, A. (1995). "Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving." *Cognition and Instruction* 13(2), 221-252.

Black, P., & Harrison, C. (2001). "Feedback in questioning and marking: the science teacher's role in formative assessment." *School Science Review* 82(301), 55-61.

Black, P., & Wiliam, D. (1998). "Inside the black box: Raising standards through classroom assessment." London: School of Education, King's College.

Chi, M. (1997). "Quantifying qualitative analysis of verbal data: A practical guide." *J. Learning Sci.* 6(3), 271-315.

Chi, M. (2000). "Self-explaining expository texts: the dual processes of generating inferences and repairing mental models." In *Advances in Instructional Psychology.* Edited by R. Glaser. Mahwah, NJ: Erlbaum Associates, pp. 161-238.

Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). "Self-explanations: How students study and use examples in learning to solve problems." *Cognitive Science* 13, 145-182.

Chi, M., Feltovich, P., and Glaser, R. (1981). "Categorization and representation of physics knowledge by experts and novices." *Cognitive Science* 5, 121-152.

Chi, M., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). "Eliciting self-explanations improves understanding." *Cognitive Science* 18, 439-477.

Chi, M., & VanLehn, K. (1991). "The Content of Physics Self-Explanations." *Journal of the Learning Sciences* 1(1), 69-105.

Collins, A., Brown, J., & Newman, S. (1989). "Cognitive Apprenticeship: Teaching the crafts of reading, writing and mathematics." In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser,* Hillsdale, NJ: Lawrence Erlbaum., 453-494.

Cohen, E., Mason, A., Singh, C., & Yerushalmi, E. (2008). "Identifying differences in diagnostic skills between physics students: Students' self-diagnostic performance given alternative scaffolding." In *Proceedings of the Phys. Ed. Res. Conference, Edmonton, AB, CA,* AIP Conf. Proc. 1064, 99-102.

Cummings, K., Marx, J., Thornton, R., & Kuhl, D. (1999). "Evaluating innovation in Studio Physics." *American Journal of Physics* 67(7), S38-S44.

Cutnell, J., & Johnson, K. (2007). *Physics, 7th Edition.* John Wiley & Sons, pp. 206-209.

Eylon, B., & Helfman, J. (1984). "Analogical problem-solving processes in physics." In *Science Teaching in Israel: Origins, Development, and Achievements.* Edited by A. M. Mayer and P. Tamir. Jerusalem: Israel Science Teaching Center, pp. 259-271. (in Hebrew)

Eylon, B., & Reif, F. (1984). "Effects of knowledge organization on task performance." *Cognition and Instruction* 1(1), 5-44.

Foster, T. (2000). "The development of students' problem-solving skill from instruction emphasizing qualitative problem-solving." Unpublished doctoral dissertation, University of Minnesota, Twin Cities.

Gick, M., & Holyoak, K. (1983). "Schema Induction and Analogical Transfer." *Cognitive Psychology* 15, 1-38.

Heller, J., and Reif, F. (1984). "Prescribing effective human problem-solving processes: Problem description in physics." *Cognition and Instruction* 1, 177-216.

Heller, P., Keith, R., & Anderson, S. (1992). "Teaching problem solving through cooperative grouping. Part I: Group versus individual problem solving." *Am. J. Phys.* 60, 627.

Heller, P., & Hollabaugh, M. (1992). "Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups." *Am. J. Phys.* 60, 637.

Itza-Ortiz, S., Rebello, S., and Zollman, D. (2004). "Students' models of Newton's second law in mechanics and electromagnetism." *Eur. J. Phys.* 25, 81–89.

Larkin, J. (1979). "Processing Information for Effective Problem Solving." *Engineering Education* 70(3), p. 285.

Larkin, J. (1980). "Enriching formal knowledge: A model for learning to solve problems in physics." In *Cognitive skills and their acquisition.* Edited by J. Anderson. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 311-334.

Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980). "Expert and novice performance in solving physics problems." *Science* 208, 1335-1342.

Larkin, J., & Reif, F. (1979). "Understanding and teaching problem solving in physics." *Eur. J. Sci. Ed.* 1(2), 191-203.

Leonard, W., Dufresne, R., & Mestre, J. (1996). "Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems." *American Journal of Physics* 64, 1495-1503.

Maloney, D. (1994). "Research in Problem Solving: Physics." In *Handbook of Research on the Teaching and Learning of Science,* D. Gabel (Ed.), MacMillan.

Mason, A., Cohen, E., Yerushalmi, E., & Singh, C. (2008). "Identifying differences in diagnostic skills between physics students: Developing a rubric." *Proceedings of the Phys. Ed. Res. Conference, Edmonton, AB, CA,* AIP Conf. Proc. 1064, 147-150.

McCloskey, M. (1982). "Naive theories about motion," In *Mental Models.* Edited by D. Gentner and A. L. Stevens. Hillsdale, NJ: Erlbaum, pp. 299–324.

Mestre, J., Dufresne, R., Gerace, W., Hardiman, P., & Touger, J. (1993). "Promoting skilled problem-solving behavior among beginning physicists." *Journal of Research in Science Teaching* 30, 303-317.

Neuman, Y., & Schwarz, B. (1998). "Is self-explanation while solving problems helpful? The case of analogical problem solving." *British Journal of Educational Psychology* 68**,** 15-24.

Perkins, D., & Swartz, R. (1992). "The nine basics of teaching thinking." In *If minds matter: A foreword to the future: Volume 2*. Edited by A. L. Costa, J. Beflanca, & R. Fogarty. Palatine, IL: Skylight Publishing, pp. 53-69.

Pirolli, P., & Recker, M. (1994). "Learning Strategies and Transfer in the Domain of Programming." *Cognition and Instruction* 12(3), 235-275.

Polya, G. (1945). *How to Solve It.* Princeton, NJ: Princeton University Press.

Reif, F. (1995). "Millikan Lecture 1994: Understanding and teaching important scientific thought processes." *Am. J. Phys.* 63, 17-32 (1995).

Reif, F., Larkin, J., & Brackett, G. (1976). "Teaching general learning and problem-solving skills." *American Journal of Physics* 44(3), 212-217.

Reif, F. & Scott, L. (1999). "Teaching scientific thinking skills: Students and computers coaching each other." *American Journal of Physics* 67, 819-831.

Schoenfeld, A. (1992). "Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics." In *Handbook for Research on Mathematics Teaching and Learning*, edited by D. Grouws (McMillan, NY), Chap. 15, pp. 334-370.

Singh, C. (2008a). "Assessing student expertise in introductory physics with isomorphic problems. I. Performance on nonintuitive problem pair from introductory physics." Physical Review Special Topic Physics Education Research 4, 010104.

Singh, C. (2008b). "Assessing student expertise in introductory physics with isomorphic problems. II. Effect of some potential factors on problem solving and transfer." Physical Review Special Topic Physics Education Research 4, 010105.

Singh, C. & Rosengrant, D. (2003). "Multiple-choice test of energy and momentum concepts." *Am. J. Phys.* 71(6), 607—617.

Singh, C., Yerushalmi, E., & Eylon, B. (2007). "Physics learning in the context of scaffolded diagnostic tasks (II): Preliminary results." *Proceedings of the Phys. Ed. Res. Conference, Greensboro, NC,* AIP Conf. Proc. 951, 31-34.

Schwartz, D., Bransford, J., & Sears, D. (2005). "Efficiency and Innovation in Transfer." In *Transfer of Learning: Research and Perspectives.* Edited by J. Mestre. Charlotte, NC: Information Age Publishing, 1-52.

Van Heuvelen, A. (1991). "Overview Case Study Physics." *American Journal of Physics* 59, 898-907.

VanLehn, K. (1996). "Cognitive Skill Acquisition." *Annual Review of Psychology* 47, 513-539.

VanLehn, K., Lynch, C., Taylor, L., Weinstein, A., Shelby, R., Schulze, K., Treacy, D., & Wintersgill, M. (2002). "Minimally invasive tutoring of complex physics problem solving." In *Intelligent Tutoring Systems: 6th International Conference.* Edited by S. Cerri, G. Gouardees, and F. Paraguacu. Berlin: Springer, 367-376.

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Functions*. Massachusetts: Harvard University Press, p. 86.

Wertsch, J. (1985). *Vygotsky and the social formation of mind.* Cambridge, MA: Harvard University Press, p. 164.

Yerushalmi, E. & Eylon, B. (2001). "Teachers' Approaches To Promoting Self-Monitoring In Physics Problem Solving By Their Students." *Proceedings of GIREP meeting, Barcelona, Spain.*

Yerushalmi, E., Mason, A., Cohen, E., & Singh, C. (2008). "Effect of self-diagnosis on subsequent problem solving performance." *Proceedings of the Phys. Ed. Res. Conference, Edmonton, CA,* AIP Conf. Proc. 1064, 53-56.

Yerushalmi, E., Singh, C., & Eylon, B. (2007). "Physics Learning in the context of scaffolded diagnostic tasks (I): The experimental setup." *Proceedings of the Phys. Ed. Res. Conference, Greensboro, NC,* AIP Conf. Proc. 951, 27-30.

4.0 REFLECTION IN UPPER-LEVEL UNDERGRADUATE QUANTUM MECHANICS

4.1 ABSTRACT

One attribute of expert physicists is that they learn readily from their own mistakes. Experts are unlikely to make the same mistakes when asked to solve a problem a second time, especially if they have had access to a correct solution. Here, we discuss a case study in which fourteen advanced undergraduate physics students taking an honors-level quantum mechanics course were given the same four problems in both a midterm and final exam. The solutions to the midterm problems were provided to students. The performance on the final exam shows that, while some advanced students performed equally well or improved compared to their performance on the midterm exam on the questions administered a second time, a comparable number performed less well on the final exam than on the midterm exam. The wide distribution of students' performance on problems administered a second time suggests that most advanced students do not automatically exploit their mistakes as an opportunity for learning, and for repairing, extending, and organizing their knowledge structure. Interviews with a subset of students revealed attitudes towards problem-solving and gave insight into their approach to learning.

4.2 INTRODUCTION

Helping students learn to "think like a physicist" is a major goal of most physics courses from the introductory to advanced level (Van Heuvelen 1991a, Reif 1986, Reif 1981, Larkin and Reif 1979, Reif 1995, Larkin 1981). Expert physicists monitor their own learning and use problem solving as an opportunity for learning, and extending and organizing their knowledge (Maloney 1994, Heller and Reif 1984). Prior research has focused on how introductory physics students differ from physics experts (Chi et al. 1981, Larkin et al. 1980, Singh 2002, Dufresne et al. 2005, Hardiman et al. 1989) and strategies that may help introductory students learn to learn (Eylon and Reif 1984, Van Heuvelen 1991b, Elby 2001, Meltzer 2005, Ozimek et al. 2004). By comparison, few investigations have focused on the learning skills of advanced physics students although some investigations have been carried out on the difficulties advanced students have with various advanced topics, e.g., in quantum mechanics (Singh 2001, Bao and Redish 2002, Wittmann et al. 2002, Singh et al. 2006, Singh 2008).

It is commonly assumed that most students who have made it through an entire undergraduate physics curriculum have not only learned a wide body of physics content but also have picked up the habits of mind and self-monitoring skills needed to build a robust knowledge structure (e.g. Chi et al. 1981). Instructors take for granted that advanced physics students will learn from their own mistakes in problem solving without explicit prompting, especially if students are given access to clear solutions. It is implicitly assumed that, unlike introductory students, advanced students have become independent learners and they will take the time out to learn from their mistakes, even if the instructors do not reward them for fixing their mistakes, e.g., by explicitly asking them to turn in, for course credit, a summary of the mistakes they made and writing down how those mistakes can be corrected (Mason et al. 2008, Cohen et al. 2008, Yerushalmi et al. 2008, Yerushalmi et al. 2007, Singh et al. 2007).

However, such assumptions about advanced students' superior learning and self-monitoring skills have not been substantiated by research. Very little is known about whether the development of these skills from the introductory level until the time the students become physics professors is a continuous process of development or whether there are some discontinuous "boosts" in this process for many students, e.g., when they become involved in graduate research or when they ultimately independently start teaching and researching. There is also no research data on the fraction of students who have gone through the "traditional" physics curriculum and have been unable to develop sufficient learning and self-monitoring skills that are the hallmark of a physicist.

Moreover, investigations in which advanced physics students are asked to perform tasks related to simple introductory physics content do not properly assess their learning and self-monitoring skills. Advanced students may have a large amount of "compiled knowledge" about introductory physics and may not need to do much self-monitoring or learning while dealing with introductory problems. For example, when physics graduate students were asked to group together introductory physics problems based upon similarity of solution, their categorization was better than that of introductory physics students, even though there is a distribution overlap (see chapter 2). While such tasks may be used to compare the grasp that introductory and advanced students have of introductory physics content, tasks involving introductory level content do not shed much light on advanced physics students' learning and self-monitoring skills.

The task of evaluating advanced physics students' learning and self-monitoring skills should involve advanced level physics topics at the periphery of advanced students' own understanding. While tracking the same student's learning and self-monitoring skills longitudinally is an extremely difficult task, taking snapshots of advanced students' learning and self-monitoring skills can be very valuable. Here, we investigate whether students in an advanced quantum mechanics course learn automatically from their own mistakes without explicit intervention.

At the University of Pittsburgh, honors-level quantum mechanics is a two-semester course sequence which is mandatory only for those students who want to obtain an honors degree in physics. It is often one of the last courses an undergraduate physics major takes. Here, we discuss a study in which we administered four quantum physics problems in the same semester both in the midterm and final exams to students enrolled in the honors-level quantum mechanics. Solutions to all of the midterm questions were available to students on a course website. Moreover, written feedback was provided to students after their midterm performance, indicating on the exams where mistakes were made and how they can be corrected.

Our goal was to explore the extent to which these advanced physics students use their mistakes as a learning opportunity (Chi et al. 1989) and whether their performance on the problems administered a second time (in the final exam) is significantly better than the performance on those problems in the midterm exams. We also interviewed a subset of students individually within two months using a think-aloud protocol (Chi 1994, Chi 1997, Ericsson and Simon 1993) to get a deeper understanding of their attitudes and approaches to problem solving and learning (Cummings et al. 2004, Marx and Cummings 2007). Moreover, an evaluation of how well students were able to retrieve relevant knowledge to solve the quantum mechanics problems during the interviews that took place a couple of months after the final exam gives us a glimpse of the robustness of students' knowledge structure (Francis et al. 1998).

We found that students' average performance in the final exam on the problems that were given a second time was not significantly better than the average performance on those problems on the midterm exams. While some students improved, others deteriorated. We conclude that advanced physics students do not routinely exploit their mistakes in problem solving as a learning opportunity. Our study suggests that many advanced physics students may be employing inferior learning strategies, e.g., "cramming" before an exam and selective memorization of content based upon their expectation that those problems are likely to show up on the exam; most do not give a high priority to building a robust knowledge structure. Prior research shows that introductory physics students benefit from explicit interventions to help them develop useful learning and self-monitoring skills (Singh et al. 2006, Singh 2008, Mason et al. 2008, Cohen et al. 2008, Yerushalmi et al. 2008, Yerushalmi et al. 2007, Singh et al. 2007, Karelina and Etkina 2007, Etkina et al. 2008). We hypothesize that similar explicit interventions will also prove useful in advanced courses and will help advanced physics students in developing habits of mind.

4.3 PROCEDURE

The honors-level quantum mechanics course had 14 students enrolled in it, most of whom were physics seniors. The class was primarily taught in a traditional lecture format but the instructor had the students work on a couple of preliminary versions of tutorials that were being developed.

Students were assigned weekly homework throughout the fifteen-week semester. In addition, there were two midterm exams and a final exam. The midterm exams covered only limited topics and the final exam was comprehensive. Students had instruction in all relevant concepts before the exams, and homework was assigned each week from the material covered in a particular week. Each week, the instructor held an optional class in which students could ask for help about any relevant material in addition to holding office hours. The first midterm took place approximately eight weeks after the semester started, and the second midterm took place four weeks after the first midterm. For our study, two problems were selected from each of the midterms and were given again verbatim on the final exam along with other problems not asked earlier. The problems given twice are listed in Appendix B.1.

Three of the problems chosen (problem 2, 3 and 4 in Appendix B.1 where the numbering of the problems corresponds to the order in the final exam) were those that several students had difficulty with; a fourth problem (problem 5) which most students found straightforward on one of the two midterm exams was also chosen. The most difficult of the four problems (based upon students' performance) was problem 4 in Appendix A.1 that was also assigned as a homework problem before the midterm exam but was perceived by students to be more abstract in nature than the other problems. The easiest of the four problems was an example that was solved within the assigned textbook (Griffiths 2005). The students had access to solutions for all homework and midterm problems. Thus students had ample opportunity to learn from their mistakes before they encountered the four problems selected from the midterm exams on their final exam (as noted earlier two problems were selected from each midterm).

A scoring rubric, developed jointly with Yerushalmi and Cohen (Mason et al. 2008, Cohen et al. 2008, Yerushalmi et al. 2008, Yerushalmi et al. 2007, Singh et al. 2007 – also see chapter 3) to assess how well the students in introductory physics courses diagnose their mistakes when explicitly prompted to do so, was adapted to score students' performance on each of the four quantum mechanics problems on both the midterm and final exams. The scoring was checked independently by another scorer and at least 80% agreement was found on the scoring for each student on each problem in each attempt (on midterm and final exams). Students were rewarded for correctly identifying and employing physical principles as well as for their presentation and problem-solving skills.

Thus the scoring rubric has two sections: one scores students on their physics performance and the other scores how well they presented their solutions. The rubric for the presentation part was somewhat different from the corresponding part for introductory physics because quantum mechanics problems often asked for more abstract answers (e.g., proving that certain energy eigenstates are equally probable) as opposed to finding a numerical answer. Therefore some categories in the introductory physics rubric (e.g. writing down units) were omitted from the presentation part of the quantum mechanics rubric and other categories were adapted to reflect the nature of the quantum problems better (e.g., checking the answer was adapted to making a conceptual connection with the results).

In-depth interviews lasting 1-1.5 hours were conducted with four paid student volunteers from the group of 14 students in the following semester within the first two months using a think-aloud protocol (Chi 1994, Chi 1997, Ericsson and Simon 1993). Three of the four interviewed students were enrolled at that time in the second semester course in honors-level quantum mechanics. The fourth student had graduated in the fall semester and was performing research with a faculty member. During these interviews we first asked students about their approaches and strategies for problem solving and learning and asked them to solve the same four problems again while thinking aloud. We did not disturb them initially when they answered the questions and only asked for clarification of points after the student had answered the questions to the best of his/her ability. These delayed interviews also provided an opportunity to understand how well students had retained relevant knowledge after the semester was over and could retrieve it to solve the problems. Two shorter interviews were conducted later with additional students, which mainly focused on students' attitudes and approaches to learning due to the time constraints.

4.3.1 Rubrics and Scoring

Figures 4.1-4.3 demonstrate the scoring rubric for three of the questions along with the score of one student for each question on both the midterm and final exams. Below, we first describe the symbols used for scoring and then give an explanation of how a quantitative score is derived after the initial scoring is assigned symbolically for each sub-part of the rubric. A "+" (worth 1 point) is given if a student correctly completes a task as defined by the criterion for a given row. A "-" (worth 0 points) is given if the student either fails to do the given task or does it incorrectly. If a student is judged to have gotten something partially correct, then the rater may assign a combination of pluses and minuses (++/-, +/-, +/--) to reflect this performance, with the understanding that such a combination represents an average score of pluses and minuses (e.g. ++/- translates to 2/3 of a point). If the student's solution cannot address a criterion then "n/a" (not applicable) is assigned and the criterion is not considered for grading purposes at all. For example, if the student does not invoke a principle, the student will receive a "-" in the invoking

General categories	Specific criteria	Sample student scores	
		Midterm diagnosis of solution	Final exam diagnosis of solution
Invoking appropriate concepts	1) Solution construction: i) spectral decomposition using identity: $\Sigma|\psi_i><\psi_i| = \mathbf{1}$ or ii) using linear combination of complete set of states: $|\psi> = \Sigma c_i|\psi_i>$, where $c_i = <\psi_i|\psi>$	+ (used i)	+ (used ii)
	2) Operator acting on an eigenstate yields the corresponding eigenvalue: $\hat{Q}|\psi_i> = \lambda_i|\psi_i>$	+	+
	3) Complex conjugate of the projection of a general state on to an eigenstate: $<\psi_i|\psi>^* = <\psi|\psi_i>$	++/-	+
Invoking inappropriate concepts	4) valid principles or concepts but not relevant for this problem	n/a	n/a
	5) invalid principles or concepts (e.g., $|\psi_i> = |\psi>$)	n/a	n/a
Applying concepts	1) Inserting identity $\Sigma|\psi_i><\psi_i| = \mathbf{1}$ into scalar product $<\psi|\hat{Q}|\psi>$	-	+
	2) Eigenvalue of an operator can be taken outside bracket	+	+
	3) Probability of measuring λ_i in terms of the projection of a general state onto the corresponding eigenstate and its complex conjugate: $<\psi_i|\psi>^*<\psi|\psi_i>$ or $|<\psi_i|\psi>|^2$	+	+
Organization	Clear/appropriate knowns	+/-	+
Plan	1) Appropriate target quantity chosen 2) Appropriate intermediate variables (λ_i, linear combination of eigenstates of $\hat{Q}$) 3) Consistent plan	+	+
Evaluation	1) completes proof 2) makes connection with results ($<\psi_i|\psi>$ or c_i)	+/-	+
Overall Scores	Physics (%) Presentation (%)	78 67	100 100

Figure 4.1. Sample scoring of student #1 in midterm and final exams on problem 2.

General categories	Specific criteria	Sample student Scores	
		Midterm diagnosis of solution	Final exam diagnosis of solution
Invoking appropriate concepts	1) Taylor expansion definition: $f(x+x_0)=\sum_{n=0}^{\infty}\frac{1}{n!}x_0^n\left(\frac{d}{dx}\right)^n f(x)$	+	-
	2) definition of momentum operator in position space in one dimension: $\hat{p}=-i\hbar(d/dx)$	+	+
	3) Expansion of exponential: $e^u=\sum_{n=0}^{\infty}\frac{1}{n!}u^n$	+	+
Invoking inappropriate concepts	4) valid principles or concepts but not relevant (e.g. expectation values)	n/a	- (Fourier transform)
	5) invalid principles or concepts (e.g. confusing position and momentum space)	n/a	- (reasoning without proof)
Applying concepts	1) Partial derivative in terms of momentum operator: $\partial/\partial x=i\hat{p}/\hbar$	+	+
	2) $e^{i\hat{p}x_0/\hbar}=\sum_n\frac{1}{n!}x_0{}^n\left(\frac{i\hat{p}}{\hbar}\right)^n$	+	++/- (constants incorrect)
	3) Taylor expansion performed correctly to obtain $f(x+x_0)=e^{i\hat{p}x_0/\hbar}f(x)$	+	-
Organization	Clear/appropriate knowns, e.g.: $e^u=\sum_{n=0}^{\infty}\frac{1}{n!}u^n$	+	-
Plan	1) Appropriate target quantity chosen 2) Appropriate intermediate variables chosen 3) Consistent plan	+	-
Evaluation	1) completes proof: $f(x+x_0)=e^{i\hat{p}x_0/\hbar}f(x)$ 2) makes connection with results (momentum operator is generator of translation in space)	+	+/-
Overall Score	Physics (%) Presentation (%)	100 100	48 17

Figure 4.2. Sample scoring of student #6 in midterm and final exams on problem 4.

General categories	Specific criteria	Sample student Scores	
		Midterm diagnosis of solution	Final exam diagnosis of solution
Invoking appropriate concepts	1) Measurement of position collapses the wave function into a delta function	+	+
	2) Expansion of general state in terms of energy eigenstates: $\Psi(x) = \Sigma c_n \psi_n$	-	-
	3) Fourier trick for probability amplitudes which are coefficients of energy eigenstates in expansion: $c_n = \int \psi_n * \psi(x) dx$ over all space (optional deriv.)	+	-
	4) solve for probability of measuring a given energy E_n	+	-
Invoking inappropriate concepts	5) valid principles or concepts but not relevant (e.g. expectation values)	n/a	n/a
	6) invalid principles or concepts (e.g. confusing position and energy eigenstates)	- (incorrect assertion)	- (same)
Applying concepts	1) Using delta function as initial state and expanding it as linear superposition of energy eigenstates: $\psi(x) = A\delta(x - a/2) = \sum_n c_n \psi_n$	+	+
	2) Dirac delta function identity $\int f(x)\delta(x - x_0) dx = f(x_0)$	+	-
	3) Probability of measuring energy E_n is $\|c_n\|^2$ which is obtained using $c_n = \int \psi_n * \psi(x) dx$ over all space	+	+
	4) using provided stationary states for infinite square well to interpret probabilities	-	-
Organization	1) diagrams: i) infinite square well and initial state ii) labels axes of diagram of well (V(x) vs. x) and axes of diagram for $\Psi(x)$ vs. x 2) Clear/appropriate knowns	+	++/-
Plan	1) Appropriate target quantity chosen $\|c_n\|^2$ 2) Appropriate intermediate variables chosen 3) Consistent plan	+/-	+/--
Evaluation	1) completes proof 2) makes connection with results	-	-
Overall Scores	Physics (%) Presentation (%)	68 50	23 33

Figure 4.3. Sample scoring of student #1 in midterm and final exams on problem 3.

row but will receive "n/a" for applying it in the apply row because the student cannot be expected to apply a principle that he/she did not invoke.

An overall or cumulative score is tabulated for each of the physics and presentation parts for each question. For the cumulative physics score, the average of the scores for each subcategory (e.g., invoking appropriate physics concepts, invoking inappropriate concepts and applying concepts) can be used in each column for each student on a given problem on the midterm or final exams. Similarly, the cumulative score for the presentation or problem solving part can be computed by averaging over the scores for each of the subcategories (organization, plan, and evaluation) in each column for each student on a given problem on the midterm or final exams.

4.4 RESULTS

Although the grading rubric allows us to assign scores to each student for performance on physics and presentation parts separately, these two scores are highly correlated, as shown in Figure 4.4. We therefore only focus on students' physics scores on each of the four questions given on the midterm and final exams. The physics scores for each student for each of the four problems were analyzed as separate data points for a total of 56 data points. A separate analysis that omitted the "easy" problem (problem 5) was also carried out (to focus only on the three problems that many students experienced difficulty with) for a total of 42 data points.

Overall, the midterm average score of all students was 66% on all four problems and 57% on the three difficult problems (omitting the "easy" problem). The average final exam score of all students was 60% on all four problems and 53% on the three difficult problems. Thus, the students' average final exam performance on these problems is slightly worse than their performance on the midterm exams. Before we focus closely on each student's change in performance from the midterm to final exams on problems given a second time, we note that this lowering of the average score in the final exam compared to the midterm exams suggests that the assumption that the senior-level physics majors will automatically learn from their mistakes may not be valid. As discussed below, some students did well both times or improved in performance but others did poorly both times or went from good to bad performance. Moreover, students struggled the most on problem 4 and regressed (from good to bad) the most on problem 2.

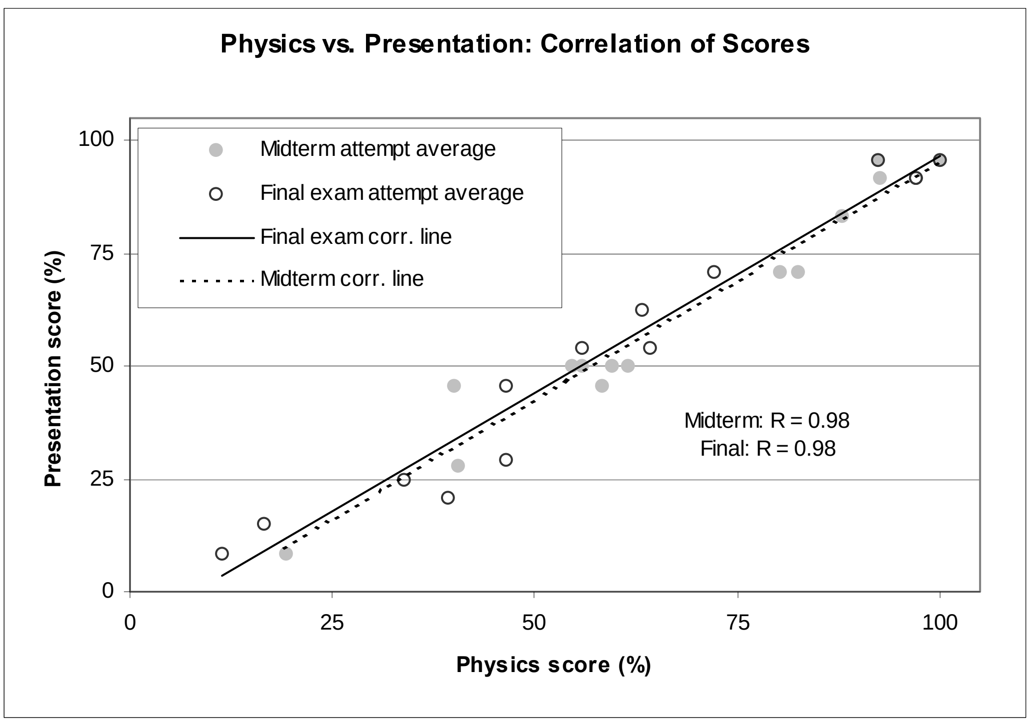


Figure 4.4. Graph showing the correlation between the average scores on the four questions for physics and presentation parts for students.
For both the average midterm and final exam attempts, the correlation is very strong (R = 0.98). The reason for this correlation is that students' presentation of the problem depended upon whether or not they understood the physical content. If a student did not know the relevant physical concepts, he/she could not set up an appropriate problem solving strategy to score well on the presentation part.

Table 4.1 summarizes how students' performance on all problems combined and on each individual problem changed from the midterm to the final exam. "Good" is defined as obtaining 60% or higher using the rubric on the physics score and "bad" is defined as getting lower than 60%. The "Good to good" category means that a student performed better than 60% on both the midterm and the final, "good to bad" means that a student performed higher than 60% on the midterm and below 60% on the final, etc. For three of the four problems, the results are mixed, and are split relatively evenly between students who performed well on both attempts, and students who performed poorly. Problem 5 is the "easy" problem that most students had very little trouble with on either attempt. Taken individually, the change in a student's performance from the midterm to the final exam has a wide distribution as shown in Tables 4.2a and 4.2b.

Table 4.1. Change of student performance from midterm to final exam per problem.
Total percentage and number of instances in which students who performed above 60% on the midterms continued to perform above that threshold (good to good) or regressed (good to bad) and the number and percentage of instances in which students who performed below 60% on midterms continued to perform below that threshold (bad to bad) or improved (bad to good). The number of instances are shown for all problems together (total instances) and also for each of the four problems separately.

	Good to good	Good to bad	Bad to good	Bad to bad
Problem 2	6	2	1	5
Problem 3	5	2	2	5
Problem 4	4	2	2	6
Problem 5	11	2	0	1
Total Instances	26	8	5	17
Total Percentage	46%	14%	9%	30%

Table 4.2a. Change of student performance from midterm to final exam per student (all 4 problems). The number of instances (out of 4 for the four common problems given on both the midterm and final exams) in which each of the 14 students who performed above 60% on the midterms continued to perform above that threshold (good to good) or regressed (good to bad) and the number of instances in which students who performed below 60% on the midterms continued to perform below that threshold (bad to bad) or improved (bad to good). An "I" next to the student ID number denotes that the student was individually interviewed as discussed later.

Student ID	Good to good	Good to bad	Bad to good	Bad to bad
1	2	1	0	1
2	0	2	0	2
3 (I)	3	1	0	0
4	1	1	0	2
5	1	0	0	3
6 (I)	1	2	0	1
7	1	0	1	2
8(I)	3	0	1	0
9(I)	4	0	0	0
10 (I)	1	1	1	1
11 (I)	3	0	1	0
12	4	0	0	0
13	2	0	1	1
14	0	0	0	4

Table 4.2b. Change of student performance from midterm to final exam per student (omitting problem 5). The number of instances (out of 3 for the three problems given on both midterm and final exams omitting problem 5 on which most students performed well) in which each of the 14 students who performed above 60% on midterms continued to perform above that threshold (good to good) or regressed (good to bad) and the number of instances in which students who performed below 60% on midterms continued to perform below that threshold (bad to bad) or improved (bad to good). An "I" next to the student ID number denotes that the student was individually interviewed as discussed later.

Student ID	Good to good	Good to bad	Bad to good	Bad to bad
1	1	1	0	1
2	0	1	0	2
3 (I)	2	1	0	0
4	0	1	0	2
5	0	0	0	3
6 (I)	0	2	0	1
7	0	0	1	2
8 (I)	2	0	1	0
9 (I)	3	0	0	0
10 (I)	1	0	1	1
11 (I)	2	0	1	0
12	3	0	0	0
13	1	0	1	1
14	0	0	0	3

4.4.1 Further analysis

Table 4.3 contains the students' individual scores on all four problems for both midterm and final exam attempts. In addition, an average midterm score (m_i) and a final exam score (f_i) is reported for each student after averaging over all four problems. The midterm and final exam averages for each student are normalized with respect to the overall average of all students for the midterm and the final exams ($\overline{m}$ and $\bar{f}$, respectively) to obtain the normalized averages for each student ($\overline{m}_i = m_i / \overline{m}$ and $\bar{f}_i = f_i / \bar{f}$ respectively). We introduce two parameters, Performance factor and Rote factor for each student, which are helpful in interpreting the students' performance from the midterm to the final exams.

Table 4.3. Individual and average physics performances for each student on each question and each exam.

	Problem 2		Problem 3		Problem 4		Problem 5		Average	
Student ID	Mid	Final	Mid	Final	Mid	Final	Mid	Final	Mid (m_i)	Final (f_i)
1	78	100	68	23	0	0	100	100	62	56
2	50	8	100	48	5	0	83	10	60	17
3 (I)	100	18	88	88	83	100	100	83	93	72
4	29	31	21	4	83	0	100	100	58	34
5	14	28	38	0	10	29	100	100	41	39
6 (I)	100	15	52	23	100	48	100	100	88	47
7	42	58	33	29	0	83	85	83	40	63
8 (I)	100	100	100	88	30	100	100	100	83	97
9 (I)	100	100	87	87	83	83	100	100	93	93
10 (I)	100	63	26	81	3	0	94	42	56	47
11 (I)	33	100	88	88	100	100	100	100	80	97
12	100	100	100	100	100	100	100	100	100	100
13	100	94	18	88	0	4	100	71	54	64
14	13	0	27	13	0	0	37	32	19	11

The Performance factor for each student (P_i) is a measure of the overall performance on both attempts at the four problems. We define mathematically the Performance factor for each student as follows: $P_i = \overline{m}_i \bar{f}_i$. A second factor, the Rote factor for each student (R_i), is a measure of how the students' performance changed from the midterm to the final exam and is mathematically defined as $R_i = \overline{m}_i / \bar{f}_i$. One possible reason for a student's rote factor being greater than unity is that a student superficially learned the limited material that was assigned for each midterm without trying to assimilate the new knowledge with prior knowledge. If a student does not attempt to build a knowledge structure, he/she may do poorly on the final exam when the student is tested on the comprehensive course material despite doing well on the midterm exam. Of course there may be other reasons why a student didn't do as well on the final exam, e.g., there may be other exams to study for during the final exam week, and the student may simply not have studied as hard for the final exam as for the midterm exam. Our assumption in defining and interpreting the Rote factor is that, if the student had integrated the relevant concepts well within his knowledge structure, he/she would not forget how to do the problem by the time the final exam came and would not need to spend too much time on those concepts again.

If the Rote factor R_i for a student is large, the student has performed poorly on the final exam compared to the midterm exam on the four questions taken together. As noted earlier, a Rote factor R_i larger than 1 might indicate a tendency to amass rote knowledge during the midterm exam without a deep integrated knowledge that lasts. The Performance factor P_i and Rote factor R_i are presented in table 4.4 for each student alongside $\overline{m}_i$, $\bar{f}_i$ and the normalized homework average for each student, $\overline{H}_i$, computed in a manner similar to $\overline{m}_i$ and $\bar{f}_i$. Thus $\overline{H}_i$

Table 4.4. The normalized average midterm score, normalized average final exam score, Performance factor, Rote factor and normalized average homework score for each student.

Student	$\overline{m}_i$	$\bar{f}_i$	P_i	R_i	$\overline{H}_i$
1	0.93	0.93	0.87	1.00	0.35
2	0.90	0.28	0.25	3.26	1.10
3	1.40	1.21	1.70	1.16	1.19
4	0.88	0.57	0.50	1.56	1.04
5	0.61	0.66	0.40	0.93	0.89
6	1.33	0.78	1.04	1.71	1.10
7	0.61	1.06	0.64	0.57	1.14
8	1.25	1.62	2.03	0.77	1.21
9	1.40	1.55	2.17	0.90	1.27
10	0.84	0.78	0.66	1.08	1.10
11	1.21	1.62	1.97	0.75	1.06
12	1.51	1.68	2.53	0.90	1.28
13	0.82	1.08	0.89	0.77	1.11
14	0.29	0.19	0.05	1.55	0.16

for each student is computed as the ratio of a student's homework score for the semester and the average homework score of all students. We compute $\overline{H}_i$ to explore any correlation between the homework score, Performance factor and Rote factor for a given student.

Figure 4.5 shows the correlation between the normalized midterm average and normalized final exam average and Figure 4.6 shows the correlation between the Performance factor and Rote factor. While the normalized midterm average and normalized final exam average are positively correlated, there is a negative correlation between the Performance factor and Rote factor. It is important to note however that the Performance factor and Rote factor together should be considered to interpret a student's overall performance in the midterm and final exams.

We also examine the correlations between each of the Performance factor P_i and Rote factor R_i with the normalized homework score $\overline{H}_i$ in order to see if the overall student performance on the homework is related to students' overall performance on the exams and their improvement from midterm to final exam. Figure 4.7 suggests that students' Performance factor is correlated with the homework performance, while Figure 4.8 suggests that students' homework performance is not strongly correlated with the Rote factor. While the differences in the homework scores of different students are not very large, students with a high Performance factor obtain better overall scores on the homework.

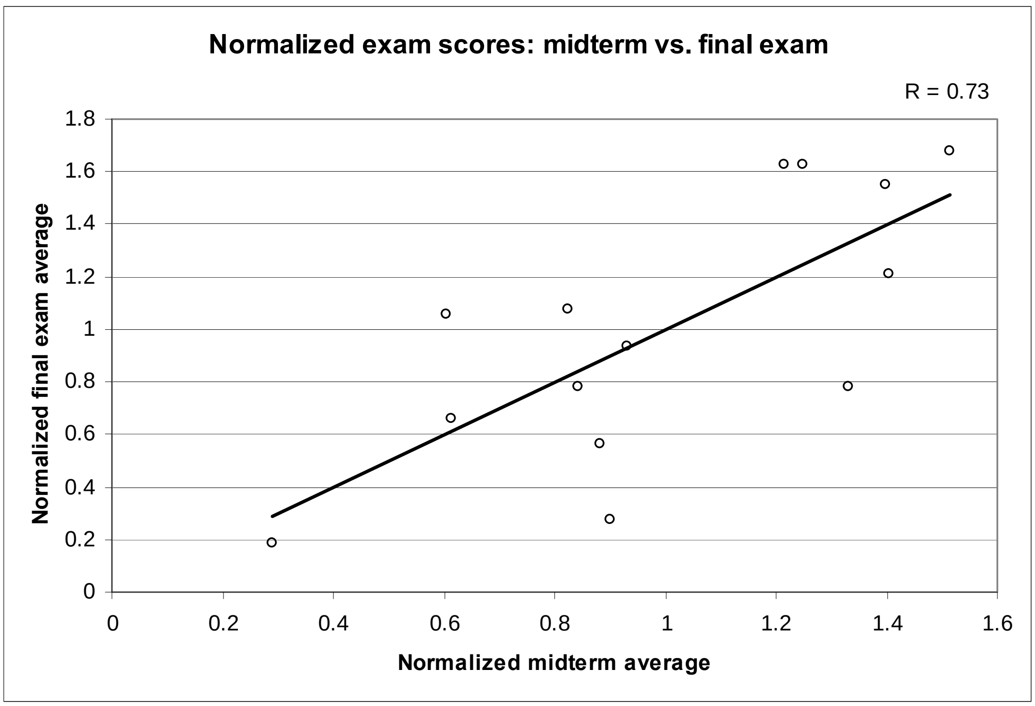


Figure 4.5. Normalized final exam average vs. normalized midterm average shows a correlation indicating students' performance in midterm exams and final exam are moderately consistent.

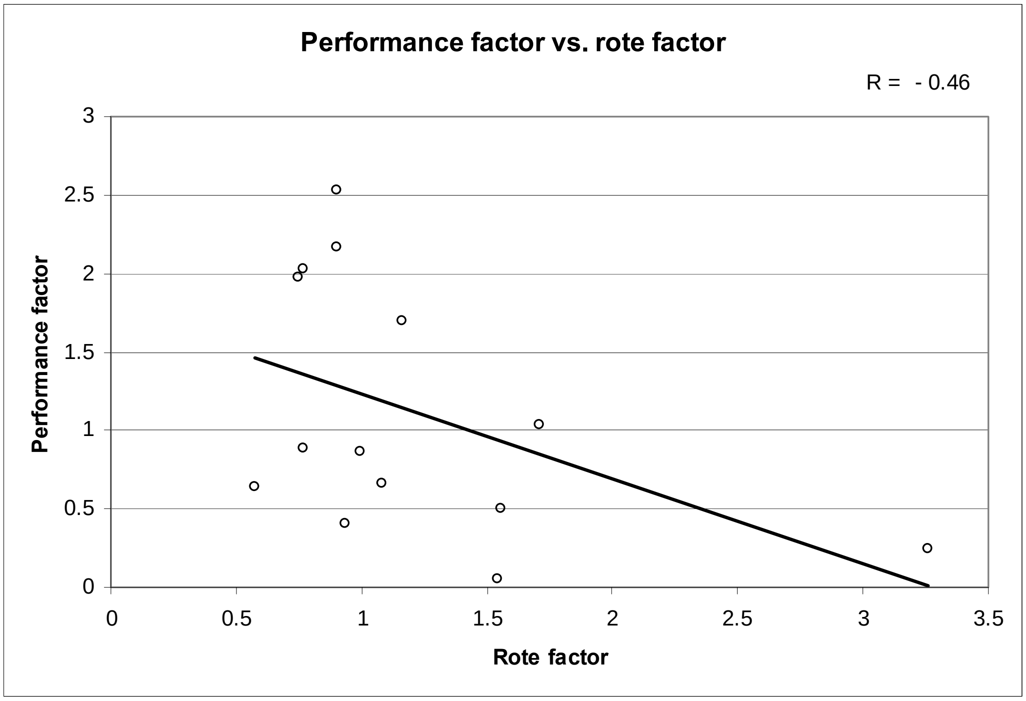


Figure 4.6. Rote factor vs. performance factor shows a negative correlation.

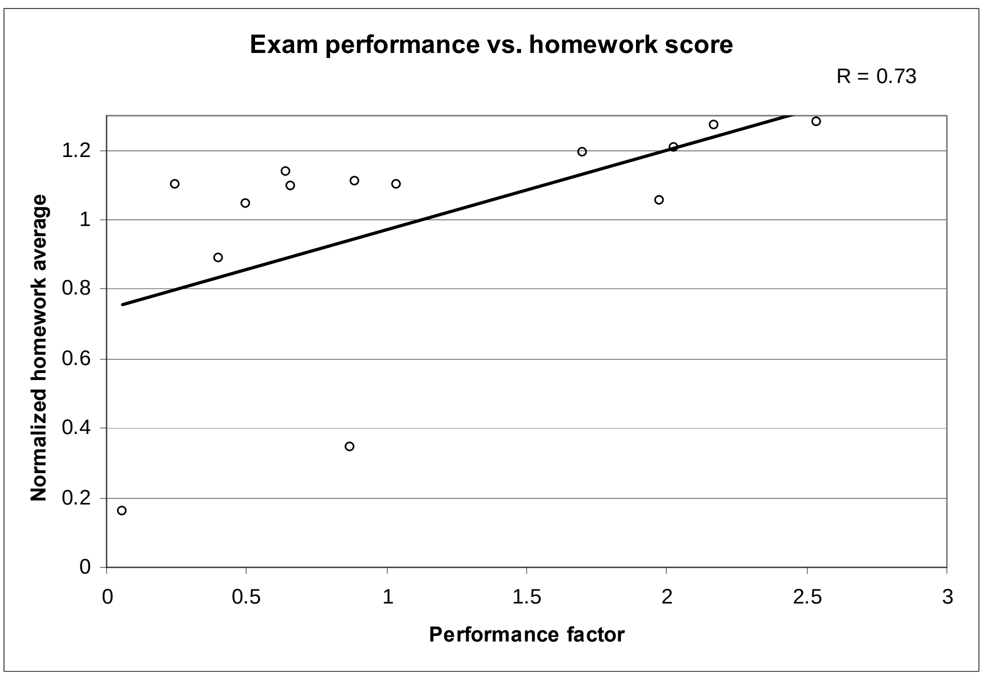


Figure 4.7. Normalized average homework score $\overline{H}_i$ vs. Performance factor P_i (excluding two outliers whose homework scores were very low).

The positive correlation implies that a student who performed better on the homework also has better performance on the exams. While the correlation line reflects all the points in the graph, the two points on the bottom are considered outliers due to the students not attempting most of their homework assignments, and therefore not used to calculate the correlation coefficient in the upper right corner.

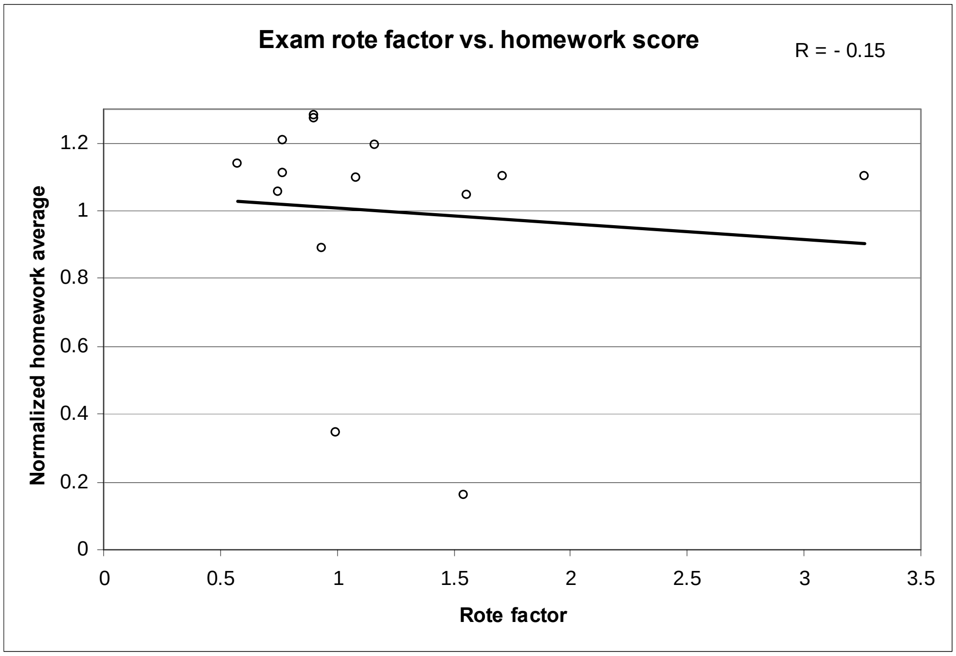


Figure 4.8. Normalized average homework score $\overline{H}_i$ vs. Rote factor R_i (excluding two outliers whose homework scores were very low).
Here, the Rote factor and homework scores have no statistically significant correlation. While the correlation line reflects all the points in the graph, the two points on the bottom are considered outliers and not used to calculate the correlation coefficient in the upper right corner due to the students not attempting most of their homework assignments.

Figures B.2.1-4 in Appendix B.2 show examples of midterm and final exam performances for students who performed poorly both times or solved a problem correctly in the midterm exam but regressed in the final exam. On the other hand, Figure B.2.5 shows an example of a student's performance with major improvement the second time. Figures B.2.1 and B.2.2 show two different students' performances on problem 3. In figure B.2.1, student 4 does not improve from a poor performance in the midterm exam and his performance on both exams is poor. In fact, both times, student 4 appears to write extensively but most of his writing is completely irrelevant to the question asked. It is hard to imagine that the student had no clue that none of the things written by him were relevant to the question asked. It is possible that the student thought that if he wrote anything that he could remember about the topic (whether relevant or not) he may get some points on the exam for trying. In fact, the irrelevant writing of the student was very different both times, e.g., in the midterm, the student attempted to calculate the expectation values of position and Hamiltonian and in the final exam, the student attempted to solve the time-independent Schrödinger equation. The poor performance of the student both times suggests that when the midterm exam was returned to the student and the correct solution was provided, the student did not use his mistake as an opportunity for learning and did not take the time to learn the relevant material even by the final exam time. As noted in Table 4.1, in 30% of all cases, the performance was poor both in the midterm and final exams.

In Figure B.2.2, Student 2's performance on problem 3 declined sharply on the final exam from a good performance on the midterm exam. In the midterm exam, the student correctly calculated the probability of measuring energy immediately after measuring position by expanding the position eigenfunction in terms of the energy eigenfunctions and finding the expansion coefficients. But Figure B.2.2b shows that in the final exam, the student did not have any clue about how to calculate the probability of measuring energy and first appears to be doing a Fourier transform of the Delta function to calculate the corresponding momentum space wave function, which is not relevant for the problem. The rest of the work on the final exam shown in Figure B.2.2b is unclear and somehow the student incorrectly ends up with the incorrect conclusion that the probability of measuring energy is zero for all odd energy states. We note that, if the student had understood either conceptually or systematically how to calculate the probability of measuring a physical observable (e.g., energy) in an experiment given the wave function and integrated this conceptual understanding into his knowledge structure, he would not be completely clueless about it a couple of months after this concept was covered right before the first midterm. This student's performance on the midterm and final exams (Figures B.2.2a and

B.2.2b) suggests that the student may have memorized a procedure for calculating the probability for measuring energy (without taking the time to reflect on how this concept fits with his knowledge structure) right before the midterm exam when the amount of material covered was limited and there was a high probability of this material showing up on the exam. If the student does not take the time to connect the various concepts learned in the course, not only is the student likely to forget over time the memorized "procedure" for calculating the probability of measuring energy which is not integrated as a "concept" in his knowledge structure, but there may also be interference from the new material covered after the midterm exam, e.g., the irrelevant calculation of the momentum space wave function (see Figure B.2.2b), which was taught after the first midterm exam.

In introductory physics courses, students who have not developed a robust knowledge structure struggle to identify, e.g., the principle of physics that may be relevant when the material for the exams spans many chapters and students have to identify the relevant concepts applicable for a particular problem. But when only a limited amount of material is covered in a quiz, they may be able to use a memorized "procedure", e.g., the fact that to find the angular speed of the ballerina when she puts her arms close to her body, one needs to realize that the initial angular momentum of the system is equal to the final angular momentum even without integrating with the knowledge structure the concept of the angular momentum conservation principle and the condition under which the angular momentum of a system is conserved. It appears that a similar lack of knowledge structure may be responsible for Student 2's unintelligible final exam performance on problem 3 although the student did the problem correctly in the midterm exam. The fact that Student 2 invoked more recently covered but irrelevant concepts such as momentum space wave function (or more likely a "procedure" instead of a "concept" in the student's mind) suggests that the student had not taken the time to build a knowledge hierarchy related to quantum mechanics concepts.

Similarly, in Figure B.2.3, Student 3 attempts problem 2, and while he does the problem correctly on the midterm exam, he seems to have no idea how to do it on the final exam. A closer look at this student's final exam performance on problem 2 (see Figure B.2.3b) also shows that, in the final exam, Student 3 invoked the more recently covered but irrelevant procedure for finding the time-dependence of the expectation value of a physical observable. Again, it is clear that Student 3 has not taken the time to reflect upon and learn these concepts and may have memorized and regurgitated a "procedure" correctly in the midterm exam. If the student had built a good knowledge structure, he would not be so off-track on the same problem in the final exam as to invoke the commutation relation between the operator Q and the Hamiltonian for calculating the expectation value of the physical observable Q at a given time (which has nothing to do with what the problem asks). However, the irrelevant procedures that Student 3 incorrectly employed for problem 2 are relevant for determining the time-dependence of the expectation value because the Hamiltonian governs the time evolution of the system according to the time-dependent Schrödinger equation. Due to a lack of good knowledge structure, the student forgot about the procedure he used in the midterm exam and invoked a procedure not related to the concepts relevant for the problem.

Figure B.2.4 shows that Student 6 solves problem 4 correctly in the midterm exam but regresses severely on the final exam. Such regression from the midterm to the final exam may again be due to the student having memorized how to do certain problems in the midterm exam because a very similar problem was part of the homework and was discussed during the homework discussion. But if the student had not integrated this new knowledge with his prior knowledge, the student forgot how to do the problem in the final exam. In fact, Figure A.2.4b shows that, in the final exam, Student 6 performs a Fourier transform to obtain a position space wave function from the momentum space wave function which is not at all relevant for the problem at hand but which was covered later in the course. The student also appears to remember vaguely (after doing the same problem in both homework and midterm exam earlier) that there was a Taylor expansion involved and there was some type of analogy with the Hamiltonian (but he writes it incorrectly as the Hamiltonian being the generator of "energy" in time which does not make any sense as opposed to being the generator of translation in time). When the student does not know how to proceed on Problem 4 in the final exam, he writes that if the momentum operator acting on "x" is not zero then the particle is moving, hence translating in space. However, this type of reasoning does not make any sense.

Finally, Figure B.2.5 shows an example in which Student 7 performed poorly on problem 4 on the midterm exam but improved significantly on the final exam. Figure B.2.5a shows that in the midterm exam, Student 7 wanders aimlessly while solving the problem. He invokes many irrelevant mathematical identities such as the eigenvalue equation for the momentum operator, writes the position space wave function in the Dirac notation, writes the identity operator as a complete set of eigenstates of momentum operator, inserts the identity operator to write the position space wave function as a Fourier transform of the momentum state wave function, etc., but none of these activities got him closer to the goal. However, the student must have reflected on this material later and he performs well on the final exam (see Figure B.2.5b).

4.5 INTERVIEWS

As noted earlier, in many instances, a student's performance on a particular problem became worse on the final exam compared to the midterm exam. This regression from the midterm to final exam is contrary to the popular belief that physics majors who are on the verge of completing the program have already learned to learn and will take the opportunity to learn from their mistakes. One hypothesis is that students who regress on problems from the midterm to the final exam might have memorized how to do certain problems (especially since the material for each of the midterm exams was not extensive) rather than actually learning the concepts, organizing and extending their knowledge structure and making sure that the new knowledge learned via homework and other course material is integrated with their prior knowledge.

Of course, one does not obtain the whole picture of what a student is thinking through examining the final written product of their work. To get a better insight, we asked the subject class for paid volunteers for individual interviews, and four of the original 14 students responded. The goal of these interviews was to learn about students' attitudes and approaches towards problem solving and learning and to track their thought processes as they attempted to solve the 4 problems chosen for the study again during the interview.

The interviews, lasting between 1-1.5 hours each, took place about two months after the end of the quantum mechanics class in which the written tests were given, and with the exception of one student (student 10), all subjects were currently enrolled in the second semester course in quantum mechanics. The timing for interviews was chosen because it ensured that the students could not simply reproduce answers they may have memorized before exams and it allowed us to test long-term learning and knowledge organization of students by determining how well the students could work these problems out a couple of months after the semester was over.

In addition to the four interviews in the following semester as described above, two other students (students 8 and 9) were later available for shorter interviews. These last two interviews focused only on their attitudes and approaches towards problem solving and learning (they were not asked to solve the problems again during the interview like the first four interviewed students due to time constraints).

4.5.1 Interview Procedure

Each of the interviewed students was first asked a series of questions that sought the student's opinion on various aspects of the course, and on the student's performance in the course. Students were also asked questions about their attitudes and approaches to problem solving and learning. In particular, they were asked about their general approach to problem solving and learning in physics, whether they prefer to do homework alone or with peers, whether they take the time to learn from their mistakes on homework and exams as soon as possible, and their performance on each question, e.g. why they struggled with one problem or did well on another. The interviews corroborate the fact that, similar to introductory physics students, many advanced students need explicit guidance and support in exploiting problem solving as an opportunity for learning and in reflecting and learning from their mistakes and in building a robust knowledge structure.

After these questions, each of the first four students interviewed was presented with the four problems of interest to solve one by one. The student first attempted to solve each problem without any input from the researcher using a think-aloud protocol. If the student noted that he did not know how to solve the problem correctly, the researcher first encouraged him to progress as far as he could. Then the researcher provided successive hints to allow the student to go back, repair mistakes and make more progress. This measure was helpful in evaluating a student's knowledge structure further since we probed a student's progress with increasing feedback. Using this process we got an insight into how well a student could self-repair on the spot and solve the problem correctly with the scaffolding provided by the researcher.

4.5.1.1 Students interviewed

Below we summarize the interviews with student volunteers. The four subjects who came forward as volunteers for the interviews were students 3, 6, 10, and 11.

Student 3 did very well on the midterm exam but appeared to have some trouble on the final exam (specifically, he regressed on problem 2). Because of this regression it was of interest to examine this student's problem solving and learning approach and explore how well he retained and could retrieve the knowledge acquired the previous semester.

Student 6 was an interesting subject because he generally did well on the problems of interest on the midterm exams but performed poorly on those same problems on the final exam, particularly with 3 of 4 problems on the final exam.

Student 10 was a fairly weak student whose performance on the four problems on the midterm and final exams was very inconsistent. In particular, he did well on problem 2 both times, improved from a bad performance on problem 3, regressed from a good performance on problem 5, and fared badly on both attempts on problem 4. It should be noted that student 10 had graduated immediately after taking the course and was not taking the second semester of quantum mechanics during the semester in which the interviews were conducted as the other three students were. He was still around conducting research with a faculty member in the department. For this reason, it was also of interest to see how well he had retained and could retrieve while solving problems based on last semester's material in comparison to the three students who were in the second semester quantum mechanics course (even though the material for the second semester course is mostly unrelated to the first semester material).

Student 11 did well except for one poor problem performance on the midterm (Problem 2); and he improved on the final exam. First, we wanted to understand how he learned from his mistakes, but we also wanted to explore how well he had retained the knowledge beyond the final exam a couple of months later when he was enrolled in the second semester of quantum mechanics.

The shorter interviews conducted later about attitudes and approaches to problem solving involved students 8 and 9 who had performed well in the final exam.

4.5.2 Interview: Survey and Attitudes/Approaches to Problem Solving

Student 3, though he got an A- in the course, felt he could have done better, and blamed his not being able to do so on being overextended during the semester he took the quantum course. Student 3's homework habits seemed to focus on working alone except when he found a problem tricky, in which case he would consult the professor or occasionally a classmate who understood the problem better. The reason he worked alone was one of convenience; he found it too much trouble to get a regular study group together. His study habits for exams focused mostly on doing practice problems and reading the book for concepts, saying that this was about all that could be done to study because problems were the main focus of study. He explicitly noted that he looked at the solutions to homework and exams provided right before the exam.

Student 6 stated that his study habits were generally consistent. He cited working with a partner (student 10) on the homework, specifically comparing notes with each other after each person worked individually on the homework, and that he made sure to speak with the instructor for any mistakes made on the homework or exams. Furthermore, he said that he created a study guide and studied equations, as well as reviewed with a partner (also student 10), for one week before each exam. His opinion of the class was very high – the quantum mechanics classes were in fact said to be his favorite of the entire physics curriculum – and he felt that the class, while difficult, could be completed successfully with effort.

Student 10 stated that, of his fall semester classes, his performance in the quantum mechanics class was his weakest performance. He found the course more difficult than usual because he considered the material unique with regard to the other senior-level course material.

Student 10's study habits for the midterms seemed fairly structured: first he would read the text material relevant to the exam; next, he would create a study sheet by writing down key sample problems and formulas (he did not study this per se, but the act of copying down key sample problems and formulas helped him to study); finally, he would re-do as many homework problems as possible. However, his style for reflecting on his mistakes on homework and exams was more suspect, as a bad performance would make him prefer to put away the assignment without looking it over. He also stated that if a problem occurred on an exam that he had seen before in the homework, he would probably do better the second time around. As for a problem that first appeared on a midterm exam and then on the final exam, he struggled because he did not necessarily expect a problem to be repeated on the final exam and did not practice them well. In fact, even Student 8 has something similar to say:

- "If I make mistakes in the homework, I look at the TA's solutions carefully because I know those problems can show up in the exams. But if I make a mistake in the midterm exam, I won't be so concerned about what I did wrong because I don't expect those questions to show up in the final exam. Also, if I don't do well on the exam, I don't feel like finding out what I did wrong because reading my mistake again would just hurt me again, and I don't want anything to ruin the after-exam happy time."

Student 9 claimed he would look back to see what mistakes he made, but he would be more careful doing this for mistakes in the homework than in the exams:

- "When I make mistakes I always look back at the work to see where I erred. In most cases I will be more careful in looking over homework than past exams as far as studying purposes go."

Student 11 seemed to place much emphasis on the available sources that he had with regard to his study habits. He especially credited the course textbook (Griffiths 2005) for providing examples and conceptual questions, as well as homework problems assigned that helped him learn the course material. In addition, he also credited homework help sessions given in the class during which any gaps in the textbook's treatment of a problem would be filled in. He worked alone on homework during the course, only interacting with other students if he got stuck on a problem, and put it off until the last moment, but would check his mistakes as a "matter of principle" (for as much effort as he put into the homework, if he got a problem wrong anyway he wanted to know why). He would also consult the internet to find hints to problems similar to the homework. It should also be noted that this student was also a mathematics major in addition to being a physics major, as he stated several times that he felt a math methods course helped him on the material.

4.5.3 Interview: Problem Performance

In general, the first four interviewed students who were also asked to solve the problems again during the interview had mixed performance on the four problems. All four students struggled to some extent on problem 4, ranging from forgetting or misunderstanding details about the Taylor expansion to having no idea how to even start the problem. Afterwards, all four students complained about problem 4, stating that it seemed to be a stand-alone topic that wasn't particularly connected to the rest of the course material and so they did not study it as thoroughly for exams because they didn't think they would be tested on it. For example, student 3 claimed that the required proof didn't seem very physical:

- "…you know, sometimes, really mathematical problems… I mean, this isn't terribly mathematical but sometimes problems like this just seem like 'oh, this is just a math thing.' You know [the book] or even the professor, maybe prior to the problem doesn't tell you the importance of the problem, like what it really demonstrates. I remember [the professor] did [afterwards]….But [the book] doesn't say anything, it just says, 'Do it.' I remember thinking, at the time I thought it was just something to make me stay up another hour and a half. And then, you know, it was on the test and I thought, 'I should've paid more attention to it,' and then it was on the final and I thought, 'jeez, I really should've paid attention to it.'"

In a similar vein, Student 6 claimed that problem 4 did not stand out as something important from the examination point of view. After writing an expression for the momentum operator he added:

- "…yeah, I totally forget... I haven't touched translator of space since the final… I remember the translator in time and that goes the same way. I remember you got to use some dummy function, some function 'f' that you can throw in, and that you can sum it…. I remember there was an exponential… it's just not ringing a bell right now."

Student 10 also noted that he did not think that problem 4 was important:

- "It was just one of the problems in the homework. It was never mentioned previously or after, so I didn't assign much importance to it in my head as far as studying goes to it."

This type of selective studying based upon the presumed probability of the material showing up on the exam suggests that even advanced upper-level students attempt to "psych out" the exams rather than trying to integrate new knowledge with their prior knowledge to build a robust knowledge structure.

Three students (3, 6, and 10) had difficulties with problem 2. For problem 2, they struggled to remember how to use Dirac notation, stating that in general they had a difficult time mastering it in the course and the second semester quantum mechanics course hardly ever used this notation. For example, about problem 2, Student 3 noted

- "So the real problem that I have is, um… I never quite mastered the whole bra and ket notation. This is what really is hanging me up here, probably if I had, I wouldn't… be hung up, because I remember it being like 4 lines to do this. I remember it not being a complicated thing."

Incidentally, Figure 11 shows that this student (Student 3) had done problem 2 involving the Dirac notation correctly on the midterm exam but did not know how to do it on the final exam. The same student continued

- "I remember doing it, I remember seeing it, I remember thinking, 'oh, it's not really that bad,' so I know it's not bad. It's just…I'm not remembering how to manipulate with bra and ket. Basically that's all it is."

Then, he added

- "…yeah, I mean that's really what the killer thing is…yeah, I mean it's just the bra and ket notation really."

Student 3 appears to have kept some procedures involving the Dirac notation in his short term memory which he used to perform well on problem 2 of the midterm exam. However, he did not remember how to do the problem either during the final exam or during the interview because the knowledge about the Dirac notation was not properly integrated into his knowledge structure. In general, it may be easier to discern if students have a robust knowledge structure when the material they are evaluated on is not very limited because limited material may make invoking of relevant concepts easy by default and all that is required to perform well is the knowledge of how to apply the procedure to solve the problem even if a conceptual understanding is lacking.

After struggling with problem 2, student 10 stated that he felt that the transition from integral notation (position representation) to Dirac notation was not clear to him. It is possible that Dirac notation could be construed as unintuitive to students learning it for the first time. Alternately, the regression on the final exam from the midterm exam by students 3 and 6 suggests they may have crammed on Dirac notation to "get by" on the midterm exams even when they did not understand it well.

For problem 3, students 3, 6, and 10 had a difficult time figuring out what they were supposed to do to solve it. Student 3 was unclear on how to get started and required a hint from the researcher to recall how to do it correctly. Student 6 displayed common errors such as confusing energy eigenstates with position eigenstates, and said that he failed to learn from mistakes on his midterm exam attempt because his solution procedure, while incorrect, gave an answer that seemed correct, i.e., that all even-symmetry energy eigenstates (i.e. states that preserve symmetry through the x-axis) are equally probable and non-zero (on the contrary, the probability of collapsing into an even-symmetry energy eigenstate is zero if the initial state is a delta function in the middle of the well because there is no overlap between the initial delta function wave function and the even-symmetry energy eigenstate wave functions which all have a value of zero in the middle of the well). Student 10 tried to recall some relevant concepts but did not remember the necessary tools to solve the problem. For example, while attempting problem 3, Student 10 noted:

- "…just from my memory, like, there's just too many holes and stuff because I haven't looked at it or thought about it in a while…"

Student 10 also had similar issues with the other problems. For example, after struggling with problem 5, Student 10 noted:

- "I feel like I might just be taking it in a different, in the wrong direction from that point, but… Basically, yeah. Like where it wouldn't take much to get me to remember how to do this again completely."

On the other hand, Student 11 generally displayed excellent physical and mathematical understanding of the problems even after a few months. He completed three of the four problems (2, 3, and 5) without getting stuck on anything and only needed a small hint in order to complete problem 4. While answering problem 2, Student 11 noted:

- "Well, we learned that in particular. It was proven to us in like, three different ways. I remember the page in Griffiths now…"

He cited that the reason for his success was that the course material and study resources were very good and that he was able to learn from them and develop a good understanding of the quantum mechanics formalism. He explicitly said that he used his mistakes in problem solving for reflecting about the holes in his understanding and for overcoming those difficulties. In addition, he majored in mathematics as well as in physics and said that the math courses, e.g., linear algebra, he took recently helped him very much in understanding the concepts in quantum mechanics.

4.6 SUMMARY AND CONCLUSIONS

Prior research on the problem solving and self-monitoring skills of introductory physics students demonstrates that the introductory students do not learn these skills automatically, e.g., by listening passively to lectures and having access to solved examples (Mason et al. 2008, Cohen et al. 2008, Yerushalmi et al. 2008, Yerushalmi et al. 2007, Singh et al. 2007). Moreover, many introductory physics students are "captive audiences" – they may not buy into the goals of the course and their main goal becomes getting a good grade even if their learning is superficial. Research shows that these introductory physics students can benefit from explicit guidance and feedback in developing problem solving and learning skills and alignment of course goals with assessment methods (Eylon and Reif 1984, Van Heuvelen 1991b, Elby 2001, Meltzer 2005, Ozimek et al. 2004, Mason et al. 2008, Cohen et al. 2008, Yerushalmi et al. 2008, Yerushalmi et al. 2007, Singh et al. 2007, Karelina and Etkina 2007, Etkina et al. 2008).

One common assumption that physics instructors often make is that the learning skills of students in advanced physics courses is significantly superior to the learning skills of students in the introductory courses. It is often assumed that advanced physics majors are in our upper-level physics classes because they are eager to learn. They will make every effort to repair, organize and extend their knowledge because they are intrinsically motivated to learn and are not grade driven. For example, instructors often believe that physics seniors in an honors-level quantum mechanics course can monitor their own learning and they will automatically take the time out to learn from their mistakes.

Contrary to this belief, we find that advanced students in the honors-level quantum mechanics sequence did not automatically improve their performance on identical questions given in a midterm and the final exam. The students were provided the correct solutions and their own graded exams. Even then, there was an apparent lack of reflective practice by supposedly mature students and many students did not take the opportunity to repair and organize their knowledge structure. This needs to be considered in that those advanced students who continue to graduate school may not have adopted the habit by that point.

In individual interviews, we asked questions about students' attitudes and approaches towards problem solving and learning, and also asked them to solve the same problems again. The statistical results were consistent with students' "self-described" approaches towards problem-solving and learning. In the interviews we also find evidence that even in these advanced courses there are students who do not use their mistakes as an opportunity for learning and for building a robust knowledge structure; they resort to rote learning strategies for getting through the course. One interviewed student alluded to the fact that he always looked at the correct homework solutions provided but did not always look up the correct midterm exam solutions partly because he did not expect those questions to be repeated on the final exam. This tendency to learn the problems that may show up on the exam without making an effort to build a knowledge structure is typically not expected of physics seniors. It would be useful to investigate if students would take the time out to learn from their mistakes on the midterm exams if they were told explicitly that some of the midterm examination questions may be repeated on the final exam.

Individual discussions with some physics faculty suggests that sometimes their incorrect inferences about advanced physics students' learning and self-monitoring skills is based on the fact that they feel that all physics majors are like them. They may not appreciate the large diversity in the population of physics majors today and may not realize that those who become college physics faculty consist of a very select group of physics majors. While longitudinal research is needed to investigate the differences between those advanced students who become physics faculty and those who do not, it is possible that those students aspiring to be physics faculty make more effort to learn from their own mistakes.

Similar to introductory students, advanced physics students will benefit from explicit scaffolding support and guidance to help them become independent learners. Students will automatically use problem solving as an opportunity for reflecting and learning if they are intrinsically motivated to learn the content and to extend and organize their knowledge. However, students who are not intrinsically motivated may need extrinsic motivation, e.g., explicit reward for developing higher order thinking and self-monitoring skills. Instructional strategies that aim to achieve these goals must ensure that the instructional design and method of assessment are aligned with these goals in order for the students to take them seriously.

There are a number of strategies based on formative assessment that can provide students explicit guidance and extrinsic motivation to learn. These deliberate instructional strategies not only show students where they need to improve, but also provide them with a path or opportunity to improve. For example, Etkina et al. (Karelina and Etkina 2007, Etkina et al. 2008) have documented that the introductory physics students can be taught the process of science and they can learn to think like a physicist with explicit intervention which provides them with scaffolding support and which lasts at least for six weeks. Moreover, as discussed earlier, similar to introductory courses, a reward system (e.g., grade incentive) is critical to help students learn to self-monitor their work. One strategy is explicitly asking students to fix their mistakes by circling what they did incorrectly in homework assignments, quizzes and exams, and explaining why it is incorrect and how it can be done correctly. Such strategies have been used for introductory courses with different levels of scaffolding provided to students for diagnosing their mistakes (Mason et al. 2008, Cohen et al. 2008, Yerushalmi et al. 2008, Yerushalmi et al. 2007, Singh et al. 2007). Asking students to develop "concept maps" (Eylon and Reif 1984) after each unit and providing feedback as they learn to connect different concepts can be a useful strategy for helping them develop robust knowledge structure which will reduce the probability of forgetting concepts. Explicitly asking students to explain in words why a certain principle or concept is relevant to solving a problem and coupling conceptual and quantitative problem solving may be effective means to force students to reflect upon what they are doing and to help them build a more robust knowledge structure (Eylon and Reif 1984, Van Heuvelen 1991b, Elby 2001, Meltzer 2005, Ozimek et al. 2004, Mason et al. 2008, Cohen et al. 2008, Yerushalmi et al.

2008, Yerushalmi et al. 2007, Singh et al. 2007). In exploiting each of these strategies to help advanced students learn to learn, assessment should be commensurate with the goals.

4.7 ACKNOWLEDGMENTS

We are grateful to F. Reif, R. Glaser, E. Yerushalmi, E. Cohen, J. Levy and R. P. Devaty for useful discussions.

4.8 REFERENCES

Bao, L., & Redish, E. (2002). "Understanding Probabilistic Interpretations of Physical Systems: A Prerequisite to Learning Quantum Physics." *Am. J. Phys.* 70(3), 210-217.

Chi, M. (1994). In *Thinking Aloud Chapter 1,* Eds. Van Someren, Barnard, and Sandberg, 1, Erlbaum.

Chi, M. (1997). "Quantifying qualitative analysis of verbal data: A practical guide." *J. Learning Sci.* 6(3), 271.

Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). "Self-explanations: How students study and use examples in learning to solve problems." *Cognitive Science* 13, 145-182.

Chi, M., Feltovich, P., and Glaser, R. (1981). "Categorization and representation of physics knowledge by experts and novices." *Cognitive Science* 5, 121-152.

Cohen, E., Mason, A., Singh, C., & Yerushalmi, E. (2008). "Identifying differences in diagnostic skills between physics students: Students' self-diagnostic performance given alternative scaffolding." In *Proceedings of the Phys. Ed. Res. Conference, Edmonton, AB, CA,* AIP Conf. Proc. 1064, 99-102.

Cummings, K., Lockwood, S., & Marx, J. (2004). "Attitudes Toward Problem Solving as Predictors of Student Success." In *2003 Physics Education Research Conference Proceedings, edited by J. Marx, S. Franklin, and K. Cummings,* AIP Conf. Proc. 720, 133-136.

Dufresne, R., Mestre, J., Thaden-Koch, T., Gerace, W., & Leonard, W. (2005). "Knowledge representation and coordination in the transfer process." In *Transfer of learning from a modern multidisciplinary perspective,* J. P. Mestre (Ed.), 393 pages, Greenwich, CT: Information Age Publishing, 155-215.

Elby, E. (2001). "Helping physics students learn how to learn." *Am. J. Phys.* 69, S54-S64.

Ericsson, K., & Simon, H. (1993). *Protocol Analysis: Verbal reports as data.* MIT Press, revised, 1993.

Etkina, E., Karelina, A., & Ruibal-Villasenor, M. (2008). "How long does it take? A study of student acquisition of scientific abilities." *Phys. Rev. St. Phys. Ed. Res.* 020108.

Eylon, B., & Reif, F. (1984). "Effects of knowledge organization on task performance." *Cognition and Instruction* 1(1), 5-44.

Francis, G., Adams, J., and Noonan, E. (1998). "Do They Stay Fixed?" *The Physics Teacher* 36, 488.

Griffiths, D. (2005). *Introduction to Quantum Mechanics, Second Edition.* Pearson Prentice Hall, Upper Saddle River, NJ.

Hardiman, P., Dufresne, R., & Mestre, J. (1989). "The relation between problem categorization and problem solving among novices and experts." *Memory and Cognition* 17, 627-638.

Heller, J., and Reif, F. (1984). "Prescribing effective human problem-solving processes: Problem description in physics." *Cognition and Instruction* 1, 177-216.

Karelina, A., & Etkina, E. (2007). "Acting like a physicist: Student approach study to experimental design." *Phys. Rev. ST. Phys. Ed. Res.* 3(2), 020106.

Larkin, J. (1981). "Cognition of learning physics." *Am. J. Phys.* 49 (6), 534-541.

Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980) "Expert and novice performance in solving physics problems." *Science* 208, 1335-1362.

Larkin, J., & Reif, F. (1979). "Understanding and teaching problem solving in physics." *Eur. J. Sci. Ed.* 1(2), 191-203.

Maloney, D. (1994). "Research in Problem Solving: Physics." In *Handbook of Research on the Teaching and Learning of Science,* D. Gabel (Ed.), MacMillan.

Marx, J., & Cummings, K. (2007). "What factors really influence shifts in students' attitudes and expectations in an introductory physics course?" In *2006 Physics Education Research*

Conference Proceedings, Syracuse, N.Y., edited by L. McCullough, L. Hsu, and P. Heron, AIP Conf. Proc. 883, 101-104.

Mason, A., Cohen, E., Yerushalmi, E., & Singh, C. (2008). "Identifying differences in diagnostic skills between physics students: Developing a rubric." *Proceedings of the Phys. Ed. Res. Conference, Edmonton, AB, CA,* AIP Conf. Proc. 1064, 147-150.

Meltzer, D. (2005). "Relation between students' problem-solving performance and representational format." *American Journal of Physics* 73, 463-478.

Ozimek, D., Engelhardt, P., Bennett, A., & Rebello, N. (2004). "Retention and transfer from trigonometry to physics." *AIP, Conf. Proc.* 173-176.

Reif, F. (1981). "Teaching problem solving – A scientific approach." *The Phys. Teach.* 33, 310-316.

Reif, F. (1986). "Scientific approaches to science education." *Phys. Today* 39, 48-54.

Reif, F. (1995). "Millikan Lecture 1994: Understanding and teaching important scientific thought processes." *Am. J. Phys.* 63, 17-32 (1995).

Singh, C. (2001). "Student understanding of quantum mechanics." *Am. J. Phys.* 69(8), 885-896.

Singh, C. (2002). "When physical intuition fails." *Am. J. Phys* 70(11), 1103-1109.

Singh, C. (2008). "Student understanding of quantum mechanics at the beginning of graduate instruction." *Am. J. Phys.* 76(3), 277-287.

Singh, C., Belloni, M., & Christian, W. (2006). "Improving student's understanding of quantum mechanics." *Physics Today* 8, 43-49, August 2006.

Singh, C., Yerushalmi, E., & Eylon, B. (2007). "Physics learning in the context of scaffolded diagnostic tasks (II): Preliminary results." *Proceedings of the Phys. Ed. Res. Conference, Greensboro, NC,* AIP Conf. Proc. 951, 31-34.

Van Heuvelen, A. (1991a). "Learning to think like a physicist: A review of research based instructional strategies." *Am. J. Phys.* 59, 891.

Van Heuvelen, A. (1991b). "Overview Case Study Physics." *Am. J. Phys.* 59, 898-907.

Wittmann, M., Steinberg, R., & Redish, E. (2002). "Investigating student understanding of quantum physics: spontaneous models of conductivity." *Am. J. Phys.*70(3), 218-226.

Yerushalmi, E., Mason, A., Cohen, E., & Singh, C. (2008). "Effect of self-diagnosis on subsequent problem solving performance." *Proceedings of the Phys. Ed. Res. Conference, Edmonton, CA,* AIP Conf. Proc. 1064, 53-56.

Yerushalmi, E., Singh, C., & Eylon, B. (2007). "Physics Learning in the context of scaffolded diagnostic tasks (I): The experimental setup." *Proceedings of the Phys. Ed. Res. Conference, Greensboro, NC,* AIP Conf. Proc. 951, 27-30.

5.0 PEER REFLECTION IN INTRODUCTORY PHYSICS

5.1 ABSTRACT

We describe a study in which algebra-based introductory physics students in the Peer Reflection or PR group were provided guidance and support to reflect upon problem solving with peers in the recitation class. The non-PR recitation class was run in a traditional manner with the TA answering students' homework questions and then giving a quiz at the end of each recitation class. Our assessment method was novel in that it involved counting the number of problems in which students drew diagrams or did scratch work on scratch books when there was no partial credit for these activities because the questions were in the multiple-choice format. We find that the PR group drew more diagrams than the traditional group even when there was no external reward for drawing them. Also, the improvement in the student performance from the average of the midterm performances to the final exam is larger for the PR group than for the traditional group. The peer-reflection process which was sustained throughout the semester requires students to evaluate their solutions and those of their peers, a task which involves a high level of mental processing (Marzano et al. 1988). The reflection process with peers can also help students monitor their learning. This method can also be beneficial in helping them develop the ability to communicate their approaches because students must discuss with each other their rationale for perceiving one student's solution as superior in terms of its problem solving

strategies than another solution. We also find that diagrams helped with problem solving and students who drew diagrams for more problems performed better than others regardless of whether they belonged to the traditional group or the PR group.

5.2 INTRODUCTION

Students must learn effective problem solving strategies in order to develop expertise in physics. Specifically, they must be able to solve problems beyond those that can be solved using a plug-and-chug approach (Reif 1986, Reif 1995, Reif 1981, Larkin and Reif 1979, Larkin 1981, Larkin et al. 1980, Maloney 1994). Research shows that converting a problem from the initial verbal representation to other suitable representations such as diagrammatic, tabular, graphical or algebraic can make further analysis of the problem easier (Larkin and Simon 1989, Larkin 1985, Kaput 1985). Similarly, using analogies or considering limiting cases are also useful strategies for solving problems (Gick and Holyoak 1987, Holyoak and Thagard 1995, Singh 2002). Many traditional courses do not explicitly teach students effective problem solving heuristics. Rather, they may implicitly reward inferior problem solving strategies in which many students engage. Instructors may implicitly assume that students appreciate the importance of initial qualitative analysis, planning, evaluation, and reflection phases of problem solving and that these phases are as important as the implementation phase (Schoenfeld 1985, Schoenfeld 1992, Schoenfeld 1989, Schoenfeld and Herrmann 1982). Consequently, they may not explicitly discuss and model these strategies while solving problems in class. Recitation is usually taught by the teaching assistants (TAs) who present homework solutions on the blackboard while students copy them in their notebooks. Without guidance, most textbook problems do not help students monitor their learning, reflect upon the problem solving process, and pay attention to their knowledge structure.

Both quantitative and conceptual problem solving both can enhance problem solving and reasoning skills, but only if students engage in effective problem solving strategies rather than treating the task purely as a mathematical chore or guess-work (Chi et al. 1981, Dufresne et al. 2005, Hardiman et al. 1989). Without guidance, many introductory physics students do not perceive problem solving as an opportunity for learning to interpret the concepts involved and to draw meaningful inferences from them. Instead, they solve problems using superficial clues and cues and apply concepts at random without concern for their applicability. With explicit training these same problem solving tasks can be turned into learning experiences that help students organize new knowledge coherently and hierarchically. The abstract nature of the laws of physics and the chain of reasoning required to draw meaningful inferences make it even more important to teach students effective problem solving strategies explicitly.

Reflection is an integral component of effective problem solving (Black and Wiliam 1998a, Black and Wiliam 1998b). While experts in a particular field reflect and exploit problem solving as an opportunity for organizing and extending their knowledge, students often need feedback and support to learn how to use problem solving as an opportunity for learning. There are diverse strategies that can be employed to help students reflect upon problem solving. One approach that has been found to be useful is “self-explanation" or explaining what one is learning explicitly to oneself. Chi et al. (1989) found that, while reading science texts, students who constantly explained to themselves what they were reading and made an effort to connect the material read to their prior knowledge performed better on related tasks given to them after the reading. Inspired by the usefulness of self-explanation, Yerushalmi et al. (2007) investigated how students may benefit from being explicitly asked to diagnose mistakes in their own quizzes with different levels of scaffolding support. They found that students benefited from diagnosing their own mistakes. The level of scaffolding needed to identify the mistakes and correct them depended on the difficulty of the problems.

Another activity that may help students learn effective problem solving strategies while simultaneously learning physics content is reflection with peers. In this approach, students reflect not only on their own solution to problems, but reflect upon their peers' solutions as well. Integration of peer interaction with traditional lectures has been popularized in the physics community by Mazur from Harvard University (Mazur 1997a, Mazur 1997b, Mazur 2009, Crouch and Mazur 2001). In Mazur's approach, the instructor poses conceptual problems in the form of multiple-choice questions to students during the lectures periodically. These multiple choice questions give students an opportunity to think about the physics concepts and principles covered in the lecture and discuss their answers and reasoning with peers. The instructor polls the class after peer interaction to obtain the fraction of students with the correct answer. On one hand, students learn about the level of understanding that is desired by the instructor by discussing with each other the concrete questions posed. The feedback obtained by the instructor is also invaluable because the instructor learns about the fraction of the class that has understood the concepts at the desired level. This peer-instruction strategy keeps students alert during lectures and helps them monitor their learning, because students not only have to answer the questions, but also must explain their answers to their peers. The method keeps students actively engaged in the learning process and lets them take advantage of each others' strengths. It helps both the low- and high-performing students because explaining and discussing concepts with peers helps students organize and solidify concepts in their minds.

Heller et al. have shown that group problem solving is especially valuable both for learning physics and for developing effective problem solving strategies (Heller et al.1992, Heller and Hollabaugh 1992). They have developed many “context-rich" problems that are close to everyday situations and more challenging and stimulating than the standard textbook problems. These problems require careful thought and the use of many problem representations. Working with peers in heterogeneous groups that include students with high, low and medium performance thus far is particularly beneficial for learning from the “context-rich" problems and students are typically assigned rotating roles of manager, time keeper and skeptic by the instructor.

Prior research has shown that, even with minimal guidance from the instructors, students can benefit from peer interaction (Singh 2005). In this study, those who worked with peers not only outperformed an equivalent group of students who worked alone on the same task, but collaboration with a peer led to “co-construction" of knowledge in 29% of the cases (Singh 2005). Co-construction of knowledge occurs when neither student who engaged in the peer collaboration was able to answer the questions before the collaboration, but after working with a peer both were able to answer them on a post-test given individually to each person.

5.3 METHODOLOGY

The investigation involved an introductory algebra-based physics course mostly taken by students interested in health related professions. The course had 200 students and was broken into two sections both of which met on Tuesdays and Thursdays and were taught by the same professor who had taught both sections of the course before. A class poll at the beginning of the course indicated that more than 80% of the students had taken at least one physics course in high school, and perhaps more surprisingly, more than 90% of the students had taken at least one calculus course (although the college physics course in which they were enrolled was an algebra-based course).

The daytime section (107 students) was the traditional recitation group whereas the evening section (93 students) was called the "Peer Reflection" (PR) group. While the lectures and all of the out of class assignments for the daytime and evening sections were identical to the best of instructor's ability, the recitations for the traditional group and PR group were structured differently.

All three of the traditional group recitations were taught in the traditional way with the TA solving selected assigned homework problems on the blackboard and fielding questions from students about their homework before assigning a quiz in the last 20 minutes of the recitation class. The quiz for the traditional group typically had three problems per week based upon the homework that was assigned for that week. For fairness, the cumulative scores in the course for the PR group and the traditional group were curved separately for purposes of determining the final course grades for the students. Each week, students were supposed to turn in answers to the assigned homework problems (based upon the material covered in the previous week) using an online homework system for some course credit. In addition, at the end of the recitation class, students were supposed to submit to the TA for some course credit a paper copy of the homework problems which had the details of the problem solving approach. While the online homework solution was graded for correctness, the TA only graded the paper copies of the submitted homework for completeness on a three point scale (full score, half score or zero).

In the two recitation sections which formed the PR group, the intervention was based upon a field-tested cognitive apprenticeship model of learning involving modeling, coaching, and fading to help students learn effective problem solving heuristics (Collins et al. 1989, Rogoff 1990). In this approach, "modeling" means that the TA demonstrated and exemplified the effective problem solving skills that the students should learn. "Coaching" meant providing students opportunity to practice problem solving skills with appropriate guidance so that they could learn the desired skills. "Fading" meant decreasing the support and feedback gradually with a focus on helping students develop self-reliance. The specific strategy used by the students in the PR group involved reflection upon problem solving with their peers in the recitations, while the TA and the undergraduate teaching assistants (UTAs) exemplified the effective problem solving heuristics. The UTAs were chosen from those undergraduate students who had earned an A+ grade in an equivalent introductory physics course previously. The UTAs had to attend all the lectures in the semester in which they were UTAs for a course and they communicated with the TA each week (and periodically with the course instructor) to determine the plan for the recitations. We note that, for effective implementation of the PR method, two UTAs were present in each recitation class along with the TA. These UTAs helped the TA in demonstrating and helping students to learn effective problem solving heuristics.

Although no pretest was given to students, there is some evidence that, over the years, the evening section of the course tends to be somewhat weaker and does not perform as well overall as the daytime section of the course. For example, the same professor had also taught both sections of the course one year before the peer reflection activities were introduced in evening recitations and thus all recitations for both sections of the course were taught traditionally that year. The midterm scores for the daytime (124 students) and evening (100 students) sections that year were 72% and 66%, respectively, and the final exam scores were 56% and 53%, respectively. The same final exam was used for both sections in this year and also in the year in which peer reflection was introduced to the evening studies. The difference between the daytime and evening sections of the course could partly be due to the fact that some students in the evening section work full-time and take classes simultaneously.

In our intervention, each of the three recitation sections in the traditional group had about 35-37 students. The two recitations for the PR group had more than 40 students each (since the PR group was the evening section of the course, it was logistically not possible to break this group into three recitations). At the beginning of each PR recitation, students were asked to form nine teams of three to six students chosen at random by the TA (these teams were generated by a computer program each week). The TA projected the names of the team members on the screen so that they could sit together at the beginning of each recitation class. Three homework questions were chosen for a particular recitation. The recitations for the two sections were coordinated by the TAs so that the recitation quiz problems given to the traditional group were based upon the homework problems selected for "peer reflection" in the PR group recitations. Each of the three "competitions" was carefully timed to take approximately 15 minutes, in order for the entire exercise to fit into the allotted fifty-minute time slot.

After each question was announced to the class, each of the nine teams was given three minutes to identify the "best" solution by comparing and discussing among the group members. If a group had difficulty coming up with a "winner", the TA/UTA would intervene and facilitate the process. The winning students were asked to come to the front of the room, where they were assembled into three second-round groups. The process was repeated, producing three finalists. These students handed in their homework solutions to the TAs, after which the TA/UTA evaluation process began. A qualitative sketch of the team structures at various stages of the competition is shown in Figure 5.1.

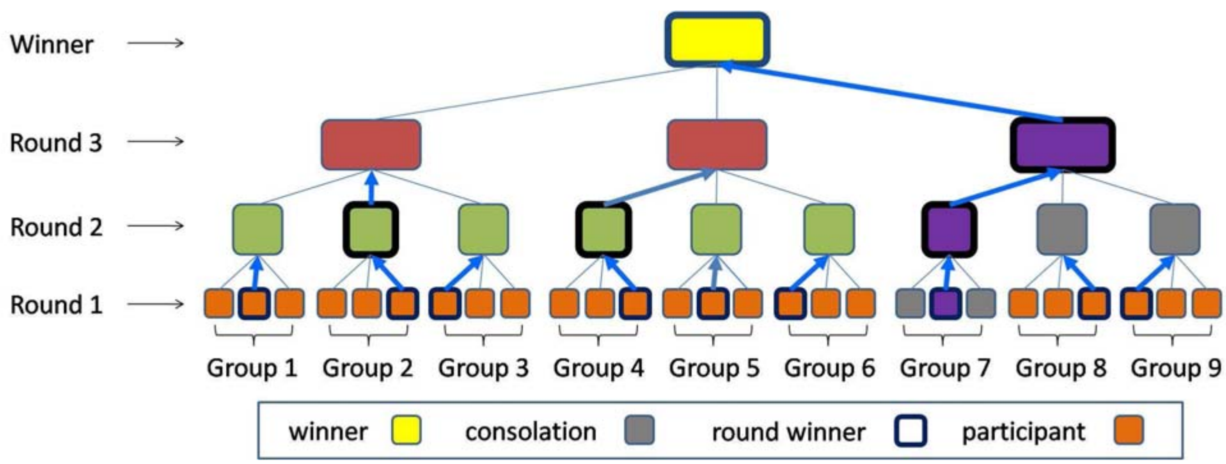


Figure 5.1. Illustration of the team structure at the three stages of peer-reflection activities. Before the students voted in the third round using the clickers, the TA and UTAs critiqued each of the three solutions at the final stage. Due to lack of space, only 3 team members per team are shown in round 1 but there were on average five members in each group in this round (as generated by a computer program each week). The consolation prize winners in gray obtained 1/3rd of the course credit awarded to the winner.

The three finalists' solutions were projected one at a time on a screen using a webcam and computer projector. Each of the three panelists (the TA and two UTAs) gave their critique of the solutions, citing what each of the finalists had done well and what could be done to further enhance the problem solving methodology in each case. In essence, the TA and UTAs were "judges" similar to the judges in the television show "American Idol" and gave their "critique" of each finalist's problem solving performance. After each solution had been critiqued by each of the panelists, the students voted on the "best" solution using the clickers. The TA and UTAs did not participate in the voting process.

Attendance was taken in the recitations using clickers for both the traditional group and the PR group. Students were given credit for attending recitation in both groups. The recitation quiz problems given to the traditional group were based upon the homework problems selected for "peer reflection" in the PR group recitation. As noted earlier, students in the PR group reflected on three homework problems in each recitation class but no recitation quiz was given to the students in the PR group at the end of the recitation classes, unlike the traditional group, primarily due to the time constraints.

In order to encourage each team in the PR group to select the student with the most effective problem solving strategy as the winner for each problem, all students from the teams whose member advanced to the final round to "win" the "competition" were given course credit (bonus points). In particular, each of these team members (consolation prize winners) earned one third of the course credit given to the student whose solution was declared to be the "winner". This reward system made the discussions very lively and the teams generally made good effort to advance the most effective solution to the next stage. Figure 5.1 shows one possible team configuration at various stages of PR activities when there are 27 students in the recitation class initially. Due to lack of space, each of the initial teams (round 1) in Figure 5.1 is shown with 3 members whereas in reality this round on average consisted of five team members. In each team, the student with the dark border in Figure 5.1 is the "winner of that round" and advances to the next stage. All the students who participated at any stage in helping select the “winner” (those shown in gray) were the consolation prize winners and obtained one third of the course credit that was awarded to the winner for that problem.

While we video-taped a portion of the recitation class discussions when students reflected with peers, a good account of the effectiveness and intensity of the team discussions came from the TA and UTAs who generally walked around from team to team listening to the discussions but not interrupting the team members involved in the discussions unless facilitation was necessary for breaking a gridlock. The course credit and the opportunity to have the finalists' solutions voted on by the whole class encouraged students to argue passionately about the aspects of their solutions that displayed effective problem solving strategies and concepts with the hope of getting partial credit for the effort. Students were constantly arguing about why drawing a diagram, explicitly thinking about the knowns and target variables, and explicitly justifying the physics principles that would be useful before writing down the equations are effective problem solving strategies.

Furthermore, the “American Idol” style recitation allowed the TAs to discuss and convey to students in much more detail what solution styles were preferred and why. Students were often shown what kinds of solutions were easier to read and understand, and which were more amenable to error-checking. Great emphasis was placed on consistent use of notation, setting up problems through the use of symbols to define physical quantities, and the importance of clear diagrams in constructing solutions.

At the end of the semester, all of the students were given a final exam consisting of 40 multiple choice questions, 20 of which were primarily conceptual in nature and 20 of which were primarily quantitative (students had to solve a numerical or symbolic problem for a target quantity). Although the final exam was all multiple-choice, a novel assessment method was used. While students knew that the only thing that counted for their grade was whether they chose the correct option for each multiple-choice question, each student was given an exam notebook which he/she could use for scratch work. We hypothesized that, even if the final exam questions were in the multiple-choice format, students who value effective problem solving strategies will take the time to draw more diagrams and do more scratch work even if there was no course credit for such activities. With the assumption that the students will write on the exam booklet and write down relevant concepts only if they think it is helpful for problem solving, a multiple-choice exam can be a novel tool for assessment. It allowed us to observe students' problem solving strategies in a more “native” form, closer to what they really think is helpful for problem solving instead of writing what the professor wants them to write down or filling the page with irrelevant equations when a free-response question is assigned.

We decided to divide the students' work in the notebooks and exam-books into two categories: diagrams and scratch work. The scratch work included everything written apart from the diagrams such as equations, sentences, and texts. Both investigators of this research (Mason and Singh) agreed on how to differentiate between diagrams and scratch work. Instead of using subjectivity in deciding how “good” the diagrams or scratch work were for each student for each of the 40 questions, we only counted the number of problems with diagrams drawn and scratch work done by each student. For example, if a student drew diagrams for 7 questions out of 40 questions and did scratch work for 10 questions out of 40 questions, we counted it as 7 diagrams and 10 instances of scratch work.

5.4 GOAL

Our goal is to examine both inter-group effects and group-independent effects of peer reflection in recitation. Inter-group effects refer to the investigation of the differences between the traditional group and the PR group because of the difference in the way recitations were structured. For example, we investigated whether there was a statistical difference in the average number of problems for which students drew diagrams and wrote scratch work in the PR group and the traditional group. Moreover, since the intervention was sustained over the whole semester (14 recitation classes), we investigated differences in how the two groups progressed. In particular, we looked for statistical differences between the PR group and the traditional group in a variable "Change" defined as the change in score from the student's average of midterm exam scores to the final exam.

We also examined group-independent effects, findings that hold for students in both the traditional group and the PR group. While investigating the differences between the PR group and the traditional group was one objective of this study, we also wanted to investigate some issues regardless of the group that have not been explicitly explored previously. One issue we examined was whether students who drew more diagrams, despite knowing that there was no partial credit for these tasks, are more likely to perform better in the final exam. We also investigated the correlation between the number of problems with diagrams or scratch work and the final exam performance when quantitative and conceptual questions were considered separately. Moreover, since students come to the college physics class with different initial preparation, another issue we examined was whether students who drew more diagrams were more likely to improve from the midterm to the final exam. We also explored whether students were more likely to draw diagrams in quantitative or conceptual questions on the final exam.

5.5 RESULTS

In the following sections, we will discuss two types of effects: the intergroup effects, i.e., whether there was a significant difference between the traditional group and the PR group on various measures, and group independent findings, e.g., the correlation between drawing diagrams and the final exam performance without regard to whether the students belonged to the traditional group or the PR group.

Before we discuss the aforementioned results, we first make note of a brief comparison of the means of both the traditional and PR groups. Table 5.1 describes the difference in means between the daytime and evening classes, and also compares the year prior to the introduction of peer reflection (fall 2006) to the year in which peer reflection was implemented in the evening classes (fall 2007). The p-values given are the results of t-tests performed between the daytime and evening classes.

While significance between groups only exists between the average midterm scores for the year in which peer reflection was implemented, the difference in average final exam scores from year to year is interesting. Traditionally, the evening class has lagged behind the daytime class by a few percentage points on the final exam, as indicated in the entry for 2006 final exams. However, after peer reflection was introduced, this difference in performance has mostly vanished.

Table 5.1. Means and p-values of comparisons for daytime and evening classes. This table compares performance during the year before peer reflection was introduced (fall 2006) to performance during the year in which it was introduced (fall 2007). An asterisk next to a mean designates the group that received peer reflection.

Daytime vs. Evening Classes	Daytime means (%)	Evening means (%)	p-value
2006: midterm exams	72.0	65.8	0.101
2006: final exams	55.7	52.7	0.112
2007: midterm exams	78.8	74.3*	0.004
2007: final exams	58.1	57.7*	0.875

For the following, we note that we will use a "null" hypothesis (van Belle 2008) for each statistical comparison, e.g., whether the average number of diagrams drawn by the students in the traditional group and the PR group is significantly different. For example, one null hypothesis we will be testing is that the population means of the number of problems with diagrams of the traditional group and PR group are equal (their difference is zero). We initially assume this null hypothesis to be true. Then, we can estimate the probability of obtaining a difference between the sample means of the number of problems with diagrams by the traditional and PR groups that is equal to or larger than the difference we observe in the data we have collected. This probability is the p-value. Consistent with the educational literature, we will regard the p-value as "small" if $p < 0.05$ in which case we will reject the null hypothesis and say that there is a significant difference between the sample means of the number of problems with diagrams by the traditional group and the PR group. Note that a small p-value results when there is a large difference between the two sample means and a large p-value results when there is a small difference between the two sample means for a fixed sample size.

The final exam had 40 multiple-choice questions, half of which were quantitative and half were conceptual. There was no partial credit given for drawing the diagrams or doing the scratch work. One issue we investigated is whether the students consider the diagrams or the scratch work to be beneficial and used them while solving problems, even though students knew that no partial credit was given for showing work. As noted earlier, our assessment method involved counting the number of problems with diagrams and scratch work. We decided to count any comprehensible work done on the exam notebook other than a diagram as a scratch work. In this sense, quantifying the amount of scratch work does not distinguish between a short and a long scratch work for a given question. If a student wrote anything other than a diagram, e.g., equations, the known variables and target quantities, an attempt to solve for unknown, etc., it was considered scratch work for that problem. Similarly, there was diversity in the quality of diagrams the students drew for the same problem. Some students drew elaborate diagrams which were well labeled while others drew rough sketches. Regardless of the quality of the diagrams, any problem in which a diagram was drawn was counted.

5.5.1 The PR Group on Average Draws More Diagrams Than the Traditional Group

Table 5.2 compares the average number of problems with diagrams or scratch work by the traditional group and the PR group on the final exam. It shows that the PR group has significantly more problems with diagrams than the traditional group. In particular, the traditional group averaged 7.00 problems with diagrams per student for the whole exam (40 problems), 4.31 problems with diagrams for the quantitative questions per student and 2.69 problems with diagrams for the conceptual questions per student. On the other hand, the PR

Table 5.2. Comparison of the average number of problems per student with diagrams and scratch work by the traditional group and the PR group in the final exam.
The PR group has significantly more problems with diagrams than the traditional group. The average number of problems with scratch work per student in the two groups is not significantly different. There are more quantitative problems with diagrams drawn and scratch work done than conceptual problems.

	Question type	Traditional group per student	PR group per student	p-value between groups
Number of problems with diagrams	All questions (40 total)	7.00	8.57	0.003
	Quantitative (20 total)	4.31	5.09	0.006
	Conceptual (20 total)	2.69	3.48	0.016
Number of problems with scratch work	All questions (40 total)	20.21	19.60	0.496
	Quantitative (20 total)	16.03	15.57	0.401
	Conceptual (20 total)	4.18	4.03	0.751

group averaged 8.57 problems with diagrams per student, 5.09 problems with diagrams for the quantitative questions per student and 3.48 diagrams for the conceptual questions per student.

5.5.2 More Diagrams for Quantitative Questions than Conceptual Questions

Table 5.2 shows that, regardless of whether they belonged to the traditional group or the PR group, students were more likely to draw diagrams for the quantitative questions than for the conceptual questions. It was not clear a priori that students will draw more diagrams for the quantitative questions than for the conceptual questions and we hypothesize that this trend may change depending on the expertise of the individuals and the type of questions asked. Table 5.2 also shows that, while there was no statistical difference between the two groups in terms of the total number of problems with scratch work performed, the students were far more likely to do scratch work for the quantitative questions than for the conceptual questions. This difference between the quantitative and the conceptual questions makes sense since problems which are primarily quantitative require calculations to arrive at an answer.

5.5.3 More Improvement from the Midterm to the Final Exam for the PR Group

60% of the questions on each of the midterm exams required students to show their work (for which students obtained partial credit) and 40% multiple-choice questions whereas the final exam only had 40 multiple-choice questions. The average midterm exam scores for the traditional and PR groups were 79% and 74%, respectively. The final exam scores for both the traditional and PR groups were 58%. We note that the final exam scores are lower than the midterm scores because there was no partial credit in the final exam. The "Change" from the average midterm to the final exam scores for the two groups suggest that the PR group improved more (statistical significance $p = 0.015$) from the averaged midterms to the final exam than the traditional group. As noted earlier, the same professor had also taught both sections of the course one year ago but all recitations for both sections of the course were taught traditionally that year. For comparison, that year, the final exam scores for the daytime and evening classes were 56% and 53%, respectively.

We note that we do not calculate the "Hake" gain (Hake 1998), (from the midterm to the final exam) which is often calculated for gain on standardized tests such as the force concept inventory (Hestenes et al. 1992) before and after instruction and is given by (Final score in % - average Midterm score in %)/(100 - average Midterm score in %). This is because the midterms and final exams in our study are not constructed in the same manner, i.e. the midterms had free-response questions and the final exam did not.

5.5.4 Change from the Average Midterm to the Final Exam is Higher for Those Who Draw More Diagrams

Now we focus on how the change from the midterm to the final exam correlates with the number of problems with diagrams or scratch work for all students without separating them into the PR group and the traditional group. In each of the cases shown in each row of Table 5.3, our null

Table 5.3. Correlations between the change from the average midterm to the final exam scores (Change) and different variables such as the number of problems with diagrams or scratch work.
Change is significantly larger for those who had more problems with diagrams (even when the number of problems with diagrams is considered separately for the quantitative or conceptual problems) or scratch work on the quantitative problems which suggests that students who improved more from the average midterm to the final exam are those who value diagrams and scratch work regardless of whether they were in the traditional group or PR group. A p-value less than 0.05 implies a positive correlation.

	R (Change)	p-value (Change)
Diagram	0.18	0.011
Diagram(Q)	0.16	0.028
Diagram(C)	0.16	0.026
Scratch work	0.18	0.012
Scratch work (Q)	0.20	0.005
Scratch work (C)	0.11	0.119

hypothesis is that there is no correlation between the variables such as the total number of problems with diagrams and Change. We reject the null hypothesis when the p-value is small and we find, e.g., that there is a weak correlation between the total number of problems with diagrams and Change. Table 5.3 shows that regardless of whether the students belonged to the traditional group or the PR group, those who had a greater "Change" were significantly more likely to have a larger number of problems with diagrams, and they were also significantly more likely to have problems with diagrams when quantitative questions or conceptual questions were considered separately (see the p values for each case in Table 5.3). Moreover, Table 5.3 shows that those who had a larger "Change" were also significantly more likely to have more problems with scratch work for the quantitative problems.

5.5.5 Final Exam Score on Average is Higher for Those Who Draw More Diagrams

Table 5.4 allows us to investigate whether the final exam score is correlated with the number of problems with diagrams or scratch work for each of the traditional group and the PR group separately. The null hypothesis in each case is that there is no correlation between the final exam score and the variable considered, e.g., the total number of problems with diagrams drawn. The Pearson coefficient R is related to the correlation of the regression line (e.g., when the final exam scores of students are plotted against the total number of problems with diagrams for students) and to the p-value. If the correlation R is larger, the deviation from the null hypothesis is larger and the p-value will be smaller. Table 5.4 shows that, for the traditional group and for the PR group, the students who had more problems with diagrams and scratch work are significantly more likely to perform well on the final exam. This result holds regardless of whether we look at all questions or the quantitative questions only (labeled diagram(Q) and scratch(Q) for the quantitative diagrams and scratch work respectively) or conceptual questions only (labeled diagram(C) and scratch(C) for the conceptual diagrams and scratch work respectively). While one may expect high-performing students to employ better problem-solving strategies, there is

Table 5.4. Correlation between the final exam scores and different variables such as the number of problems with diagrams or scratch work for each of the traditional group and the PR group.
The students who performed well on the final exam in both the traditional group and the PR group had significantly more problems with diagrams and scratch work than the students who performed poorly on the final exam. In each row, the Pearson coefficient R and the p-value are dependent variables and a higher correlation R implies a greater deviation from the null hypothesis and hence a smaller p-value.

	Traditional: Final		PR: Final	
	R	p-value	R	p-value
Diagram	0.24	0.014	0.40	0.000
Diagram(Q)	0.19	0.046	0.36	0.000
Diagram(C)	0.20	0.042	0.36	0.000
Scratch work	0.39	0.000	0.53	0.000
Scratch work (Q)	0.42	0.000	0.59	0.000
Scratch work (C)	0.28	0.004	0.32	0.002

no prior research related to physics learning at any level that shows that students who perform well will draw more diagrams even when answering multiple-choice questions where there is no partial credit for drawing diagrams. Future research will examine whether these results are dependent on whether the students are introductory or advanced physics students and the type of questions asked.

5.5.6 Diagrams Help for Both Quantitative and Conceptual Questions

We also investigated the correlations between the number of problems with diagrams drawn (or the amount of scratch work) and the final exam scores on the quantitative questions alone and the conceptual questions alone. Table 5.4 shows that within each group (traditional or PR), the correlation between the number of diagrams drawn and the final exam score is virtually identical for the quantitative and conceptual questions (R = 0.19 for quantitative vs. R=0.20 for conceptual in the traditional group, R=0.36 for quantitative vs. R=0.36 for conceptual in the peer reflection group). It is evident that for both conceptual and quantitative questions, diagrams help.

5.5.7 Diagrams Drawn by the PR Group Explain More of the Exam Performance

Since it is difficult to rate objectively the quality of each diagram or scratch work for each problem, we only counted the number of problems in which any diagram was drawn or scratch work was done. But the comparison of the traditional group and the PR group in Table 5.4 shows that for each case shown in the different rows, the correlation between the number of diagrams or amount of scratch work and the final exam score is stronger for the PR group than for the traditional group. For example, the regression coefficient R for the number of problems with diagrams drawn vs. the final exam score is higher for the PR group compared to the traditional group (R=0.40 vs. R=0.24). The PR group also showed a stronger correlation than the traditional group even when the quantitative and conceptual questions were considered separately for the correlation between the number of problems with diagrams drawn and the final exam scores. Similarly, the regression coefficient for the number of scratch work vs. the final exam score is higher for the PR group compared to the traditional group (R=0.53 vs. R=0.39). One possible interpretation of the number of diagrams or scratch work being more strongly correlated with the final exam performance for the PR group compared to the traditional group is that for the PR group the diagrams or the scratch work explain more of the final exam performance.

5.5.8 Scratch Work Explains More of the Performance for Quantitative Questions Than Conceptual Questions

Table 5.4 shows that within each group (traditional and PR) the correlation between the amount of scratch work and the final exam score was stronger for the quantitative questions than for the conceptual questions. This implies that the scratch work explains more of the performance for the quantitative problems than for the conceptual questions. This was again followed up by a further examination of the quantitative-only correlations between the scratch work and the final exam scores (R = 0.42 for the traditional group, R = 0.59 for the PR group) and the conceptual-only correlations (R = 0.28 for the traditional group, R = 0.32 for the PR group). Students wrote scratch work much more often on the quantitative problems than on the conceptual questions. This finding is reasonable because students do not necessarily have to perform algebraic manipulations for conceptual questions but it may be a pre-requisite for identifying the correct answer for a quantitative question. While it is true that a student can convert even a primarily conceptual question into a quantitative question if the student is not sure about the answer, our prior research shows that students are reluctant to convert a conceptual question into a quantitative problem and prefer to use their "gut feelings" (Singh 2008).

5.6 SUMMARY AND CONCLUSIONS

In this study, the recitation classes for an introductory physics course primarily for the health science majors were broken into a traditional group and a "peer reflection" or PR group. We investigated whether students in the PR group use better problem-solving strategies such as drawing diagrams than students in the traditional group and also whether there are differences in the performance of the two groups. We also explored whether students who perform well are the ones who are more likely to draw diagrams or write scratch work even when there is no partial credit for these activities.

In the PR group recitation classes, students reflected about their problem solving in the homework with peers each week. Appropriate guidance and support provided opportunities for learning effective problem solving heuristics to the students in the PR group. In particular, students in the PR group reflected in small teams on selected problems from the homework and discussed why solutions of some students employed better problem solving strategies than others. The TA and UTAs in the PR group recitations demonstrated effective approaches to problem solving and coached students so that they learn those skills. Each small team in the PR group discussed which student's homework solutions employed the most effective problem solving heuristics and selected a “winner”. Then, three teams combined into a larger team and repeated the process of determining the “winning” solution. Typically, once three “finalists” were identified in this manner, the TA and UTAs put each finalist's solution on a projector and discussed what they perceived to be good problem-solving strategies used in each solution and what can be improved. Finally, each student used clickers to vote on the “best” overall solution with regard to the problem solving strategies used. There was a reward system related to course credit that encouraged students to be involved to select the solution with the best problem solving strategy in each round. Students in the traditional group had traditional recitation classes in which they asked the TA questions about the homework before taking a quiz at the end of each recitation class. Each problem selected for "peer reflection" from the homework assignment for the previous week was adapted into a quiz problem for the traditional group.

The assessment of the effectiveness of the intervention was novel. The final exam had 40 multiple-choice questions, half of which were quantitative and half were conceptual. For the multiple-choice questions, students technically do not have to show their work to arrive at the answer. However, students may use effective approaches to problem solving such as drawing a diagram or writing down their plan if they believe it may help them answer a question correctly. Although students knew that there was no partial credit for drawing diagrams or doing scratch work, we compared the average number of problems with diagrams or scratch work in the traditional and experimental groups. Our hypothesis was that students who value effective problem solving strategies will have more problems in which diagrams are drawn or scratch work is done despite the fact that there is no partial credit for these activities. The fact that there was no partial credit for the diagrams or the scratch work helped eliminate the possibility of students drawing the diagrams or doing scratch work for the sole purpose of getting partial credit for the effort displayed (even if it is meaningless from the perspective of relevant physics content).

Our findings can be broadly classified into inter-group and group-independent categories. The inter-group findings that show the difference between the traditional group and PR group can be summarized as follows:

- On the multiple-choice final exam where there was no partial credit for drawing diagrams, the PR group drew diagrams in more problems than the traditional group.
- The improvement from the average midterm to the final exam was larger for the PR group than for the traditional group.

- Since it is difficult to rate the quality of each diagram or the scratch work for each problem, we only counted the number of problems in which any diagram was drawn or scratch work was done. However, the regression coefficient R for the number of problems with diagrams vs. the final exam score is somewhat higher for the PR group compared to the traditional group (R=0.40 vs. R=0.24). Similarly, the regression coefficient for the number of problems with scratch work vs. the final exam score is somewhat higher for the PR group compared to the traditional group (R=0.53 vs. R=0.39). One possible interpretation of the higher regression coefficient of the PR group compared to the traditional group is that the diagrams explain more of the exam performance for the students in the PR group.

Findings that are independent of group (which are true even when the traditional group and PR group are not separated and all students are considered together) can be summarized as follows:

- There is a significant positive correlation between how often students did scratch work or drew diagrams and how well they performed on the final exam regardless of whether they were in the traditional group or the PR group. In particular, those who performed well in the multiple-choice final exam (in which there was no partial credit for showing work) were much more likely to draw diagrams than the other students. While one may assume that high-performing students will draw more diagrams even when there is no partial credit for it, no prior research that we know of has explicitly demonstrated a correlation between the number of "genuinely drawn" diagrams and student performance at any level of physics instruction.
- Students who drew more diagrams or did more scratch work were more likely to show improvement from the average midterm to the final exam than the other students.

- The correlations between the number of problems with diagrams drawn and the final exam scores on the quantitative questions alone and the conceptual questions alone are comparable. Evidently, for both conceptual and quantitative questions, drawing diagrams helped students.
- Students in both groups were more likely to draw diagrams or do scratch work for quantitative problems than for the conceptual questions. While more scratch work is expected on quantitative problems, it is not clear a priori that more diagrams will be drawn for the quantitative problems than for the conceptual questions. We hypothesize that this trend may depend upon the expertise of the individuals, explicit training in effective problem solving strategies and the difficulty of the problems.

Finally, the students in the traditional group were given weekly recitation quizzes in the last 20 minutes of the recitation class based upon that week's homework. There were no recitation quizzes in the PR group due to the time constraints. It is sometimes argued by faculty members that the recitation quizzes are essential to keep students engaged in the learning process during the recitations. However, this study shows that the PR group was not adversely affected by not having the weekly quizzes that the traditional group had and instead having the peer reflection activities. The mental engagement of students in the PR group throughout the recitation class may have more than compensated for the lack of quizzes. The students in the PR group were evaluating their peer's work along with their own which requires a high level of mental processing (Marzano et al. 1988). They were comparing problem solving strategies such as how to do a conceptual analysis and planning of the solution, why drawing two separate diagrams may be better in certain cases (e.g., before and after a collision) than combining the information into one diagram, how to define and use symbols consistently etc. In addition, after the active engagement with peers, students got an opportunity to learn from the TA and UTAs about their critique of each solution highlighting the strengths and weaknesses. Reflection with peers gives students an opportunity to explain to the peers the aspects of their problem solving approach which are effective. This type of opportunity to communicate with peers can help students monitor their own learning and improve their ability to articulate their scientific point of view.

5.7 ACKNOWLEDGMENTS

We are very grateful to Dr. L. Pingel from the School of Education and Dr. Allan R. Sampson and Huan Xu from the Department of Statistics at the University of Pittsburgh for their help in data analysis. We thank F. Reif, R. Glaser, R. P. Devaty and J. Levy for very useful discussions. We thank J. Levy for his enthusiastic support of helping us evaluate the "peer reflection" activities in his class.

5.8 REFERENCES

Black, P., & Wiliam, D. (1998a). "Inside the black box: Raising standards through classroom assessment." *Phi Delta Kappan*, 80(2), 139-148.

Black, P., & Wiliam, D. (1998b). "Assessment and classroom learning", *Assessment in Education*, 5 (1), 7-74.

Chi, M., Bassok, M., Lewis, M., Reiman, P., & Glaser, R. (1989). "Self explanations: How students study and use examples in learning to solve problems." *Cognitive Science*, 13, 143-182.

Chi, M., Feltovich, P., & Glaser, R. (1981). "Categorization and representation of physics knowledge by experts and novices." *Cog. Sci.* 5, 121-152.

Collins, A., Brown, J., & Newman, S. (1989). "Cognitive Apprenticeship: Teaching the crafts of reading, writing and mathematics." In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser,* Hillsdale, NJ: Lawrence Erlbaum., 453-494.

Crouch, C., & Mazur, E. (2001). "Peer Instruction: Ten years of experience and results." *Am. J. Phys.*, 69, 970-977.

Dufresne, R., Mestre, J., Thaden-Koch, T., Gerace, W., & Leonard, W. (2005). "Knowledge representation and coordination in the transfer process." In *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. P. Mestre (Information Age Publishing, Greenwich, CT), pp. 155-215.

Gick, M., & Holyoak, K. (1987). "The cognitive basis of knowledge transfer." In *Transfer of learning: Contemporary research and applications*, Cornier & Hagman (Eds.), New York, Academic Press.

Hake, R. (1998). "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses." *Am. J. Phys.* 66, 64.

Hardiman, P., Dufresne, R., & Mestre, J. (1989). "The relation between problem categorization and problem solving among novices and experts." *Memory and Cognition* 17, 627-638.

Heller, P., Keith, R., & Anderson, S. (1992). "Teaching problem solving through cooperative grouping. Part I: Group versus individual problem solving." *Am. J. Phys.* 60, 627.

Heller, P., & Hollabaugh, M. (1992). "Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups." *Am. J. Phys.* 60, 637.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). "Force Concept Inventory." *Phys. Teach.* 30, 141.

Holyoak, K., & Thagard, P. (1995). "Mental Leaps: Analogy in creative thought." Cambridge, MA: MIT press.

Kaput, J. (1985). "Representation and problem solving: Methodological issues related to modeling." In *Teaching and Learning Mathematical Problem Solving: Multiple Research Perspectives,* Edited by E. A. Silver. Lawrence Erlbaum, Hillsdale, NJ, pp. 381-398.

Larkin, J. (1981). "Cognition of learning physics." *Am. J. Phys.* 49 (6), 534-541.

Larkin, J. (1985). "Understanding, problem representations, and skill in physics." In *Thinking and learning skills.* Edited by S. F. Chipman, J. W. Segal and R. Glaser. Lawrence Erlbaum, Hillsdale, NJ, pp. 141-159.

Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980). "Expert and novice performance in solving physics problems." *Science* 208, 1335-1362.

Larkin, J., & Reif, F. (1979). "Understanding and teaching problem solving in physics." *Eur. J. Sci. Educ.* 1(2), 191-203.

Larkin, J., & Simon, H. (1989). "Why a diagram is (sometimes) worth ten thousand words." In H.A. Simon, *Models of Thought (Yale Univ. Press, New Haven), Vol. II,* pp. 413-437.

Maloney, D. (1994). "Research in problem solving: Physics." In *Handbook of Research on the Teaching and Learning of Science*, edited by D. Gabel (MacMillan, New York).

Marzano, R., Brandt, R., Hughes, C., Jones, B., Presseisen, B., Rankin, S., & Suhor, C. (1988). *Dimensions of thinking: A framework for curriculum and instruction.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mazur, E. (1997a). *Peer Instruction: A User's Manual*. Series title: *Prentice Hall series in educational innovation.* Upper Saddle River, Prentice Hall, N.J.

Mazur, E. (1997b). "Understanding or memorization: Are we teaching the right thing?" In *Conference on the Introductory Physics Course on the occasion of the retirement of Robert Resnick,* Ed. J. Wilson, 113-124, Wiley, NY.

Mazur, E. (2009). "Farewell, Lecture?" *Science*, 323, 50-51.

Reif, F. (1981). "Teaching problem solving – A scientific approach." *Phys. Teach.* 19, 310-316.

Reif, F. (1986). "Scientific approaches to science education." *Phys. Today* 39 (11), 48-54.

Reif, F. (1995). "Millikan Lecture 1994: Understanding and teaching important scientific thought processes." *Am. J. Phys.* 63, 17.

Rogoff, B. (1990). *Apprenticeship in thinking.* NY: Oxford University Press.

Schoenfeld, A. (1985). *Mathematical Problem Solving.* NY: Academic Press.

Schoenfeld, A. (1989). "Teaching mathematical thinking and problem solving." In *Toward the Thinking Curriculum: Current Cognitive Research*, edited by L. B. Resnick and B. L. Klopfer (ASCD, Washington, DC,), pp. 83-103.

Schoenfeld, A. (1992). "Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics." In *Handbook for Research on Mathematics Teaching and Learning*, edited by D. Grouws (McMillan, NY), Chap. 15, pp. 334-370.

Schoenfeld, A. & Herrmann, D. (1982). "Problem perception and knowledge structure in expert-novice mathematical problem solvers." *J. Exp. Psych. Learning, Memory, and Cognition* 8, 484-494.

Singh, C. (2002). "When physical intuition fails." *Am. J. Phys* 70 (11), 1103-1109.

Singh, C., (2005). "Impact of peer interaction on conceptual test performance." *Am. J. Phys.,* 73(5), 446-451.

Singh, C. (2008). "Assessing student expertise in introductory physics with isomorphic problems, Part II: Examining the effect of some potential factors on problem solving and transfer." *Phys. Rev. ST Phys. Educ. Res.* 4, 010105(1-10).

Van Belle, G. (2008). *Statistical rules of thumb: Second Edition, Wiley's series in Probability and Statistics.* Wiley, NY.

Yerushalmi, E., Singh, C., & Eylon, B. (2007). "Physics learning in the context of scaffolded diagnostic tasks (I): The experimental setup." In *Proceedings of the Phys. Ed. Res. Conference, Greensboro, NC*, edited by L. Hsu, L. McCullough, and C. Henderson, AIP Conf. Proc. 951, 27-30.

6.0 ATTITUDES AND APPROACHES TO PROBLEM SOLVING (AAPS) SURVEY

6.1 ABSTRACT

Students' attitudes and approaches to problem solving in physics can profoundly influence their motivation to learn and development of expertise. We administered an Attitudes and Approaches to Problem Solving (AAPS) survey to physics graduate students. Comparison of their responses to the survey questions about problem solving in their own graduate-level courses vs. problem solving in the introductory physics courses provides insight into their expertise in introductory and graduate level physics. The physics graduate students' responses to the survey questions were also compared with those of introductory physics and astronomy students and physics faculty members. We find that, even for problem solving in introductory physics, graduate students' responses to some survey questions are less expert-like than those of the physics faculty members. A comparison of survey responses shows more expert-like responses from graduate students than introductory students about introductory physics. However, survey responses suggest that graduate level problem solving by graduate students on several measures of the survey has remarkably similar trends to introductory level problem solving by introductory students.

6.2 INTRODUCTION

Students' attitudes and approaches towards learning can have a significant impact on what they actually learn (Schommer 1990, Halloun 1996, Redish et al. 1998, Adams et al. 2006, Gray et al. 2008). Mastering physics amounts to developing not only a robust knowledge structure of physics concepts but also positive attitudes about the knowledge and learning in physics. In essence, it is impossible to become a true physics expert without a simultaneous evolution of expert-like attitudes about the knowledge and learning in physics. If students think that physics is a collection of disconnected facts and formulas rather than seeing the coherent structure of the knowledge in physics, they are unlikely to see the need for organizing their knowledge hierarchically. Similarly, if students believe that only a few smart people can do physics, the teacher is the authority and the students' task in a physics course is to take notes, memorize the content and reproduce it on the exam and then forget it, they are unlikely to make an effort to synthesize and analyze what is taught, ask questions about how concepts fit together or how they can extend their knowledge beyond what is taught. Similarly, if students believe that if they cannot solve a problem within 10 minutes, they should give up, they are unlikely to persevere and make an effort to explore strategies for solving challenging problems.

The Maryland Physics Expectation Survey (MPEX) was developed to explore students' attitudes and expectations related to physics (Redish et al. 1998). When the survey was administered before and after instruction in various introductory physics courses, it was found that students' attitudes about physics after instruction deteriorated compared to their expectations before taking introductory physics. Very few carefully designed courses and curricula have shown major improvements in students' expectations after an introductory physics course (Elby

2001, Laws 1999, Marx and Cummings 2007). The Colorado Learning Assessment Survey (CLASS) is another survey which is similar to the MPEX survey and explores students' attitudes about physics (Adams et al. 2006, Gray et al. 2008). The analysis of CLASS data yields qualitatively similar results to those obtained using the MPEX survey. Moreover, when introductory physics students were asked to answer the survey questions twice, once providing the answers from their perspective and then from the perspective of their professors, introductory students' responses to many questions were very different from their perspective compared to what they claimed would be their professors' perspective (Gray et al. 2008). Thus, introductory students maintained their views although they knew that the physics professors would have different views about some of the survey questions.

Cummings et al. (2004, Marx and Cummings 2007) developed the Attitudes towards Problem Solving Survey (APSS) which is partially based upon MPEX. The original APSS survey has 20 questions and examines students' attitudes towards physics problem solving (Cummings et al. 2004). The survey was given to students before and after instruction at three types of institutions: a large university, a smaller university and a college. It was found that students' attitudes about problem solving did not improve after instruction (deteriorated slightly) at the large university and the attitudes were least expert-like (least favorable) at the large university with a large class.

Students' attitudes and approaches to learning and problem solving can affect how they approach learning and how much time they spend repairing, extending and organizing their knowledge. If instructors are aware of students' attitudes and approaches to problem solving, they can explicitly exploit strategies to improve them. For example, knowing students' beliefs about mathematics learning (which is similar to students' beliefs about physics learning in many aspects) motivated Schoenfeld to develop a curriculum to improve students' attitude (Collins et al. 1989, Schoenfeld 1985, Schoenfeld 1989, Schoenfeld 1992). In particular, based on the knowledge that students in the introductory mathematics courses often start looking for formulas right away while solving a mathematics problem instead of performing a careful conceptual analysis and planning, Schoenfeld used an explicit strategy to change students' approach. He routinely placed students in small groups and asked them to solve problems. He would move around and ask them questions such as “What are you doing? Why are you doing it? How does it take you closer to your goals?” Very soon, students who were used to immediately looking for formulas were embarrassed and realized that they should first perform conceptual analysis and planning before jumping into the implementation of the problem solution. Schoenfeld's strategy helped most students adopt an effective problem solving approach within a few weeks and they started to devote time to qualitative analysis and decision-making before looking for equations (Collins et al. 1989).

Another unfavorable attitude about mathematical problem solving that Schoenfeld wanted students to change was that students often felt that if they could not solve a problem within 5-10 minutes, they should give up (Collins et al. 1989, Schoenfeld 1985, Schoenfeld 1989, Schoenfeld 1992). Schoenfeld realized that one reason students had such an attitude was that they observed their instructor solving problems during the lectures without faltering or spending too much time thinking. To bust this myth about problem solving, Schoenfeld began each of his geometry classes with the first 10 minutes devoted to taking students' questions about challenging geometry problems (often from the end of the chapter exercises) and thus attempting to solve them without prior preparation. Students discovered that Schoenfeld often struggled with the problems and was unable to solve them in the first 10 minutes and asked students to continue to think about the problems until one of them had solved it and shared it with others. This approach improved students' attitude and raised their self-confidence in solving mathematics problems.

Here, we discuss responses of physics graduate students on the Attitudes and Approaches to Problem Solving (AAPS) survey, a modified version of APSS survey (Cummings et al. 2004) that includes additional questions related to approaches to problem solving. We explore how graduate students differ in their attitudes and approaches while they solve graduate level problems versus introductory level problems. We find that, on some measures, graduate students have very different attitudes and approaches about solving introductory physics problems compared to their own graduate level problems. The attitudes and approaches of graduate students on the AAPS survey were also compared to those of introductory physics and astronomy students and to physics faculty. We find that the attitudes and approaches of graduate students differ significantly from introductory students and physics faculty on several measures. Overall, the patterns that emerge from the graduate students' responses to the AAPS survey require reflection on whether their attitudes and approaches to problem solving in the graduate courses are commensurate with the goals of those courses.

6.3 DEVELOPMENT OF THE AAPS SURVEY

The AAPS survey is given in section C.1 of Appendix C. In order to develop the survey, we selected 16 questions from the APSS survey (Cummings et al. 2004) and tweaked some of them for clarity based upon in-depth interviews with five introductory physics students and discussions with some faculty. These 16 questions constitute the first 14 questions and the last two questions of the AAPS survey. We also developed 17 additional questions, many of which focused on approaches individuals take to problem solving, and modified them based upon the feedback from introductory students during interviews and discussions with three physics faculty. The reason introductory physics students and faculty were sought for this purpose is that we hypothesized that the responses of these two groups would be the most disparate and would provide the most diverse feedback for improving the preliminary survey. Some of the themes in the additional questions are related to the use of diagrams and scratch work in problem solving, use of "gut" feeling vs. using physics principles to answer conceptual questions, reflection on one's solution after solving a problem to learn from it, giving up on a problem after 10 minutes, preference for numerical vs. symbolic problems, and enjoying solving challenging physics problems.

The in-depth interviews with five students from a first-semester algebra-based class, which was part of the validation of the survey, were helpful in ensuring that the questions were interpreted clearly by the students. Of approximately 40 introductory students responding to the invitation for paid interviews, five were selected. Since we wanted all students to be able to interpret the problems, two students were randomly chosen for interviews from those who scored above 70% and three students were chosen who obtained below 70% on their first midterm exam.

The survey questions were administered to all interviewed students in the form of statements that they could agree or disagree with on a scale of 1 (strongly agree) to 5 (strongly disagree) with 3 signifying a neutral response. During the individual interviews, students were also asked to solve some physics problems using a think-aloud protocol to gauge whether their responses to the survey questions about their attitudes and approaches to problem solving were consistent with the attitudes and approaches displayed while actually solving problems. In this protocol, we asked individuals to talk aloud while answering the questions. We did not disturb them while they were talking and only asked for clarifications of the points they did not make clear on their own later. While it is impossible to grasp all facets of problem solving fully by having students solve a few problems, a qualitative comparison of their answers to the survey questions and their actual approaches to solving problems was done after the interviews using the think aloud protocol. This comparison shows that students were consistent in their survey responses in many cases but in some instances they selected more favorable (expert-like) responses to the survey questions than the expertise that was explicitly evident from their actual problem solving. In this sense, the favorable responses (at least for the introductory students) should be taken as the upper limit of the actual favorable attitudes and approaches to problem solving.

6.4 EXPERIMENT – ADMINISTRATION OF THE AAPS SURVEY

The final version of the AAPS survey was first administered anonymously to 16 physics graduate students enrolled in a graduate level course. The expert (favorable) responses are given in section C.2 of Appendix C. Discussions with the graduate students after they took the survey showed that all of them interpreted that the survey was asking about problem solving in their own graduate courses and that they would have answered the questions differently if they were asked about their attitudes and approaches to solving introductory physics problems. We then administered the survey a second time to 24 graduate students (there was overlap between the first cohort of 16 graduate students and this cohort) with the questions explicitly asking them to answer each question about their attitudes and approaches to introductory physics problem solving. Due to lack of class time, this second round of survey was administered online. We had individual discussions with 4 graduate students about the reasoning for their AAPS survey responses and invited all who had answered the questions to write a few sentences explaining their reasoning for selected survey questions online. We explicitly asked them to explain their reasoning when they answered the survey questions about problem solving in the graduate level courses and separately for introductory physics. Ten graduate students (out of 24 who took the survey online) provided written reasoning for their responses.

We also administered the AAPS survey to several hundred introductory students in two different first-semester and second-semester algebra-based physics courses and to students in the first and second-semester calculus-based courses. Specifically, there were two sections of the first-semester algebra-based physics course with 209 students, two sections of the second-semester algebra-based physics course with 188 students, one first-semester calculus-based course section with 100 students, and a second-semester calculus-based course section with 44 students. In all of these courses, students were given a small number of bonus points for taking the survey. In addition, the survey was given to 31 students in an astronomy course which is the first astronomy course taken by students who plan to major in Physics and Astronomy (but fewer than 20% of the students in this course actually major in Physics and Astronomy). The students in the astronomy course who did not want to major in physics and astronomy were not required to take that course, unlike the students in the introductory physics courses (the calculus-based introductory courses are dominated by engineering majors and the algebra-based courses by those interested in health-related professions who must take two physics courses to fulfill their requirements). For the astronomy course, the word “physics” in the survey was replaced throughout by “astronomy,” e.g., “in solving astronomy problems...” Also, the contexts (which were not related to astronomy) were removed from questions 32 and 33 of the survey for astronomy students. Finally, the survey was given to 12 physics faculty who had taught introductory physics recently. Half of the faculty members were those who also gave the survey to their students. We also discussed faculty responses to selected questions individually with some of them.

6.4.1 Scoring Systems

We did not differentiate between “agree” and “strongly agree” in interpreting the data. Similarly, “disagree” and “strongly disagree” were combined for streamlining the data and their interpretation. A favorable response refers to either “agree” or “disagree” based upon which one was chosen by a majority of physics faculty. As we will discuss later, for some questions, the favorable response is supported by almost the entire faculty, but for a few questions the favorable response may only have the support of 60 – 70% of the faculty.

We report the data in two ways. First, the “net” average favorable response, shown in Table 6.1, was calculated as defined by Cummings et al. (2004): a value of +1 is assigned to each favorable response, a value of -1 is assigned to each unfavorable response, and a 0 is assigned to neutral responses. We then averaged these values for everybody in a particular group (e.g., faculty) to obtain a “net” favorable response for that group. Thus, the net favorable response for a group indicates how expert-like the survey response of the group is to each survey question. A second method for representing the data is by separately showing the average percentage of favorable and unfavorable responses for each question for each group (the neutral responses are 100% minus the percentage of favorable and unfavorable responses). We will use this second method of data representation for all of our graphical representations of data below.

6.5 RESULTS

Table 6.1 shows the net average responses for all groups to each individual question and averaged over all questions (see the last entry in Table 6.1). As noted earlier, the survey questions were designed in an "agree" or "disagree" format. There is leeway to agree or disagree "somewhat" or "strongly" (although we did not distinguish between these in our analysis presented here), and one may also select a neutral response. The favorable (expert) responses for each question based upon the responses chosen by most physics faculty are given in Appendix C.2. In Table 6.1, the introductory physics students from both semesters of algebra-based and calculus-based courses were lumped into one group because there was no significant difference between the "net" average survey responses of these groups.

6.5.1 Graduate students: graduate-level vs. introductory-level problem solving

Figure 6.1 compares the AAPS survey responses of graduate students to selected questions for which differences were observed when they answered the questions about problem solving in their graduate courses and problem solving in introductory physics. The error bars shown on the

Table 6.1. Average responses for students and faculty over each individual question, and averaged over all survey questions (see the last entry).

Problem number	1	2	3	4	5	6	7
Faculty	0.83	1.00	0.50	0.92	0.92	1.00	0.83
Graduate students - intro	0.71	0.42	-0.04	0.83	0.17	0.75	0.83
Graduate students - self	0.40	0.63	-0.13	0.75	0.63	0.25	0.88
Astronomy 113 students	0.45	0.48	-0.16	0.58	0.13	0.71	0.84
All introductory students	0.14	0.19	0.15	0.41	0.16	0.24	0.61
Problem number	8	9	10	11	12	13	14
Faculty	0.92	0.58	0.92	0.67	0.83	1.00	0.50
Graduate students - intro	0.83	0.46	0.88	0.67	0.54	0.88	0.88
Graduate students - self	1.00	0.31	0.69	0.33	0.44	0.94	0.81
Astronomy 113 students	0.77	0.35	0.94	0.23	0.10	0.74	0.77
All introductory students	0.67	0.24	0.58	-0.03	-0.06	0.56	0.32
Problem number	15	16	17	18	19	20	21
Faculty	1.00	0.67	1.00	0.92	1.00	0.75	1.00
Graduate students - intro	0.96	0.50	0.79	0.96	0.88	0.38	0.92
Graduate students - self	0.94	0.31	0.50	0.88	0.56	0.25	0.94
Astronomy 113 students	0.29	0.52	0.06	0.19	0.84	0.32	0.90
All introductory students	0.74	0.23	0.55	0.69	0.77	-0.19	0.71
Problem number	22	23	24	25	26	27	28
Faculty	1.00	0.92	0.42	0.92	1.00	0.92	1.00
Graduate students - intro	1.00	1.00	0.21	0.54	0.71	0.67	0.96
Graduate students - self	1.00	0.75	0.19	0.38	0.50	0.63	0.88
Astronomy 113 students	0.77	0.74	0.06	0.68	0.55	0.74	0.87
All introductory students	0.52	0.40	0.43	0.56	0.37	0.03	0.75
Problem number	29	30	31	32	33	All	
Faculty	1.00	1.00	1.00	1.00	1.00	0.87	
Graduate students - intro	1.00	0.92	0.92	1.00	0.83	0.73	
Graduate students - self	1.00	0.69	1.00	0.88	0.19	0.62	
Astronomy 113 students	0.68	-0.19	-0.13	0.68	0.50	0.49	
All introductory students	0.74	-0.04	0.08	0.70	0.46	0.38	

histograms (and in all the other figures in this chapter) indicate the standard error. One typical difference between introductory and graduate level problem solving is that the graduate students display more expert-like (favorable) attitudes and approaches while solving introductory level problems than while solving graduate level problems. For example, in response to question 1 for problem solving in their graduate level courses, approximately 40% of the graduate students felt that if they were not sure about the right way to start a problem, they would be stuck unless they got help but only 20% felt this way when solving introductory physics problems. Also, they were more likely to reflect upon physics principles that may apply and see if they yield a reasonable solution when not sure about the approach while solving introductory problems than while solving graduate level problems (see response to question 10 in Figure 6.1). They were also more likely to be able to tell that their answer was wrong without external input while

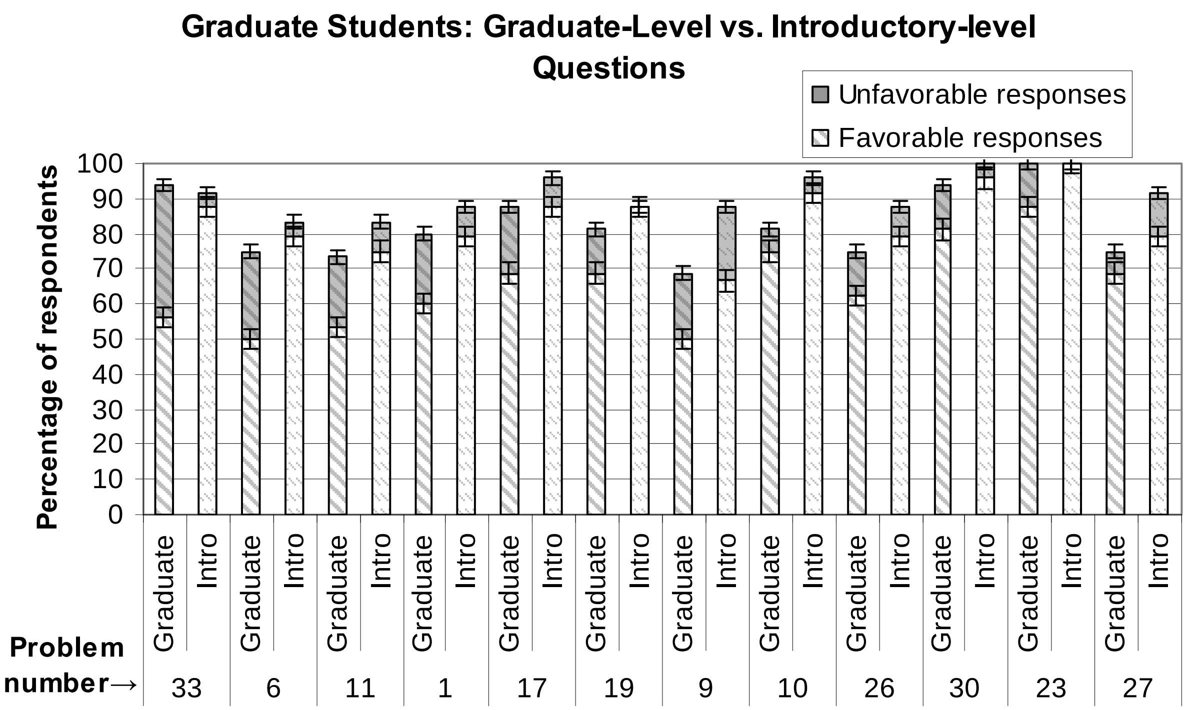


Figure 6.1. Comparison of graduate students' survey responses to 12 selected questions when considering introductory-level problem solving and graduate-level problem solving.
The order of the questions was selected such that the difference between the introductory-level problem solving ("Intro" in the figure) and graduate-level problem solving ("Graduate") is largest for the first question (question 33), second largest for the second question (question 6), etc. Error bars shown here (and in all other figures) are the standard errors. The responses to these questions are more favorable for introductory level problem solving than for graduate-level problem solving.

solving introductory problems than graduate level problems (see response to question 6 in Figure 6.1). Graduate students were approximately 20% more likely to claim that they routinely use equations to calculate answers even if they are non-intuitive while solving graduate level problems than while solving introductory level problems (see response to question 11 in Figure 6.1).

While none of the graduate students claimed they would give up solving an introductory physics problem if they could not solve it within 10 minutes, approximately 15% claimed they would give up after 10 minutes while solving a graduate level problem (see response to question 23 in Figure 6.1). While approximately 80% of the graduate students claimed they enjoy solving introductory physics problems even though it can be challenging at times, less than 70% of them said the same about the graduate level problems (see response to question 27 in Figure 6.1). Also, more graduate students claimed that it is useful for them to solve a few difficult problems using a systematic approach and learn from them rather than solving many similar easy problems one after another when solving introductory level problems than for graduate level problems (see response to question 26 in Figure 6.1).

As shown in Table 6.1, the introductory physics students enjoyed solving challenging problems even less than the graduate students and were also less likely to find solving a few difficult problems more useful than solving many easy problems based upon the same principle. One introductory student stated in an interview that he feels frustrated with an incorrect problem solution and feels satisfied when he gets a problem right, which motivates him to continue to do problem solving. Therefore, he likes easier problems.

In response to survey question 33, close to 90% of the graduate students agreed that two introductory level problems, both of which involve conservation of energy, can be solved using similar methods whereas only approximately 55% of them agreed that both problems can be solved using similar methods when solving graduate level conservation of energy problems (see Figure 6.1). Individual discussions with a subset of graduate students suggest that they felt that since air-resistance and friction were involved, they might have to use different methods to solve the problems. In particular, they noted that they often use different methods involving Lagrangians and Hamiltonians to solve complicated problems in graduate level courses and they were not sure if the same technique will be useful in problems involving friction and air-resistance. In response to survey question 33, the entire physics faculty noted that both problems can be solved using similar methods (see Table 1). When we asked some physics faculty individually whether their responses would be different if the survey question 33 was for introductory level problems vs. graduate level problems, they noted that their responses would not differ. Different responses from graduate students points to the fact that graduate students who are taking graduate level physics courses are immersed in learning complicated mathematical techniques and they are evaluating their survey responses in light of their experiences with mathematical tools. When some physics faculty members were shown the responses of graduate students to question 33, they commented that sometimes graduate students are so focused on mathematical manipulations in the graduate-level courses, they tend to use unnecessarily complicated techniques even when they are asked questions which can be solved using introductory level techniques, e.g., Gauss's law.

6.5.2 Graduate students: Comparison with other survey groups

Figure 6.2 shows that on question 5 of the survey, while no faculty agreed with the statement (no unfavorable response) that problem solving in physics basically means matching problems with the correct equations and then substituting values to get a number, more than 30% of the graduate students agreed with the statement (unfavorable) in the context of introductory level problem solving and approximately 20% agreed with the statement for graduate level problem solving. Incidentally, the responses of introductory physics students to this question were indistinguishable from those of the graduate students for introductory physics problem solving. But individual discussions show that there is difference in the reasoning of many graduate students and introductory physics students. For example, individual discussions show that for introductory level problem solving, many graduate students felt so comfortable with the applications of basic principles that not much explicit thought was involved in solving the introductory level problems. For example, in response to question 5, one graduate student noted:

- "Well for introductory physics this is true. But, in more advanced problems you kind of have to setup the equations."

Some graduate students reflected explicitly on their introductory physics experiences and compared it to the graduate level experiences in problem solving. For example, one graduate student noted:

- "…you can get an expression from two others without understanding how or why. As an introductory student I probably did this more because the expressions were simpler and easier to manipulate without a 100% understanding. My motives were also more to get the work done than to learn every detail."

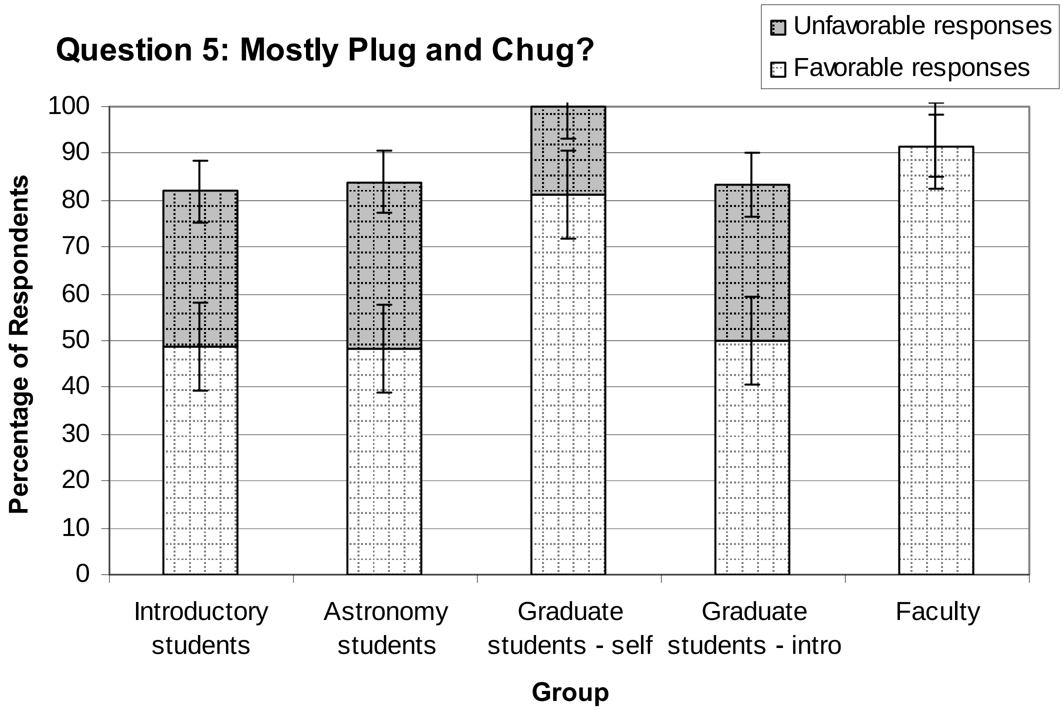


Figure 6.2. Histogram showing favorable (disagree) and unfavorable (agree) responses for survey question 5 about whether problem solving in physics is mainly an exercise in finding the right formula. The histogram shows that a large number of non-faculty respondents from all groups agreed with the statement or were neutral.

On the other hand, prior research shows that many introductory physics students think that physics is a collection of disconnected facts and formulas and use a "plug and chug" approach without thinking if a principle is applicable in a particular context (Redish et al. 1998, Adams et al. 2006).

Figure 6.3 shows that, in response to question 6, all of the physics faculty noted that while solving physics problems they could often tell when their work and/or answer is wrong even without external resources but only approximately 50% of the graduate students could do so while solving graduate level problems and approximately 80% of the graduate students could do so for introductory level problem solving. Moreover, the response of the graduate students to this question for graduate level problems is similar to that of the introductory physics students for introductory level problems. Such similarity suggests that while graduate students may be experts in solving introductory problems, they are not experts in their own graduate level courses.

Figure 6.4 shows that, in response to question 11 about whether equations must be intuitive in order to use them or whether they routinely use equations even if they are non-intuitive, graduate students' responses while solving introductory physics problems were similar to those of faculty and approximately 75% disagreed with the statements (favorable response) in question 11. When answering graduate level problems, only slightly more than 50% of the graduate students noted that equations must be understood in an intuitive sense before being

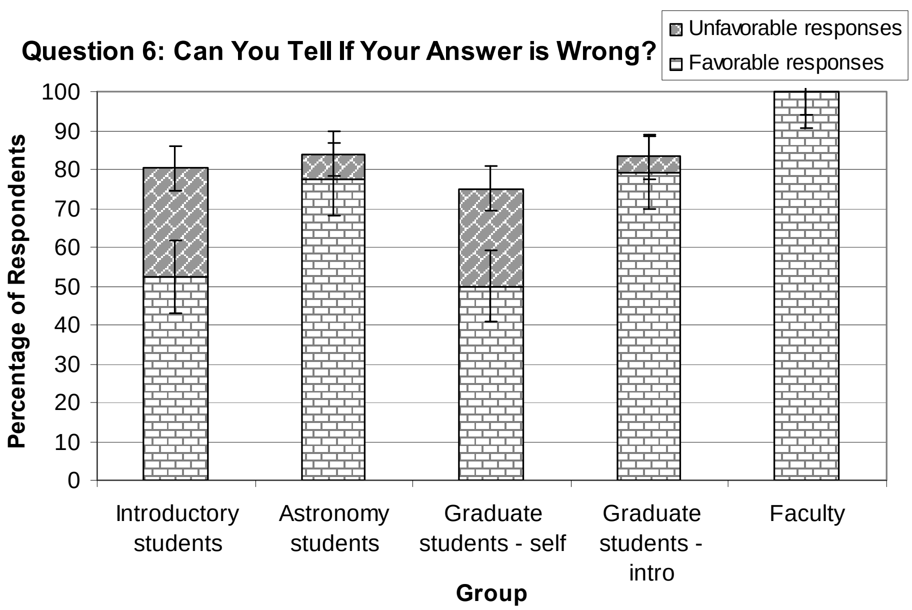


Figure 6.3. Histogram showing favorable (agree) and unfavorable (disagree) responses for survey question 6. The histograms show that faculty members were always aware of when they were wrong in problem solving but other respondents were less certain. Only about 50% of graduate students could tell that their answers were wrong in graduate level problem solving.

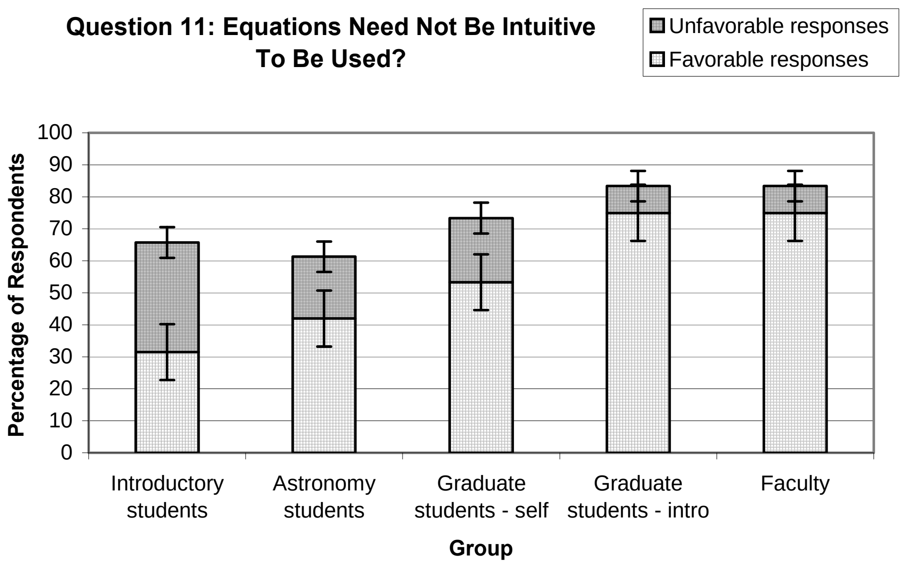


Figure 6.4. Histogram showing favorable (disagree) and unfavorable (agree) responses for survey question 11. The histogram shows that many more graduate students disagreed that they routinely use equations to calculate answers even if they are non-intuitive for introductory level problem solving than graduate level problem solving. An almost equal number of introductory physics students agreed and disagreed with the statement.

used. Individual discussions show that the graduate students felt that sometimes the equations encountered in the graduate courses are too abstract and they do not have sufficient time to make sense of them and ensure that they have built an intuition about them. The following responses from some graduate students reflect their sentiments:

- “...you just cannot understand everything. So it's ok to deal with the homework first. But I really feel bad when I do plug and chuck (sic).”
- “I am often still presented with equations to calculate something without enough motivation to understand the process, even at the graduate level, and being able to use the equation and accept that you'll understand it later is often necessary. For students' first course in physics, this is more the rule than the exception at some level...”

- “I remember physics via the equations, so I try my best to always understand the meaning. But if I can't, I fall back on ‘this is the equation, use it’.”
- “As an introductory student I had the point of view that the equations are right so my intuition must be wrong. I used equations to get the answer whether it made sense at first or not, but I trained my intuition with every such result. I had more faith in the physics that is taught to me than the physics intuition I acquired just by observation. As a graduate student, one is already used to the unintuitive results being the correct one, they have by then become intuitive."

The last graduate student quoted above expresses an interesting view that by the time one goes to graduate school in physics, one may have learned to accept non-intuitive results and unintuitive results start appearing intuitive. In contrast to the graduate students, the responses of the introductory physics students show that they are even more likely than graduate students to use equations to calculate answers even if they are non-intuitive (see Figure 6.4). This finding is consistent with the prior results that show that many introductory students view problem solving in physics as an exercise in finding the relevant equations rather than focusing on why a particular physics principle may be involved and building an intuition about a certain type of physics problems (Redish et al. 1998, Adams et al. 2006).

Figure 6.5 shows that, in response to question 12 regarding whether physics involves many equations each of which applies primarily to a specific situation, all but one physics faculty members disagreed with the statement (favorable) but less than 70% of the graduate students disagreed with it when solving graduate level problems. The percentage of introductory physics students who disagreed with the statement was slightly more than 35% and slightly more than 40% for the introductory astronomy students. These responses are commensurate with the

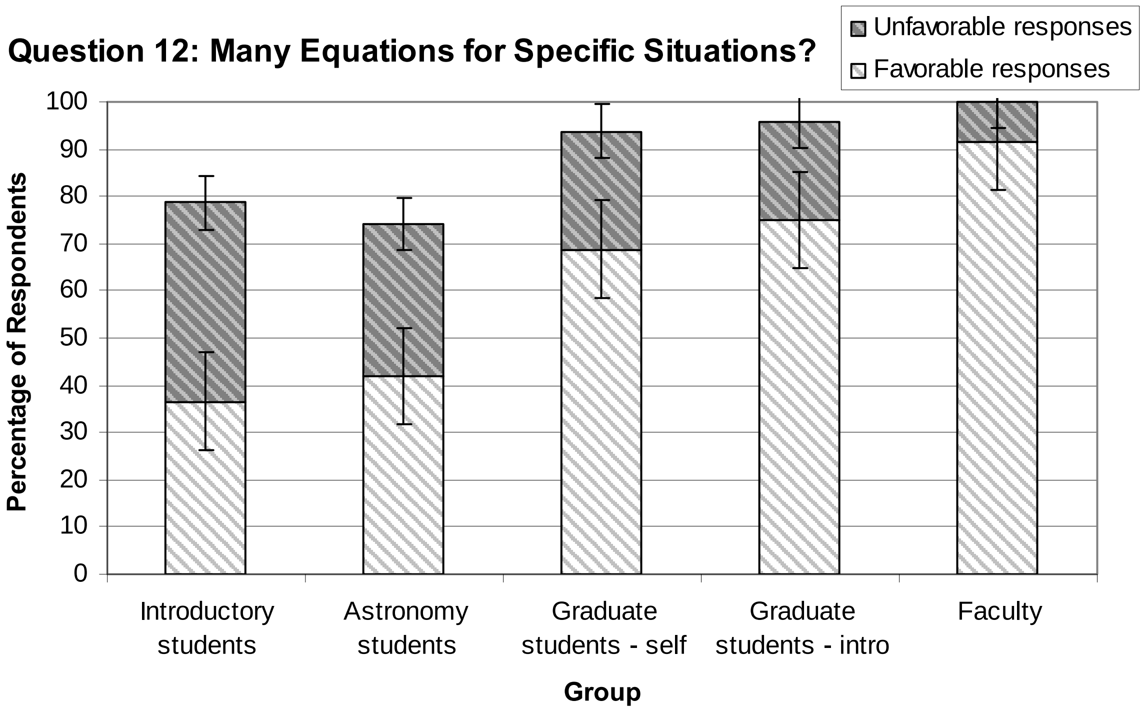


Figure 6.5. Histogram showing favorable (disagree) and unfavorable (agree) responses for survey question 12 about whether physics involves many equations that each apply primarily to a specific situation. As we go from the introductory students to the faculty, the disagreement with the statement increases.

expertise of each group and the fact that experts are more likely to discern the coherence of the knowledge in physics and appreciate how very few laws of physics are applicable in diverse situations and can explain different physical phenomena.

Figure 6.6 shows that, in response to question 14 regarding whether they always explicitly think about concepts that underlie the problems when solving physics problems, close to 90% of the graduate students agreed (favorable) that they do so both in the context of introductory and graduate level problem solving. However, only approximately 65% and 55% of the physics faculty and introductory physics students agreed, respectively. The non-monotonic nature of the responses in Figure 6.6 going from the introductory students to faculty may seem surprising at first, but individual discussions show that some faculty do not always explicitly

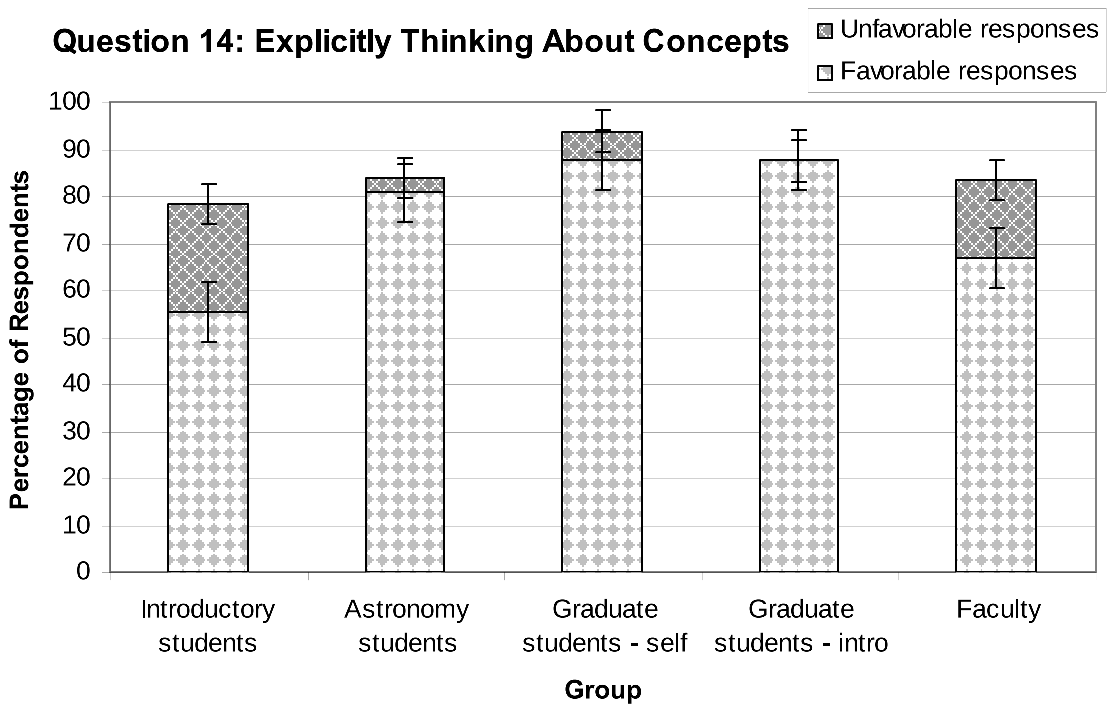


Figure 6.6. Histogram showing favorable (agree) and unfavorable (disagree) responses for survey question 14. While the non-monotonic trend in favorable responses from introductory students to faculty may seem surprising, some faculty noted that they do not explicitly think about concepts that underlie the problem while solving problems because the concepts have become obvious to them whereas introductory students often do not think about concepts because they believe in a plug and chug approach.

think about the concepts that underlie the problem because the concepts have become obvious to them due to their vast experience. They are able to invoke the relevant physics principles, e.g., conservation of mechanical energy or conservation of momentum, automatically when solving an introductory problem without making a conscious effort. In contrast, prior research shows that introductory physics students often do not explicitly think about the relevant concepts because they often consider physics as consisting of disconnected facts and formulas and associate physics problem solving as a task requiring looking for the relevant formulas without performing a conceptual analysis and planning of the problem solution (Redish et al. 1998, Adams et al. 2006). Thus, the reasoning behind the less favorable responses of faculty to question 14 is generally very different from the reasons behind the introductory physics students' responses.

Problem solving is often a missed learning opportunity because, in order to learn from problem solving, one must reflect upon the problem solution (Black and Wiliam 1998a, Black and Wiliam 1998b, Yerushalmi et al. 2007, Singh et al. 2007, Mason et al. 2009; see chapter 4). For example, one must ask questions such as "what did I learn from solving this problem?", "why did the use of one principle work and not the other one?" or "how will I know that the same principle should be applicable when I see another problem with a different physical situation?" Unfortunately, the survey results show a general lack of reflection by individuals in each group after solving problems.

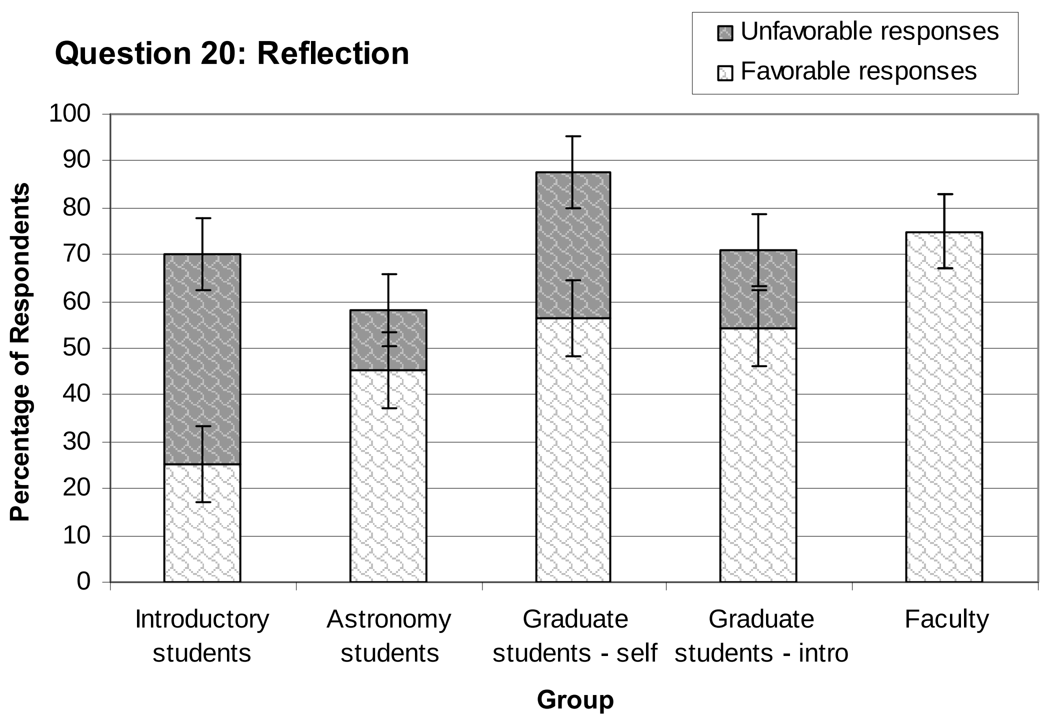


Figure 6.7. Histogram showing favorable (agree) and unfavorable (disagree) responses for survey question 20. The histogram shows that none of the groups had 80% individuals who agreed that they take the time to reflect and learn from the problem solutions after solving problems but the reasons for the lack of reflection varied across different groups.

Figure 6.7 shows that in response to question 20, only approximately 55% of the graduate students (for both introductory and advanced level problems) noted that, after they solve homework problems, they take the time to reflect and learn from the solution. The number of faculty members who noted that they take the time to reflect is close to 75%. Individual discussions show that, for introductory level problems, both physics faculty and graduate students felt that they monitor their thought processes while solving the problems since the problems are relatively simple. Therefore, reflection at the end of problem solving is not required. In contrast, while solving graduate level homework problems, some graduate students pointed to the lack of time for why they do not take the time to reflect after solving problems. Following are some explanations from graduate students for their responses:

- "If I have enough time, then I would like to reflect and learn from the problem solution after I struggle with it for a long time and then finally solve it successfully."
- "If the solution or the problem is interesting, then I would take time to reflect and learn from it. This usually happens in more challenging problems."
- "To be honest, I didn't do this when I was in college. But now I realized it's helpful."

Only approximately 25% of introductory physics students noted that they reflect and learn from problem solutions. Since reflection is so important for learning and building a robust knowledge structure, these findings suggest that instructors should consider giving students explicit incentive to reflect after they solve physics problems (Black and Wiliam 1998a, Black and Wiliam 1998b, Yerushalmi et al. 2007, Singh et al. 2007, Mason et al. 2009; also see chapter 4).

In response to question 25 about whether individuals make sure they learn from their mistakes and do not make the same mistakes again, all but one physics faculty members agreed with the statement (favorable) and one was neutral. On the other hand, only slightly more than 60% and 70% of the graduate students agreed with the statement when pertaining to solving graduate-level problems and introductory level problems, respectively. In individual discussions, some of the graduate students said that they do not have the time to learn from their mistakes. The response of introductory physics students to question 25 was comparable to that of graduate students. One introductory student said he did not review errors on the midterm exam as much as he would on homework, partly because the homework problems may show up on a future test but partly because he didn't like staring at his bad exam grade. The reluctance to reflect upon tests is consistent with our earlier findings for an upper-level undergraduate quantum mechanics course which demonstrated that many students did not reflect automatically on their mistakes in the midterm exams for similar reasons (Mason et al. 2009; see chapter 4).

Manipulation of symbols rather that numbers increases the difficulty of a problem for many introductory physics students. Question 30 asked whether symbolic problems were more difficult than identical problems with numerical answers and question 31 asked if individuals preferred to solve a problem with a numerical answer symbolically first and only plug in the numbers at the very end. Figures 6.8 and 6.9 show that the responses of graduate students for both introductory and graduate level problem solving are comparable to physics faculty but introductory students' responses are very different. Only approximately 35% of the introductory physics students disagreed with the statement that it is more difficult to solve a problem symbolically and 45% agreed with the statement that they prefer to solve the problem symbolically first and only plug in the numbers at the very end.

Individual discussions with some introductory physics students show that they have difficulty keeping track of the variables they are solving for if they have several symbols to

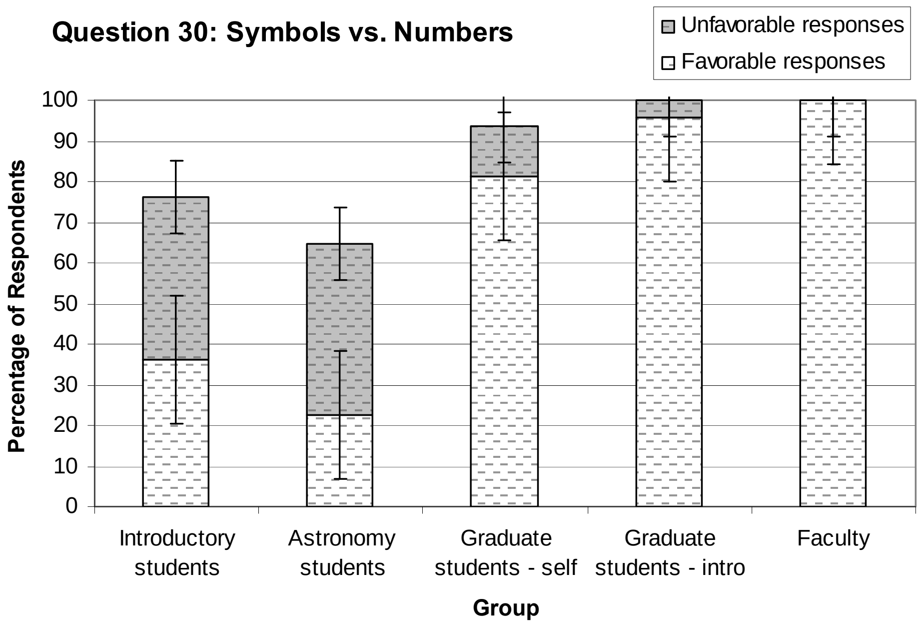


Figure 6.8. Histogram showing favorable (disagree) and unfavorable (agree) responses for survey question 30. The histogram shows that faculty and graduate students did not believe that it is more difficult to solve a physics problem with symbols than solving an identical problem with numerical answer but introductory physics and astronomy students often did.

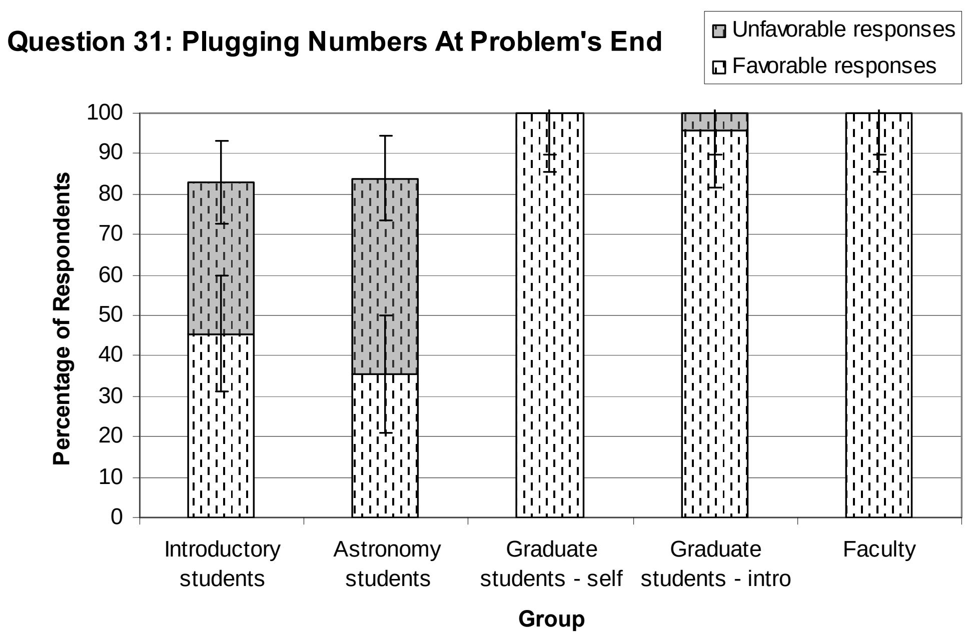


Figure 6.9. Histogram showing favorable (agree) and unfavorable (disagree) responses for survey question 31. The histogram shows that faculty and graduate students preferred to solve a problem symbolically first and only plug in the numbers at the very end but less than half of the introductory physics and astronomy students agreed with them.

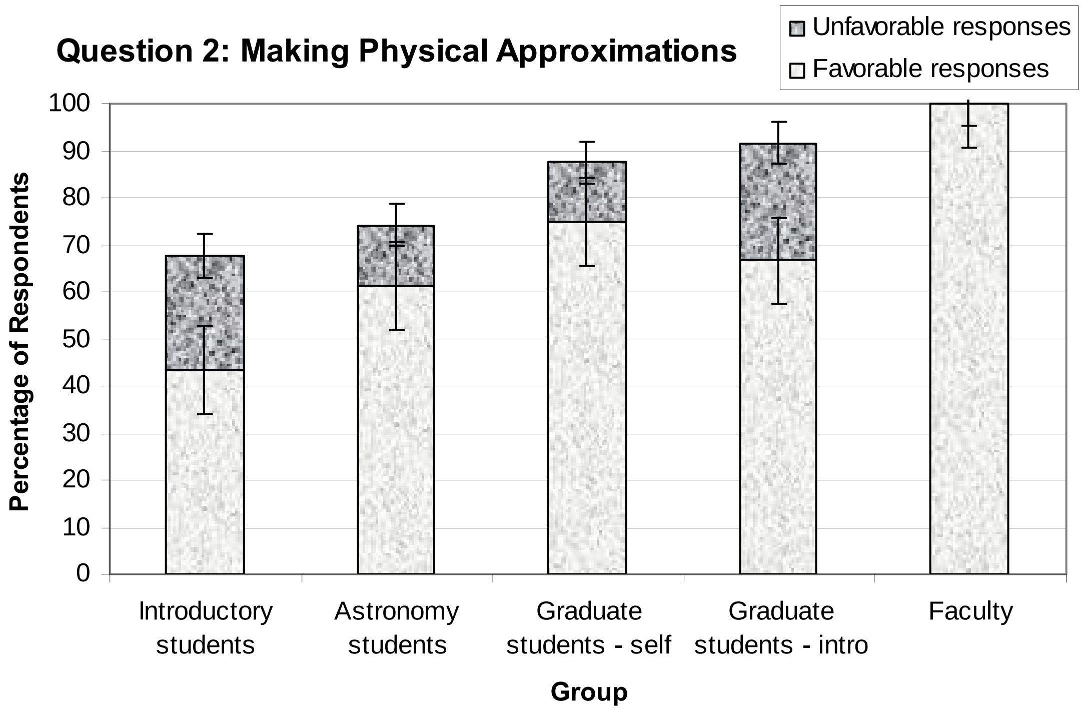


Figure 6.10. Histogram showing favorable (agree) and unfavorable (disagree) responses for survey question 2. The histogram shows that all faculty members agreed that they often make approximations about the physical world but other respondents, including physics graduate students, were not always in agreement.

consider, which motivates them to substitute numbers at the beginning of the solutions. One strategy to help introductory physics students feel more confident about using symbols is to ask them to underline the variable they are solving for so as to keep it from getting mixed with the other variables. Some introductory students noted that they did not like carrying expressions involving symbols from one equation to another because they were afraid that they will make mistakes in simplifying the expressions. Developing mathematical facility can help students develop the confidence to solve the problems symbolically first before substituting values. In addition, instructors should help students understand why it is useful to keep the symbols till the end, including the fact that it can allow them to check the solution, e.g., by checking the dimension, and it can also allow them to check the limiting cases which is important for developing confidence in one's solution.

Figure 6.10 shows that, in response to question 2 about whether they often make approximations about the physical world when solving physics problems, all faculty members noted that they do so. However, only approximately 75% and 65% of graduate students noted they do so for graduate level problem solving and introductory level problem solving, respectively. Individual discussions and written explanations suggest that the graduate students have different views on this issue as illustrated by the comments below:

- “I don't connect the physics problems to real world very much.”
- “…it's stat mech, in which I do whatever I have to [including approximations], to make the answer come out (and usually that is correct).”
- “Solving physics problems as an introductory physics student I was perhaps more prone to this, thinking about how a block would slide down an incline. As I became more familiar with the extent of “non-physical” approximations we made such as a frictionless world, I learned to separate problem solving space and real life space. I find that this is one aspect of physics problem solving that is harder for introductory level courses than graduate courses, the problems we solve [in introductory physics] are farther away from the physical world than graduate level problems. It keeps the math manageable and the physics concepts manageable but it makes them less intuitive.”
- “Many introductory-level problems are well defined and ideal, which doesn't require approx.”

On the other hand, individual discussions with some faculty showed that they considered the idealization of the problem in introductory and graduate level physics (e.g., framing problems without friction or air resistance, considering spherical cows or point masses, the infinite square well or the hydrogen atom with only the Coulomb force, etc.) as making approximations about the physical world and they felt that such approximations were helpful for getting an analytical answer and for building intuition about physical phenomena. It is possible that more than 25% of the graduate students who said that they don't make approximations had not carefully thought about the role of approximations about the physical world in physics problem solving.

6.5.3 Some questions may not have a “strong" expert-like response

If faculty members are taken as experts, their responses to the survey questions indicated that there were some questions whose answers may not necessarily represent favorable/unfavorable traits without disagreement from other faculty members. For example, in response to question (24), less than 70% of the faculty (and an even smaller percentage of graduate students) noted that they like to think through a difficult physics problem with a peer. Individual discussions with some of the faculty members suggested that the choice of whether one continues to persevere individually or works with a peer to solve a challenging problem depends on an individual's personality. The graduate students’ explanations for whether or not they preferred working with a peer were varied as illustrated by the following examples:

- “This is not true for introductory-level problems because, typically these types of problems are quite direct. Thinking about them for a little while always produces some result. For graduate-level problems, it is almost essential to work with others because of the complex line of thought it takes to solve some problems.”

- “Bouncing ideas with someone gives me sort of a chance to see the problem from the outside. You somehow see another point of attack. If you are stuck on a problem, usually the reason is that the approach is a dead end or too complex. Having someone to talk to forces you to think with a different perspective.”
- “I would like to think out by myself.”
- “As an introductory student it (working with peers) can make things more complicated, I would rather ask the TA. Other students have the same misconceptions as I do so they aren't a good source. As a graduate student, I saw that everyone was benefiting from collaboration. I know I would too. I just don't like to do anything with a peer but that's purely a social issue, I believe it is useful to work with peers."

Thus, while many graduate students agree that talking to peers is helpful, at least for challenging problems, some of them are inherently more averse to discussions with peers than others.

Next, we compare graduate students' responses on selected questions on the AAPS survey with those of the physics faculty and introductory physics and astronomy students. Figure 6.11 shows that, in response to question 3, regarding whether being able to handle the mathematics is the most important part of the process in solving a physics problem, less than 60% of the physics faculty were favorable and disagreed with the statement and approximately 35% were neutral. However, amongst graduate students, less than 40% of the students disagreed with the statement (favorable) both for problem solving at the introductory and graduate levels. Roughly 50% of the graduate students agreed (unfavorable) that mathematics is the most important part of the process in problem solving at the graduate level, and more than 40% agreed (unfavorable) with it for introductory level problem solving. Individual discussions with the graduate students show that in response to question (3) some students felt that facility with high-

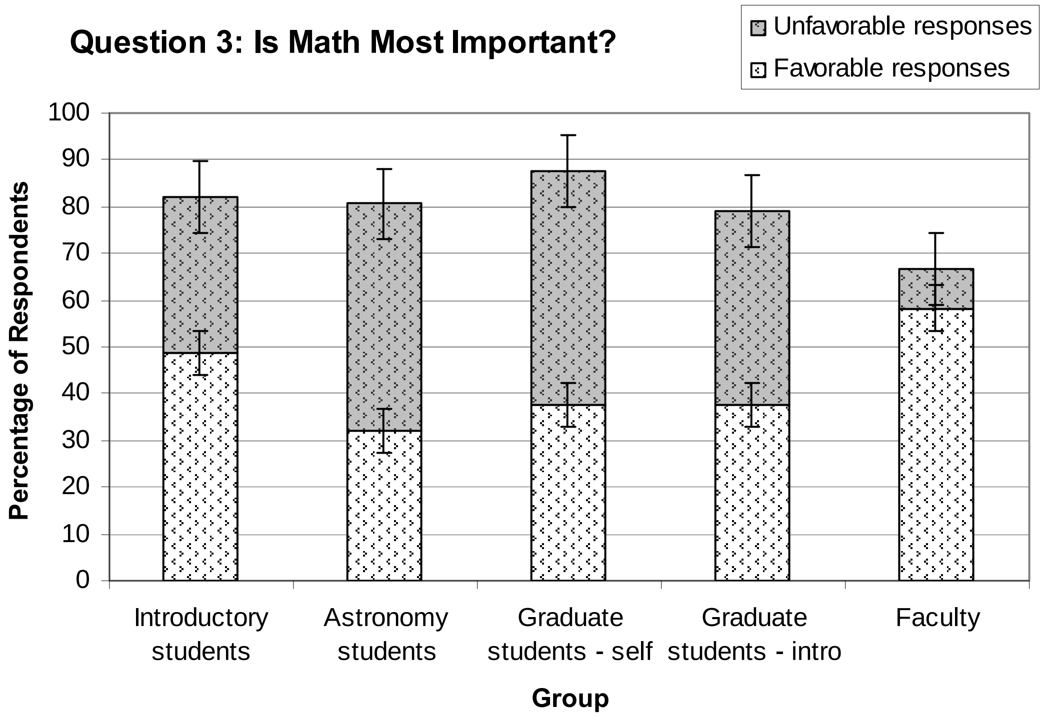


Figure 6.11. Histogram showing favorable (disagree) and unfavorable (agree) responses for survey question 3 about whether mathematics is the most important part of the problem solving process. The neutral percent responses can be found by subtracting from 100, the percentage of favorable and unfavorable responses. The histograms show that a large number of non-faculty respondents from all groups agreed with the statement. "Graduate students – self" and "Graduate Students – intro" refer to graduate students' response for graduate level problem solving and introductory level problem solving, respectively.

level mathematics is the most important skill for excelling in their graduate courses. Some graduate students felt that basic mathematics is very important for doing well in introductory problem solving as well, whereas others did not think mathematics was as important, especially for the algebra-based courses. The following are examples of responses from the graduate students that convey such sentiments:

- "If I was teaching a class of med. students, the concepts are certainly most important... However, a class of engineers really need to know how to get the right answer so the things that they build, function as they are supposed to. I would say in this case that math and concepts are equally important. For graduate-level problems, I believe mathematics becomes more essential over introductory-level problems.”

- “From my point of view, the introductory physics concepts are very easy to understand. It's in the details of problem solving that you could get stuck like encountering a difficult integration for example or some tricky algebra."
- “In introductory physics, this was not at all the case, the math was easy enough, I needed little more than high school calculus, so it was getting the physics down and understanding the language of physics, the new jargon, new concepts etc. Once the concepts became familiar enough and I moved on to graduate school, math became my biggest problem. From vector calculus to advanced linear algebra to special functions to group theory, the math is often harder. I find it a lot easier to think about physics and the universe conceptually (now that I am armed with such intuition and interest) but trying to actually ‘solve’ a physics problem comes down to the math, which I find hard.”

Individual discussions with some physics faculty about question 3 suggests that they believed that conceptual knowledge in physics was the central aspect of physics problem solving in both introductory and graduate level problem solving. But some faculty who were neutral in response to question 3 felt that the students may not excel in physics without a good grasp of mathematics even though concepts are very important. The views of more graduate students (compared to faculty) about mathematics being the most important aspect of physics problem solving may stem from the fact that graduate students have recently taken graduate and undergraduate level courses in which their grades often depend not as much on their conceptual knowledge but on their mathematical facility. However, question 3 is one of the survey questions for which several faculty members were not in good agreement and there isn't a strong agreement on expert-like response. Some faculty members even disliked having to agree or disagree with the question, stating that they felt both mathematics and physical concepts were too vital to problem solving to be compared to each other in such a way

Similarly, in response to question 16, only 75% of faculty (and an even smaller percentage for graduate students) noted that, while answering conceptual physics questions, they use the physics principles they usually think about when solving quantitative problems rather than mostly using their "gut" feeling. Discussions elucidated that the faculty members' use of their "gut" feeling to answer conceptual questions (rather than explicitly invoking physics principles) was often due to the fact that they had developed good intuition about the problems based upon their vast experience (Singh 2002). They did not need to explicitly think about the physical principles involved. Incidentally, 50% of the introductory physics students claimed that they use their "gut" feeling to answer conceptual questions rather than invoking physics principles. Our earlier research and those of others show that introductory students often view conceptual questions as guessing tasks and use their "gut" feeling rather than explicitly considering how the physical principles apply in those situations (Redish et al. 1998, Mazur 1997, Singh 2008). One interviewed introductory student stated that he would not consider principles when answering a conceptual question because over-analyzing the problem is more likely to make his answer wrong. When Mazur from Harvard University gave the Force Concept Inventory Conceptual standardized test (Hestenes et al. 1992) to his introductory students, a student asked if he should do it the way he really thinks about it or the way he has been taught to think about it (Mazur 1997). It appears that students sometimes hold two views simultaneously, one based upon their gut feeling and another is based upon what they learned in the physics class, and their views coexist and are difficult to merge.

6.6 SUMMARY AND CONCLUSIONS

We administered the "Attitudes and Approaches to Problem Solving" (AAPS) survey to physics graduate students, who answered the survey questions about problem solving in their graduate courses and in introductory physics. Their survey responses were also compared to those of introductory students in physics and astronomy courses and to physics faculty. We discussed the responses individually with some students and faculty members and obtained written explanations from some graduate students on selected questions.

There were major differences on some measures in graduate students' responses about problem solving in the graduate courses compared to problem solving in introductory physics. In general, graduate students' responses about problem solving in the graduate courses were less favorable (less expert-like) than their responses about solving introductory physics problems. For example, graduate students were more likely to feel stuck unless they got help while solving graduate level problems than on introductory level problems. Similarly, for problem solving in their graduate level courses, fewer graduate students could tell when their work and/or answer was wrong without talking to someone else but many more could tell when their solution was not correct when solving introductory physics problems. Also, more graduate students noted that they routinely use equations even if they are non-intuitive while solving graduate level problems than while solving introductory physics problems. In addition, fewer graduate students noted that they enjoy solving challenging graduate level physics problems than solving challenging introductory physics problems (perhaps because introductory physics problems are more doable for them).

Comparison of graduate students' responses with faculty responses shows that, on many measures, graduate students' responses to the AAPS survey are less favorable than faculty responses. For example, unlike the graduate students, all physics faculty members noted that they enjoy solving challenging physics problems. The less favorable response of graduate students while solving graduate level problems is partly due to the fact that the graduate students are not yet experts in their own graduate course content. Due to lower expertise in solving graduate level problems, graduate students are more likely to feel stuck unless they get help, not know whether their solution is right or wrong, use equations that are not intuitive and not enjoy solving challenging graduate level problems on which their grade depends and for which they have a limited time to solve.

We find that, on some survey questions, the graduate students' and faculty members' responses to the survey questions must be interpreted carefully. For example, only two thirds of the faculty noted that they always think about the concepts that underlie the problem explicitly while solving problems, which is lower than the fraction of graduate students who noted that they do so while solving both introductory level and graduate level problems. Individual discussions with faculty members shows that they felt that, after years of teaching experience, the concepts that underlie many of the problems have become "automatic" for them and they do not need to explicitly think about them. The fact that, in contrast to most faculty, many graduate students always think explicitly about the concepts that underlie the problems both while solving introductory and graduate level problems suggests that the graduate students have not developed the same level of expertise and efficiency in solving introductory level problems as physics faculty have.

Comparison of graduate students' AAPS responses with introductory physics students' responses shows that, on some measures, graduate students have more favorable attitudes and approaches to solving introductory physics problems due to their higher level of expertise than the introductory students. However, on other questions, the responses must be interpreted carefully in light of the explanations provided by the graduate students. For example, in response to whether the problem solving in physics is essentially “plug and chug,” the response of the graduate students while solving introductory physics problems and those of introductory physics students is indistinguishable. However, interviews and written responses from the graduate students shows that they have developed sufficient expertise in introductory physics so that solving such problems does not require much explicit thought and they can often immediately tell which principle of physics is applicable in a particular situation. On the other hand, prior research show that many introductory physics students forgo the conceptual analysis and planning of the problem solution and immediately look for the formulas (Larkin and Reif 1979, Reif 1986, Reif 1995).

Also, due to their higher level of expertise, graduate students find introductory physics equations more intuitive and are better able to discern the applicability of a physics principle epitomized in the form of a mathematical equation to diverse situations than the introductory students. In solving both introductory and graduate level problems, the fraction of graduate students who noted that they reflect and learn from the problem solution after solving a problem is significantly larger than the fraction of introductory physics students who noted doing so in their courses. While we may desire a higher percentage of graduate students to reflect and learn from their problem solving, interviews suggest that, in the graduate courses, students felt they did not have the time to reflect. Also, they often did not reflect on the exam solutions even after they received the solutions because they did not expect those problems to show up again on another exam. Some graduate students explained that the reason they do not reflect after solving an introductory physics problem is that the solutions to those problems are obvious to them and do not require reflection. There was a large difference between the introductory physics students' and graduate students' responses in their facility to manipulate symbols (vs. numbers) with introductory physics students finding it more difficult to solve problems given in symbolic form. In problems where numbers were provided, many introductory students noted that they prefer to plug numbers at the beginning rather than waiting till the end to do so.

In general, the more favorable responses of graduate students on the AAPS survey towards attitudes and approaches to introductory problem solving compared to those of the introductory physics students and less favorable responses compared to the faculty imply that graduate students have a higher level of expertise in introductory physics but less expertise than physics faculty. Moreover, graduate students' responses to graduate level problem solving in many instances are comparable to introductory students' responses to introductory level problem solving, implying that the graduate students are still developing expertise in their own graduate level courses. As noted earlier, the survey results also show that many graduate students are more likely to enjoy solving difficult introductory physics problems than graduate level problems, they are more likely to feel stuck while solving graduate level problems, less likely to find graduate level equations intuitive (but they still use them freely to solve problems), less likely to predict whether their problem solution is correct and to not give a high priority to learning after solving a problem. While one can rationalize these unfavorable responses of graduate students to graduate level problem solving by claiming that these are reflections of the fact that they are not “experts” in graduate level courses, they force us to think about whether we are achieving the goals of the graduate courses and giving graduate students an opportunity to learn effective approaches and attitudes to problem solving. Graduate instructors should consider whether assessment in those courses should include both quantitative and conceptual questions to motivate students to reflect on problem solving and to help them develop intuition about the equations underlying the problems.

6.7 ACKNOWLEDGMENTS

We thank F. Reif, R. Glaser, R. P. Devaty and J. Levy for very useful discussions. We thank the faculty who administered the survey to their class and took the survey for their help.

6.8 REFERENCES

Adams, W., Perkins, K., Podolefsky, N., Dubson, M., Finkelstein, N., & Wieman, C. (2006). "New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey." *Phys. Rev. ST Phys. Educ. Res.* 2, 010101.

Black, P., & Wiliam, D. (1998a). "Assessment and classroom learning." *Assessment in Education* 5(1), 7-74.

Black, P., & Wiliam, D. (1998b). "Inside the black box: Raising standards through classroom assessment." *Phi Delta Kappan* 80(2), 139-148.

Collins, A., Brown, J., & Newman, S. (1989). "Cognitive Apprenticeship: Teaching the crafts of reading, writing and mathematics." In *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Edited by L. B. Resnick. Hillsdale, NJ: Lawrence Erlbaum, 453-494.

Cummings, K., Lockwood, S., & Marx, J. (2004). "Attitudes Toward Problem Solving as Predictors of Student Success." In *2003 Physics Education Research Conference Proceedings*. Edited by J. Marx, S. Franklin, and K. Cummings. *AIP Conf. Proc.* 720, 133-136.

Elby, E. (2001). "Helping physics students learn how to learn." *Am. J. Phys.* 69, S54-S64.

Gray, K., Adams, W., Wieman, C., & Perkins, K. (2008). "Students know what physicists believe, but they don't agree: A study using CLASS survey." *Phys. Rev. St. Phys. Educ. Res*. 4, 020106.

Halloun, I. (1996). "Views about science and physics achievement: the VASS story." In *Proceedings of the International Conference on Undergraduate physics education,* College Park, MD. Available at http://www.eric.ed.gov (available on http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content storage 01/0000019b/80/14/af/e7.pdf).

Hestenes, D., Wells, M., & Swackhamer, G. (1992). "Force Concept Inventory." *Phys. Teach.* 30, 141.

Schommer, M. (1990). "Effects of beliefs about the nature of knowledge on comprehension." *J. Ed. Psych.* 82(3), 406-411.

Redish, E., Saul, J., & Steinberg, R. (1998). "Student Expectations in Introductory Physics." *Am. J. Phys.* 66, 212-224.

Larkin, J., & Reif, F. (1979). "Understanding and teaching problem solving in physics." *Eur. J. Sci. Educ.* 1(2), 191-203.

Laws, P. (1999). *Workshop Physics.* NY: Wiley.

Marx, J., & Cummings, K. (2007). "What factors really influence shifts in students' attitudes and expectations in an introductory physics course?" *In 2006 Physics Education Research Conference Proceedings, Syracuse, N.Y.* Edited by L. McCullough, L. Hsu, and P. Heron. *AIP Conf. Proc.* 883, 101-104.

Mason, A., & Singh, C. (2009). "Do advanced students learn from their mistakes without explicit intervention?" Submitted; also see chapter 4.

Mazur, E. (1997). *Peer Instruction: A User's Manual.* Series title: *Prentice Hall series in educational innovation.* Upper Saddle River, Prentice Hall, N.J., and video that accompanies the book.

Reif, F. (1986). "Scientific approaches to science education." *Phys. Today* 39(11), 48-54.

Reif, F. (1995). "Millikan Lecture 1994: Understanding and teaching important scientific thought processes." *Am. J. Phys.* 63, 17.

Schoenfeld, A. (1985). *Mathematical Problem Solving*. NY: Academic Press.

Schoenfeld, A. (1989). "Teaching mathematical thinking and problem solving." In *Toward the Thinking Curriculum: Current Cognitive Research.* Edited by L. B. Resnick and B. L. Klopfer. Washington, DC: ASCD, pp. 83-103.

Schoenfeld, A. (1992). "Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics." In *Handbook for Research on Mathematics Teaching and Learning.* Edited by D. Grouws. NY: McMillan, Chap. 15, pp. 334-370.

Singh, C. (2002). "When physical intuition fails." *Am. J. Phys.* 70(11), 1103-1109.

Singh, C. (2008). "Assessing student expertise in introductory physics with isomorphic problems, Part II: Examining the effect of some potential factors on problem solving and transfer." *Phys. Rev. ST Phys. Educ. Res.* 4, 010105(1-10).

Singh, C., Yerushalmi, E., & Eylon, B. (2007). "Physics Learning in the context of scaffolded diagnostic tasks (II): The preliminary results." In *Proceedings of the Phys. Educ. Res. Conference, Greensboro, NC.* Edited by L. Hsu, L. McCullough, and C. Henderson. *AIP Conf. Proc.* 951, 31-34.

Yerushalmi, E., Singh, C., and Eylon, B. (2007). "Physics Learning in the context of scaffolded diagnostic tasks (I): The experimental setup." In *Proceedings of the Phys. Ed. Res. Conference, Greensboro, NC.* Edited by L. Hsu, L. McCullough, and C. Henderson. *AIP Conf. Proc.* 951, 27-30.

7.0 CONCLUSIONS AND FUTURE CONSIDRATIONS

In this thesis, we discuss five investigations designed to shed light on the role of reflection in problem solving by students in introductory and advanced level physics courses. A major goal of many physics courses at different levels is to make students good problem solvers and independent learners and activities that require students to reflect about problem solving can help students develop these skills. Therefore, reflection upon problem solving should be incorporated as an integral component of all physics courses in different contexts.

7.1 INTRODUCTORY PHYSICS

7.1.1 Categorization Study

The categorization study requiring students to group together problems based upon similarity of solution suggests that there is a large overlap between the expertise of physics graduate students and introductory calculus-based students with regard to their ability to categorize problems. Our study which was carried out in the classroom environment therefore suggests that unlike the conclusion of Chi et al. (1981 – see section 2.7 for reference), there is a large diversity in the expertise of physics graduate students and introductory calculus-based students and it is not appropriate to call all introductory physics students “novices” and all physics graduate students “experts”.

The categorization study also shows that introductory physics students in the calculus-based courses perform significantly more expert-like categorizations than those in the algebra-based courses. This may initially seem puzzling because, while the calculus-based students have better mathematical skills, the categorization task is conceptual in nature. Moreover, the algebra-based students are typically aspiring for careers in health-care professions and they are generally more motivated to learn than those in the calculus-based courses who are mostly engineers. One possible reason for this difference, which is supported by prior research, is that the calculus-based students have better scientific thinking skills. While learning physics concepts, the calculus-based introductory students are less likely to have cognitive overload than the algebra-based introductory physics students and hence they may learn better and have more organized and robust knowledge structure. These differences can affect their ability to group together problems based upon similarity of solution.

Future studies can focus on investigating the reasons for the difference between the calculus-based and algebra-based students’ categorization in more depth. It will also be useful to investigate how the graduate students and introductory physics students perform when asked to outline the problem solutions that go beyond students’ ability to categorize.

7.1.2 Approaches and attitudes towards problem solving

The survey related to approaches and attitudes towards problem solving shows that physics professors’ approaches and attitudes on average are most expert-like, followed by the graduate students and introductory physics students. There wasn't a major difference between the calculus-based and algebra-based students with regard to their approaches and attitudes to problem solving. When graduate students expressed their approaches and attitudes towards problem solving while solving introductory physics problems compared to graduate level physics problems, in many cases their approaches and attitudes were more novice-like while solving graduate-level physics problems than while solving introductory level physics problems. Graduate instructors should reflect upon these findings and the goals of the graduate physics courses and modify the graduate curriculum accordingly.

Several approaches and attitudes were unexpected especially from graduate students. For example, more graduate students agreed (than disagreed) with the statement that "being able to handle the mathematics is the most important part of problem solving process" both while solving problems for their own graduate courses and introductory physics problems. Moreover, the survey shows that a large fraction of graduate students and introductory physics students do not take time to reflect on the mistakes in their homework and examination solutions. This finding suggests that it may be useful to explicitly ask students to diagnose mistakes in their homework and exams in order to learn from them.

7.1.3 Self-diagnosis and Peer Reflection

The in-recitation self-diagnosis and peer reflection studies indicate that students in the algebra-based introductory courses can benefit from reflection on their problem solving in both an individual and group exercise. Of course, the amount of scaffolding support that students receive should be commensurate with their prior knowledge and skills because the guidance and feedback can strongly influence how students benefit from these activities. In the case of more difficult problems, more structured and active scaffolding is necessary to learn self-diagnosis skills and benefit from diagnostic activities. In the study involving reflection with peers, we found that those in the “peer reflection” group drew more diagrams and gained more on the final exam compared to the midterm exams.

Future studies on self-diagnosis can focus on providing more sustained intervention and explore the level of scaffolding necessary to help students become better at self-diagnosis. A more consistent and sufficiently structured exercise in self-diagnosis throughout the semester will be a more effective strategy for helping students develop higher order learning skills and habits of mind. Further studies on self-diagnosis could determine more clearly how the difficulty of the problem that students self-diagnose, the level of difference between the transfer problem and the self-diagnosed problem, and the level of scaffolding each influence students' performance on transfer tasks. In the case of reflection with peers, strategies for more active participation by students who are in the bottom one fourth of the class should be explored in future research.

7.2 DO STUDENTS AUTOMATICALLY REFLECT IN ADVANCED COURSES?

The study in advanced quantum mechanics discussed in this thesis suggests that even the more experienced students do not necessarily reflect on their mistakes in the midterm exams automatically. It is therefore necessary to consider whether implementation of explicit reflection tasks for advanced students may be helpful. The fact that certain subjects like quantum mechanics are mostly novel to upper-level students also suggests that some cognitive apprenticeship support may be necessary that takes into account the deeper theoretical and mathematical structure of advanced problems and their solutions.

7.3 OTHER FUTURE CONSIDERATIONS

Overall, the studies describe a cross-section of student reflection on problem-solving. Reflection has potential to be implemented in a physics curriculum at all levels of instruction, e.g., students at all levels can be explicitly rewarded and prompted to consider errors in their work using the various strategies described in this thesis.

It should be noted that if students perform well on one reflection task, e.g., self-diagnosis, it does not automatically imply deep cognitive engagement. Therefore, there may not be a good correlation between how well students perform on self-diagnosis and another reflection task, e.g., categorization. As discussed in the self-diagnosis study, self-diagnosis could be superficial or meaningful. If the self-diagnosis is superficial, the score on self-diagnosis will not reflect actual learning from the diagnostic activities. Assessment tools need to be built so that achievement is understood beyond the superficial level. Future studies should examine the correlation between students' performance on different reflection tasks.

APPENDIX A

CHAPTER 2: CATEGORIZATION QUESTIONNAIRES

A.1 VERSION I OF THE QUESTIONNAIRE

This version of the survey contains 25 questions given to all groups in the categorization except for one of the algebra-based introductory physics sections.

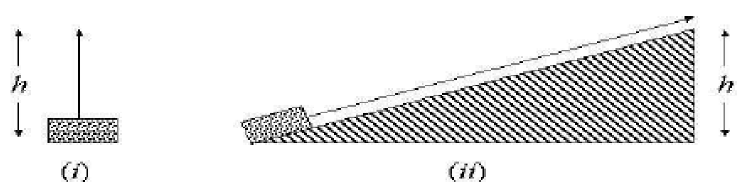
Instructions

• Your task is to group the 25 problems below based upon similarity of solution into various groups on the sheet of paper provided. Problems that you consider to be similar should be placed in the same group. You can create as many groups as you wish. The grouping of problems should NOT be in terms of "easy problems", "medium difficulty problems" and "difficult problems" but rather it should be based upon the features and characteristics of the problems that make them similar. A problem can be placed in more than one group created by you. Please provide a brief explanation for why you placed a set of questions in a particular group. You need NOT solve any problems.
• Ignore the retarding effects of friction and air resistance unless otherwise stated.

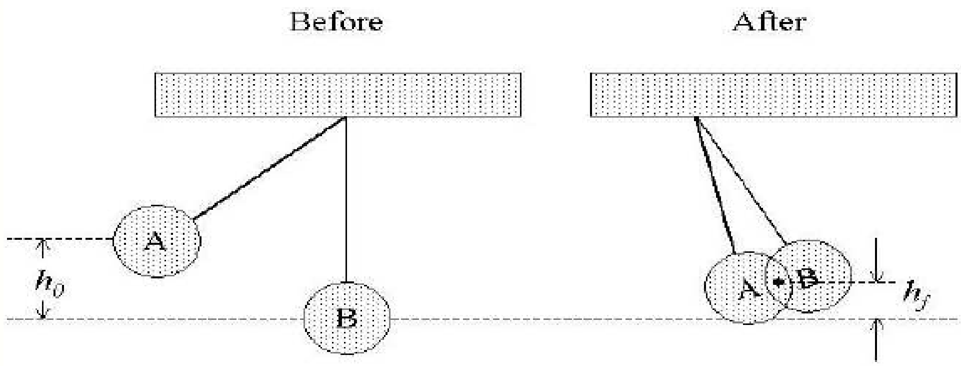
1. Harry Potter and Voldemort are wrestling inside a cart traveling east at a speed of 45 m/s directly toward an abyss. Harry then notices the danger and jumps backward due west off the cart. Ron who stands on safe ground in the back notices that Harry's velocity due west at the jump is 15 m/s relative to the ground. What is the speed of the cart after Harry jumps off it? The mass of the cart is 200 kg, Harry's mass is 60 kg, and Voldemort's mass is 80 kg.

2. Two identical stones, A and B, are shot from a cliff from the same height h and with identical initial speeds v_0. Stone A is shot vertically up, and stone B is shot vertically down (see Figure). Which stone has a larger speed right before it hits the ground?

3. You want to lift a heavy block through a height h by attaching a string of negligible mass to it and pulling so that it moves at a <u>constant speed</u> v. You have the choice of lifting it either by pulling the string vertically upward or along a frictionless inclined plane (see Figure). How much is the work done by the gravitational force in the two cases?

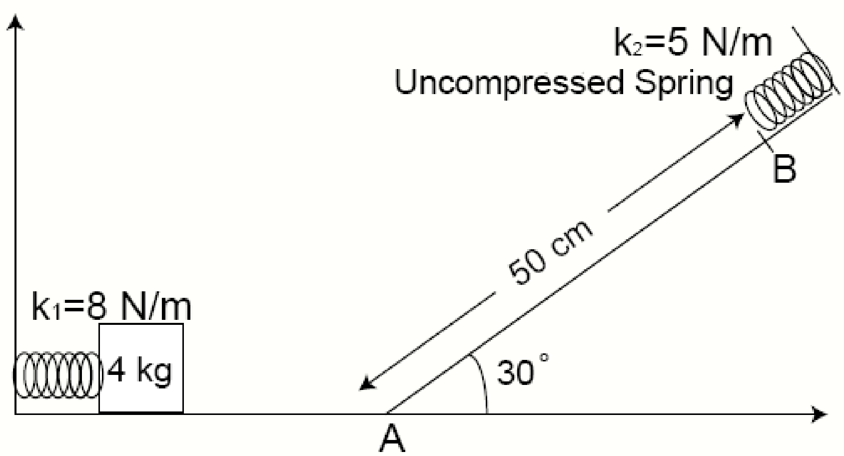


4. Two small spheres of putty, A and B, of equal mass, hang from the ceiling on massless strings of equal length. Sphere A is raised to a height h_0 as shown below and released. It collides with sphere B (which is initially at rest); they stick and swing together to a maximum height h_f. Find the height h_f in terms of h_0.

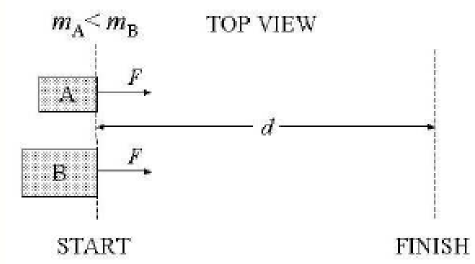


5. A family decides to create a tire swing in their backyard for their son Ryan. They tie a nylon rope to a branch that is located 16 m above the earth, and adjust it so that the tire swings 1 meter above the ground. To make the ride more exciting, they construct a launch point that is 13 m above the ground, so that they don't have to push Ryan all the time. You are their neighbor, and you are concerned that the ride might not be safe, so you calculate the maximum tension in the rope to see if it will hold. Calculate the maximum tension in the rope, assuming that Ryan (mass 30 kg) starts from rest from his launch pad. Is it greater than the maximum rated value of 2500 N?

6. In the figure below, a horizontal spring with spring constant $k_1 = 8$ N/m is compressed 20 cm from its equilibrium position by a 4 kg block. Then, the block is released. What would be the maximum compression of a spring ($k_2 = 5$ N/m) on the inclined plane when the 4 kg block presses against it? Assume that the track is frictionless and the distance from A to B is 50 cm where B is the edge of the uncompressed spring on the inclined plane.



7. Two blocks are initially at rest on a frictionless horizontal surface. The mass m_A of block A is less than the mass m_B of block B. You apply the same constant force F and pull the blocks through the same distance d along a straight line as shown below (force F is applied for the entire distance d). Compare the speed of the blocks after you pull them the same distance d.

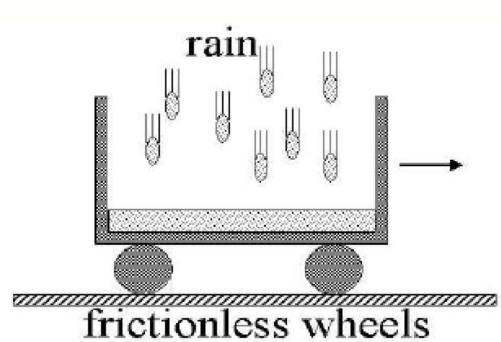


8. Your friend Dan, who is in a ski resort, competes with his twin brother Sam on who can glide higher with the snowboard. Sam, whose mass is 60 kg, puts his 15 kg snowboard on a level section of the track, 5 meters from a slope (inclined plane). Then, Sam takes a running start and jumps onto the stationary snowboard. Sam and the snowboard glide together till they come to rest at a height of 1.8 m above the starting level. What is the minimum speed at which Dan should run to glide higher than his brother to win the competition? Dan has the same weight as Sam and his snowboard weighs the same as Sam's snowboard.

9. Two identical stones, A and B, are shot from a cliff from the same height h and with identical initial speeds v_0. Stone A is shot at an angle of 30^0 above the horizontal and stone B is shot at an angle of 30^0 below the horizontal. Which stone takes a longer time to hit the ground?

10. At amusement parks, there is a popular ride in which the floor of a rotating cylindrical room falls away, leaving the backs of the riders "plastered" against the wall. Suppose the radius of the room is 3.3 m and the speed of the wall is 10 m/s when the floor falls away. What is the minimum coefficient of friction that must exist between a rider's back and the wall, if the rider is to remain in place when the floor drops away?

11. Rain starts falling vertically down into a cart (of mass M) with frictionless wheels which is initially moving at a constant speed V on a horizontal surface. The rain drops fall on the car with a speed v and come to rest with respect to the cart after striking it. Find the speed of the cart when m grams of rain water accumulate in the cart.



12. In the track shown below, section AB is a quadrant of a circle of 1 m radius. A block is released at A and slides without friction until it reaches point B. The horizontal part is not smooth. If the block comes to rest 3 m from B, what is the coefficient of kinetic friction?

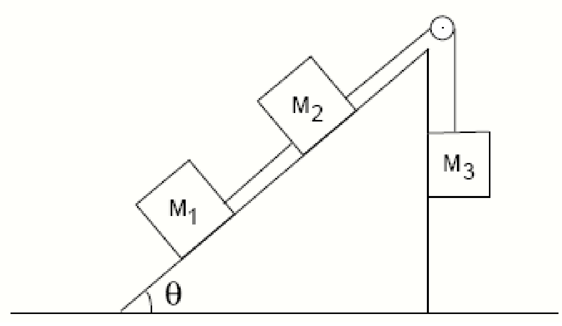
13. Three blocks (m_1=1 kg, m_2=2 kg, m_3=3kg) are in a straight line in contact with each other on a frictionless horizontal table (block with mass m_2 is in the middle). A constant horizontal force F_H=3 N is applied to the block with mass m_1. Find the forces exerted on m_1 by m_2 and on m_2 by m_3.

14. A ball is thrown from the top of a 35 m high building with an initial speed of 80 m/s at an angle of 25^0 above the horizontal. Find the time it takes to reach the ground.

15. A cyclist approaches the bottom of a gradual hill at a speed of 15 m/s. The hill is 5 m high, and the cyclist estimates that she is going fast enough to coast up and over it without peddling. Ignoring friction and air resistance, find the speed at which the cyclist crests the hill? Neglect the kinetic energy of the rotating wheels.

16. A slingshot fires a pebble from the top of a building at a speed of 10 m/s. The building is 20 m tall. Ignoring air resistance, find the speed with which the pebble strikes the ground when the pebble is fired (I) horizontally, (II) vertically straight up.

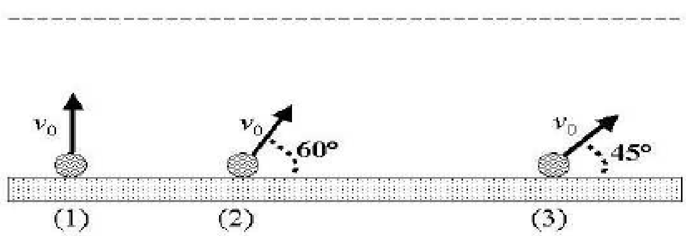
17. The figure below shows two blocks on a frictionless inclined plane with an angle of inclination $\theta = 40^0$ and the two connected to each other via a massless rope. The rope that connects the two blocks goes around a frictionless massless pulley and is connected to a third block as shown. Find the magnitude of the tension force in the rope between blocks with mass M_1 and M_2 and the acceleration of the blocks.



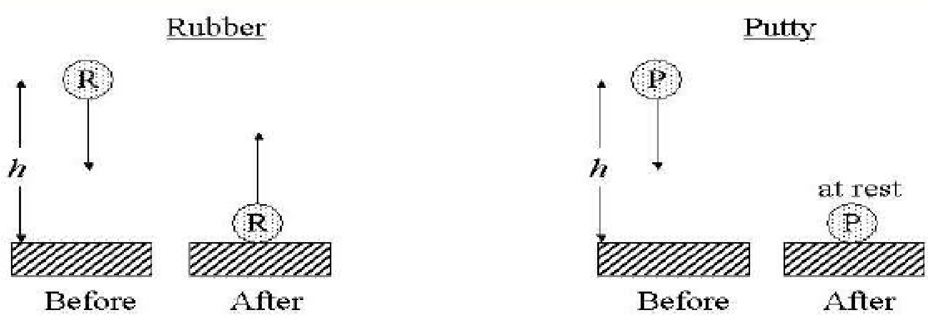
18. Two frictionless inclined planes have the same height but have different angles of inclinations of 45^0 and 60^0 with respect to the horizontal. You slide down from the top which is at a height h above the ground on each inclined planes starting from rest. Find your speed at the bottom of the inclined planes in the two cases.

19. The brakes of your bicycle have failed, and you must choose between slamming into either a haystack or a concrete wall. Explain why hitting a haystack is a wiser choice than hitting a concrete wall.

20. Three balls are launched from the same horizontal level with identical speeds v_0 as shown below. Ball (1) is launched vertically upward, ball (2) at an angle of 60^0, and ball (3) at an angle of 45^0. In order of decreasing speed (fastest first), rank the speed each one attains when it reaches the level of the dashed horizontal line. All three balls have sufficient speed to reach the dashed line.

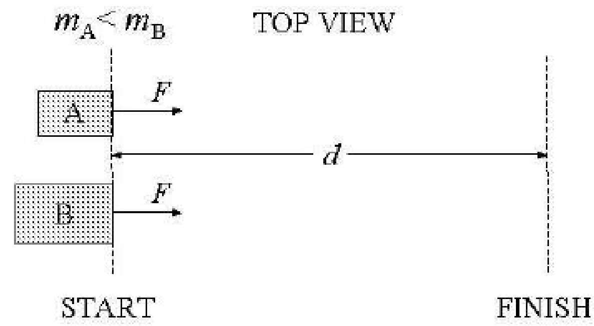


21. You drop two balls of equal mass, made of rubber and putty, from the same height h above a horizontal surface (see Figure). The rubber ball bounces up after it strikes the surface while the putty ball comes to rest after striking it. Assume that in both cases the velocity of the ball takes the same time Δt to change from its initial to its final value due to contact with the surface. During time Δt, which of the average forces $\overline{F}_R$ or $\overline{F}_P$ exerted on the surface by the rubber and putty balls, respectively, is greater?



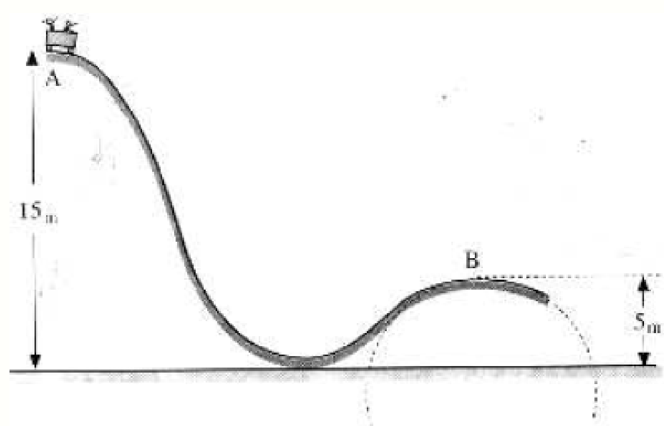
22. Two frictionless inclined planes have the same height but have different angles of inclinations of 45^0 and 60^0 with respect to the horizontal. You slide down from the top which is at a height h above the ground on each inclined planes starting from rest. Find the time taken to reach the bottom in the two cases.

23. Two blocks are initially at rest on a frictionless horizontal surface. The mass m_A of block A is less than the mass m_B of block B. You apply the same constant force F and pull the blocks through the same distance d along a straight line as shown below (force F is applied for the entire distance d). Rank the time taken to pull the two blocks by the same distance d.



24. You are standing at the top of an incline with your skateboard. After you skate down the incline, you decide to "abort", kicking the skateboard out in front of you such that you remain stationary afterwards. How fast is the skateboard travelling with respect to the ground after you have kicked it? Assume that your mass is 60 kg, the mass of the skateboard is 10 kg, and the height of the incline is 10 cm.

25. A friend told a girl that he had heard that if you sit on a scale while riding a roller coaster, the dial on the scale changes all the time. The girl decides to check the story and takes a bathroom scale to the amusement park. There she receives an illustration (see below), depicting the riding track of a roller coaster car along with information on the track (the illustration scale is not accurate). The operator of the ride informs her that the rail track is smooth, the mass of the car is 120 kg, and that the car sets in motion from a rest position at the height of 15 m. He adds that point B is at 5m height and that close to point B the track is part of a circle with a radius of 30 m. Before leaving the house, the girl stepped on the scale which indicated 55 kg (the scale is designed to be used on earth and displays the mass of the object placed on it). In the rollercoaster car the girl sits on the scale. According to your calculation, what will the scale show at point B?



A.2 VERSION II OF THE QUESTIONNAIRE

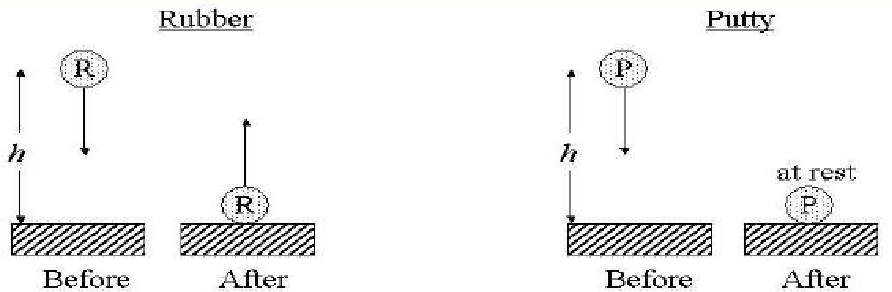
This version replaces 10 problems with the 7 problems that could be recovered from the original Chi et al. (1981) study and 3 problems that had to do with conservation of angular momentum. In addition, the other 15 problems were rearranged in order. This version of the questionnaire was given to one set of algebra-based introductory physics students

Instructions

• Your task is to place the 25 problems below into various groups based upon similarity of solution on the sheet of paper provided. Problems that you would solve in a similar way should be placed in the same group. You can create as many groups as you wish. The grouping of problems should NOT be in terms of "easy problems", "medium difficulty problems" and "difficult problems" but rather it should be based upon the features and characteristics of the problems that make them similar. A problem can be placed in more than one group created by you. Please provide a brief explanation for why you placed a set of questions in a particular group. You need not solve any problems.

• Ignore the retarding effects of friction and air resistance unless otherwise stated.

1. You drop two balls of equal mass, made of rubber and putty, from the same height h above a horizontal surface (see Figure). The rubber ball bounces up after it strikes the surface while the putty ball comes to rest after striking it. Assume that in both cases the velocity of the ball takes the same time Δt to change from its initial to its final value due to contact with the surface. Compare the average forces $\overline{F}_R$ and $\overline{F}_P$ exerted on the surface by the rubber and putty balls, respectively, during time Δt.



2. An ice skater is spinning on essentially frictionless ice with her arms extended. Then, she pulls her arms in close to her body, cutting her moment of inertia in half. There are no net external forces or torques on her. How will her angular speed change when she puts her arms in?

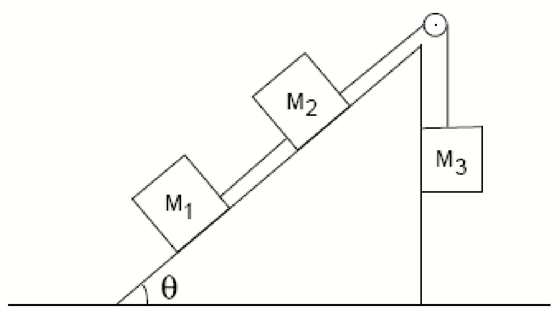
3. Your friend Dan, who is on vacation in a ski resort, competes with his twin brother Sam on which of them can glide higher with their snowboards. Sam, who weighs 60 kg, puts his 15 kg snowboard on a level section of track, 5 meters far from a slope (inclined plane), then he takes a running start and jumps onto the stationary snowboard. Sam and the snowboard glide together till they come to rest at a height of 1.8 m above the starting level (as determined by the height signs attached to the slope). What is the minimum speed at which Dan should run to glide higher than his brother if he has a chance to win the competition? Dan has the same weight as Sam.

4. Three blocks (m_1=1 kg, m_2=2 kg, m_3=3kg) are in a straight line in contact with each other on a frictionless horizontal table (block with mass m_2 is in the middle). A constant horizontal force F_H=3 N is applied to block with mass m_1. Find the forces exerted on m_1 by m_2 and on m_2 by m_3.

5. A ball is thrown from the top of a 35 m high building with an initial speed of 80 m/s at an angle of 25^0 above the horizontal. Find the time it takes to reach the ground.

6. A cylist approaches the bottom of a gradual hill at a speed of 15 m/s. The hill is 5 m high, and the cyclist estimates that she is going fast enough to coast up and over it without peddling. Ignoring friction and air resistance, find the speed at which the cyclist crests the hill.

7. The figure below shows two blocks on a frictionless inclined plane with an angle of inclination $\theta = 20^0$ and the two connected to each other via a massless rope. The chord that connects the two blocks goes around a frictionless massless pulley and is connected to a third block as shown. Find the magnitude of the tension force and the acceleration of the blocks.



8. The brakes of your bicycle have failed, and you must choose between slamming into either a haystack or a concrete wall. Explain why hitting a haystack is a wiser choice than hitting a concrete wall.

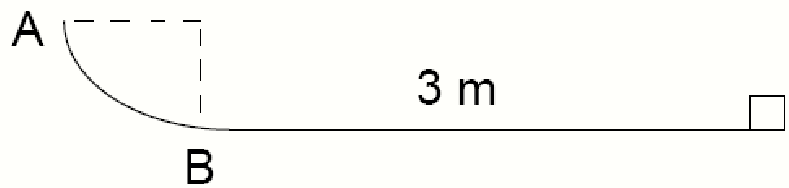
9. You are standing at the top of an incline with your skateboard. After you roll down the incline, you decide to "abort", kicking the skateboard out in front of you such that you remain stationary afterwards. How fast is the skateboard travelling (with respect to the ground) after you have kicked it? For concreteness, assume that your mass M=80 kg, the mass of the skateboard m=10 kg, and the height of the incline h=1 m. You cannot neglect the mass of the skateboard.

10. A heavy flywheel rotating on its axis is slowing down because of friction in its bearings. At the end of the first minute its angular speed is 0.90 of its angular speed ω_0 at the start. Assuming constant frictional forces, find its angular speed at the end of the second minute.

11. A girl (mass M) stands on the edge of a frictionless merry-go-round (mass 10 M, radius R, rotational inertia I) that is not moving. She throws a rock (mass m) in a horizontal direction that is tangent to the outer rim of the merry-go-round. The speed of the rock, relative to the ground, is v. What is the linear speed of the girl after she throws the rock?

12. Two frictionless inclined planes have the same height but have different angles of inclinations of 45^0 and 60^0 with respect to the horizontal. You slide down from the top which is at a height h above the ground on each inclined planes starting from rest. Find your speed at the bottom of the inclined planes in the two cases.

13. In the track shown below, section AB is a quadrant of a circle of 1 m radius. A block is released at A and slides without friction until it reaches point B. The horizontal part is not smooth. If the block comes to rest 3 m from B, what is the coefficient of kinetic friction?



14. A man of mass M_1 lowers himself to the ground from a height X by holding onto a rope passed over a massless frictionless pulley and attached to another block of mass M_2 on the other side. The mass of the man is greater than the mass of the block. With what speed does the man hit the ground after falling through a distance X?

15. A man of mass M_1 lowers himself to the ground from a height X by holding onto a rope passed over a massless frictionless pulley and attached to another block of mass M_2 on the other side. The mass of the man is greater than the mass of the block. What is the tension in the rope?

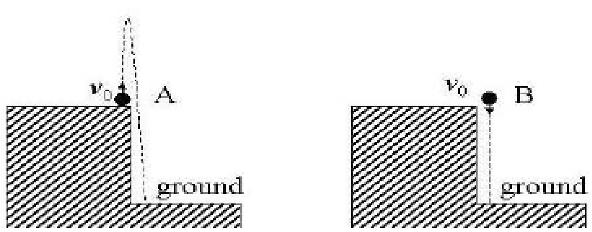
16. A 2-kg block is forced against a horizontal spring of negligible mass, compressing the spring by 15cm. When the block is released, it moves 60 cm across a horizontal tabletop before coming to rest. The force constant of the spring is 200 N/m. What is the coefficient of friction between the block and the table?

17. A 4-kg block starts up a 30^0 inclined plane with a kinetic energy of 128 J. How far will it slide up the plane if the coefficient of kinetic friction is 0.3?

18. A 2 kg block is given an initial speed of 4 m/s up an inclined plane starting from a point 2m from the bottom as measured along the plane. If the plane makes an angle of 30^0 with the horizontal and the coefficient of friction is 0.2, with what speed will it reach the bottom of the plane?

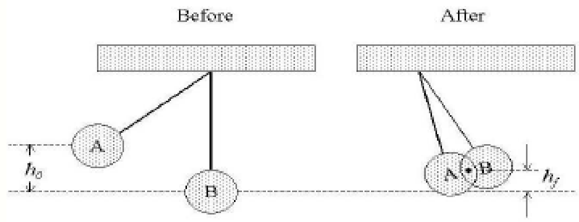
19. After the fusion reaction is over, a star collapses under its own gravitational force into a neutron star, shrinking to 1/100th of its initial radius. Assume that there are no external torques on the star and that it loses no mass as it collapses. You may treat it as a uniform sphere. If its initial angular momentum and angular speed are L_0 and ω_0, respectively, what are those variables after the collapse?

20. Two identical stones, A and B, are shot from a cliff from the same height h and with identical initial speeds v_0. Stone A is shot vertically up, and stone B is shot vertically down (see Figure). Which stone has a larger speed right before it hits the ground?



21. You dive from a diving board into the swimming pool. You want to make several sommersaults before landing in the water, should you put your arms close to your body or spread it out as you fall from the diving board? Explain your reasoning.

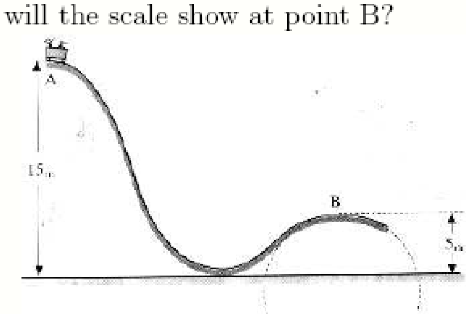
22. Two small spheres of putty, A and B, of equal mass, hang from the ceiling on massless strings of equal length. Sphere A is raised to a height h_0 as shown below and released. It collides with sphere B (which is initially at rest); they stick and swing together to a maximum height h_f. Find the height h_f in terms of h_0.



23. A family decides to create a tire swing in their back yeard for their son Ryan. They tie a nylon rope to a branch that is located 16 m above the earth, and adjust it so that the tire swings 1 meter above the ground. To make the ride more exciting, they construct a launch point that is 13 m above the ground, so that they don't have to push Ryan all the time. You are their neighbor, and you are concerned that the ride might not be safe, so you calculate the maximum tension in the rope to see if it will hold. Calculate the maximum tension in the rope, assuming that Ryan (mass 30 kg) starts from rest from his launch pad. Is it greater than the rated value of 2500 N?

24. At amusement parks, there is a popular ride in which the floor of a rotating cylindrical room falls away, leaving the backs of the riders "plastered" against the wall. Suppose the radius of the room is 3.3 m and the speed of the wall is 10 m/s when the floor falls away. What is the minimum coefficient of friction that must exist between a rider's back and the wall, if the rider is to remain in place when the floor drops away?

25. A friend told a girl that he had heard that if you sit on a scale while riding a roller coaster, the dial on the scale changes all the time. The girl decides to check the story and takes a bathroom scale to the amusement park. There she receives an illustration (see below), depicting the riding track of a roller coaster car along with information on the track (the illustration scale is not accurate). The operator of the ride informs her that the rail track is smooth, the mass of the car is 120 kg, and that the car sets in motion from a rest position at the height of 15 m. He adds that point B is at 5m height and that close to point B the track is part of a circle with a radius of 30 m. Before leaving the house, the girl stepped on the scale which indicated 55 kg. In the rollercoaster car the girl sits on the scale. According to your calculation, what will the scale show at point B?



APPENDIX B

CHAPTER 4: SUPPLEMENTAL MATERIALS

B.1 MIDTERM AND FINAL EXAM QUESTIONS

The following problems were given both in the midterm and final exams. The problem numbers refer to the numbering of the problems in the final exam. Students were given an additional sheet on which useful information was provided. For example, they were given the stationary state wave functions and energies for a one-dimensional infinite square well. For the one-dimensional harmonic oscillator, they were also given energies in terms of quantum number n, how the ladder operators relate to the position and momentum operators, the commutation relation between the raising and lowering operators, how a raising or lowering operator acting on the nth energy eigenstate of a one dimensional Harmonic Oscillator changes that state, etc.

Problem 2) The eigenvalue equation for an operator $\hat{Q}$ is given by $\hat{Q}\,|\,\psi_i\rangle = \lambda_i\,|\,\psi_i\rangle$, i = 1,…, N. Find an expression for $\langle\psi\,|\,\hat{Q}\,|\,\psi\rangle$, where $|\,\psi\rangle$ is a general state, in terms of $\langle\psi_i\,|\,\psi\rangle$.

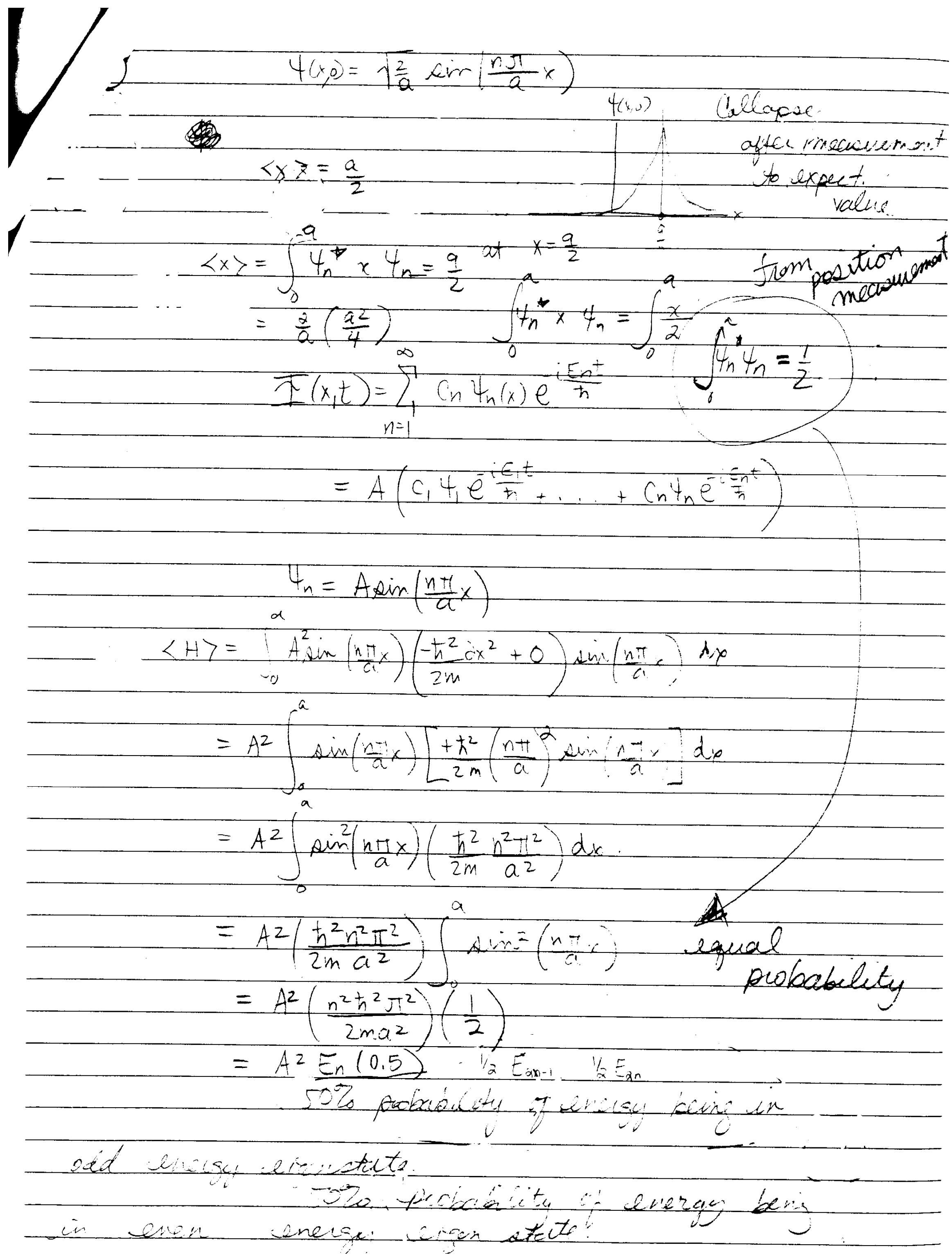
Problem 3) For an electron in a one-dimensional infinite square well with well boundaries at x=0 and x=a, measurement of position yields the value $x = a/2$. Write down the wave function immediately after the position measurement and without normalizing it show that if energy is measured immediately after the position measurement, it is equally probable to find the electron in any odd-energy stationary state.

Problem 4) Write an expression to show that the momentum operator $\hat{P}$ is the generator of translation in space. Then prove the relation. (Simply writing the expression is not sufficient…you need to prove it.)

Problem 5) Find the expectation value of potential energy in the nth energy eigenstate of a one dimensional Harmonic Oscillator using the ladder operator method.

B.2 SAMPLE STUDENT WORK

Examples of students' work on various problems on both the midterm and final exams.



$\psi(x,0) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi}{a}x\right)$

Collapse after measurement to expect. value

$\langle x \rangle = \frac{a}{2}$

$\langle x \rangle = \int_0^a \psi_n^* x \psi_n = \frac{a}{2}$ at $x = \frac{a}{2}$

$= \frac{2}{a}\left(\frac{a^2}{4}\right)$ $\int_0^a \psi_n^* x \psi_n = \int_0^a \frac{x}{2}$ $\int_0^a \psi_n^* \psi_n = \frac{1}{2}$ from position measurement

$\Psi(x,t) = \sum_{n=1}^{\infty} c_n \psi_n(x) e^{-\frac{iE_n t}{\hbar}}$

$= A\left(c_1 \psi_1 e^{-\frac{iE_1 t}{\hbar}} + \ldots + c_n \psi_n e^{-\frac{iE_n t}{\hbar}}\right)$

$\psi_n = A\sin\left(\frac{n\pi}{a}x\right)$

$\langle H \rangle = \int_0^a A^2 \sin\left(\frac{n\pi}{a}x\right)\left(\frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + 0\right)\sin\left(\frac{n\pi}{a}x\right) dx$

$= A^2 \int_0^a \sin\left(\frac{n\pi}{a}x\right)\left[\frac{+\hbar^2}{2m}\left(\frac{n\pi}{a}\right)^2 \sin\left(\frac{n\pi}{a}x\right)\right] dx$

$= A^2 \int_0^a \sin^2\left(\frac{n\pi x}{a}\right)\left(\frac{\hbar^2}{2m}\frac{n^2\pi^2}{a^2}\right) dx$

$= A^2\left(\frac{\hbar^2 n^2 \pi^2}{2m a^2}\right)\int_0^a \sin^2\left(\frac{n\pi}{a}x\right)$ equal probability

$= A^2\left(\frac{n^2\hbar^2\pi^2}{2ma^2}\right)\left(\frac{1}{2}\right)$

$= A^2 E_n (0.5)$ $\frac{1}{2} E_{2n-1}$ $\frac{1}{2} E_{2n}$

50% probability of energy being in odd energy eigenstate.

50% probability of energy being in even energy eigenstate!

Figure B.2.1a. Sample problem 3 solution of student 4 who performed poorly on the midterm exam. Here, the student confuses the collapse of the wave function into a position eigenstate after the position measurement with collapse into an energy eigenstate. He then tries to compute the position and energy expectation values instead of finding the probability of measuring different values of energy after the position measurement. Note that the state of the system after the position measurement is also not drawn correctly in the diagram.

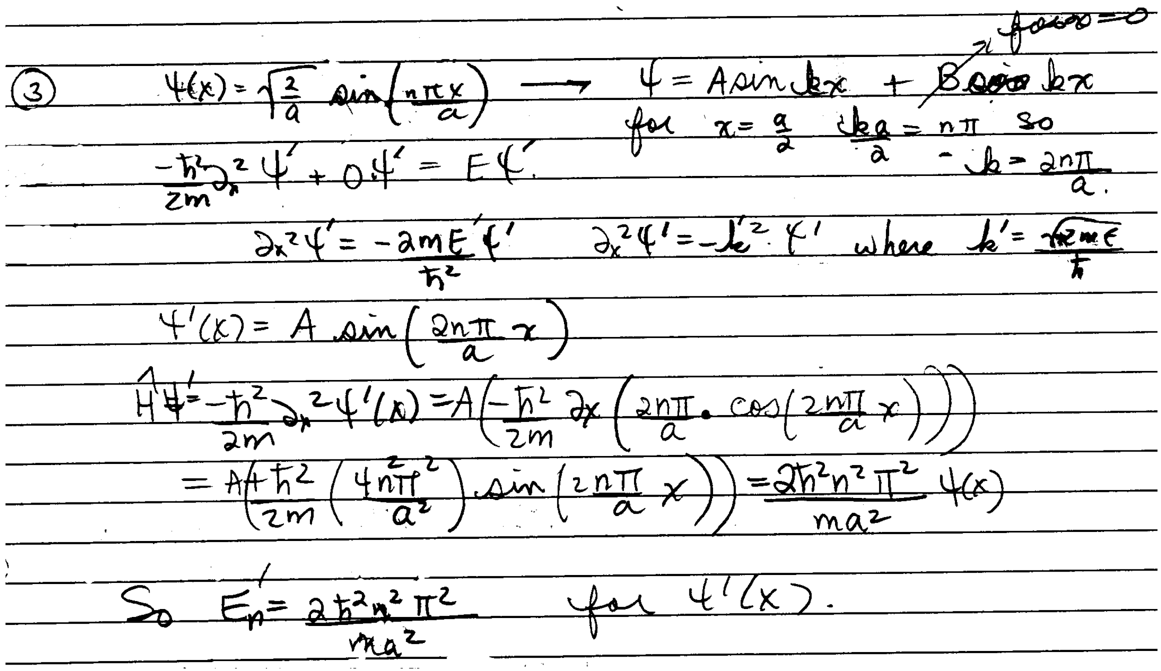
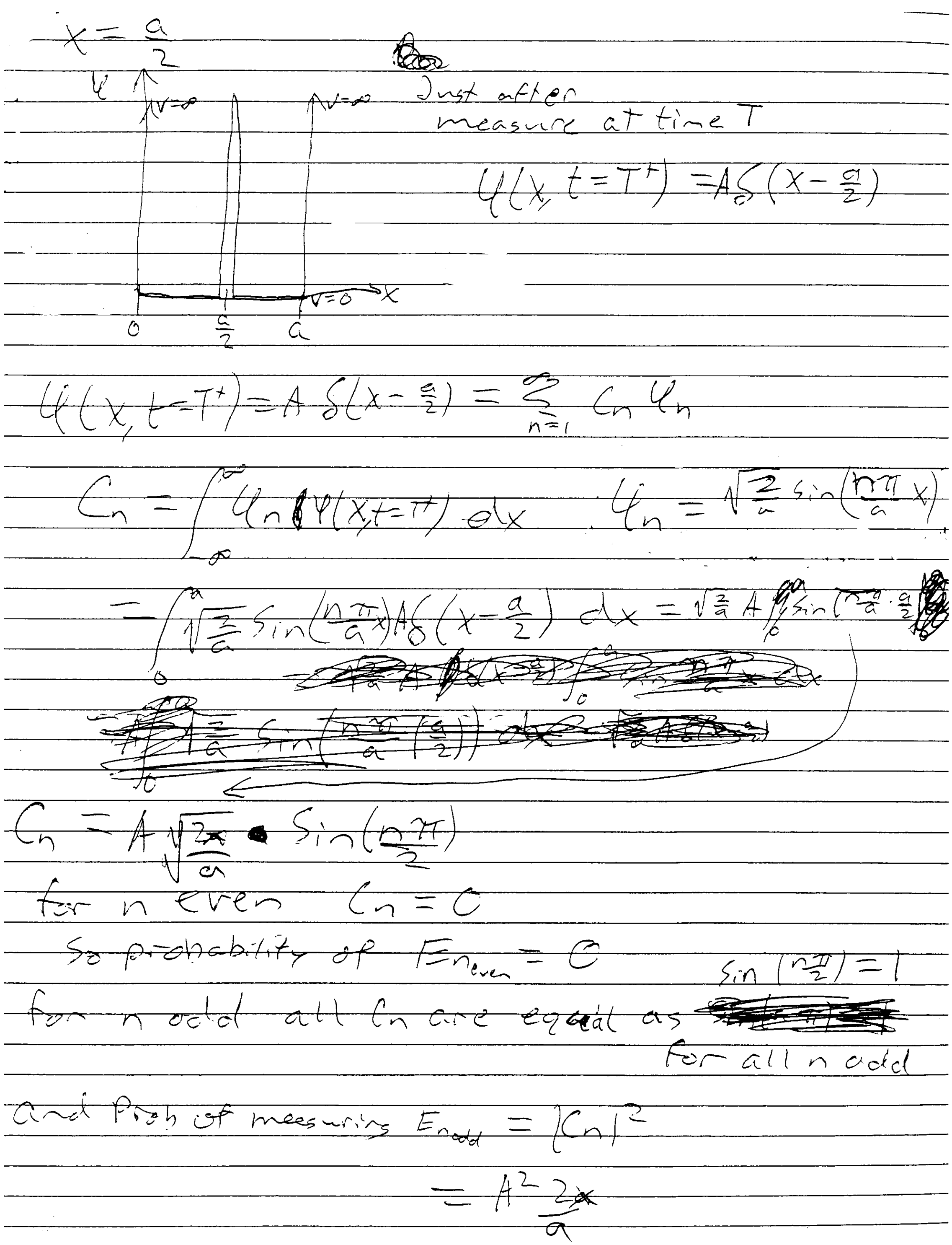


Figure B.2.1b. Sample problem 3 solution of student 4 on the final exam.

This time, the student attempts to solve the time-independent Schrodinger equation for the one-dimensional infinite square well, but fails to address issues related to position and energy measurements. While the mistake the student makes in the final exam is different from the mistake made in the midterm exam, the student has not learned from his mistakes on the midterm exam. In both midterm and final exams, the student's work is not relevant to the given problem.



$x = \frac{a}{2}$

ψ, $V=\infty$, $V=\infty$, $V=0$, x, 0, $\frac{a}{2}$, a

Just after measure at time T

$\psi(x, t=T^+) = A\delta(x-\frac{a}{2})$

$\psi(x, t=T^+) = A\,\delta(x-\frac{a}{2}) = \sum_{n=1} c_n \psi_n$

$C_n = \int_{-\infty}^{\infty} \psi_n \Psi(x,t=T^+)\,dx \qquad \psi_n = \sqrt{\frac{2}{a}}\sin\left(\frac{n\pi}{a}x\right)$

$= \int_0^a \sqrt{\frac{2}{a}}\sin\left(\frac{n\pi}{a}x\right) A\delta\left(x-\frac{a}{2}\right)dx = \sqrt{\frac{2}{a}}\,A\sin\left(\frac{n\pi}{a}\cdot\frac{a}{2}\right)$

$C_n = A\sqrt{\frac{2}{a}}\,\sin\left(\frac{n\pi}{2}\right)$

for n even $C_n = 0$

So probability of $E_{n_{even}} = 0$

for n odd all C_n are equal as $\sin\left(\frac{n\pi}{2}\right) = 1$ for all n odd

and Prob of measuring $E_{n_{odd}} = |C_n|^2$

$= A^2\frac{2}{a}$

Figure B.2.2a. Sample problem 3 solution of student 2 on the midterm.
The student recognizes that the initial state after the position measurement is a delta function in position space. He expands this initial state in terms of a complete set of energy eigenstates and performs the proof correctly.

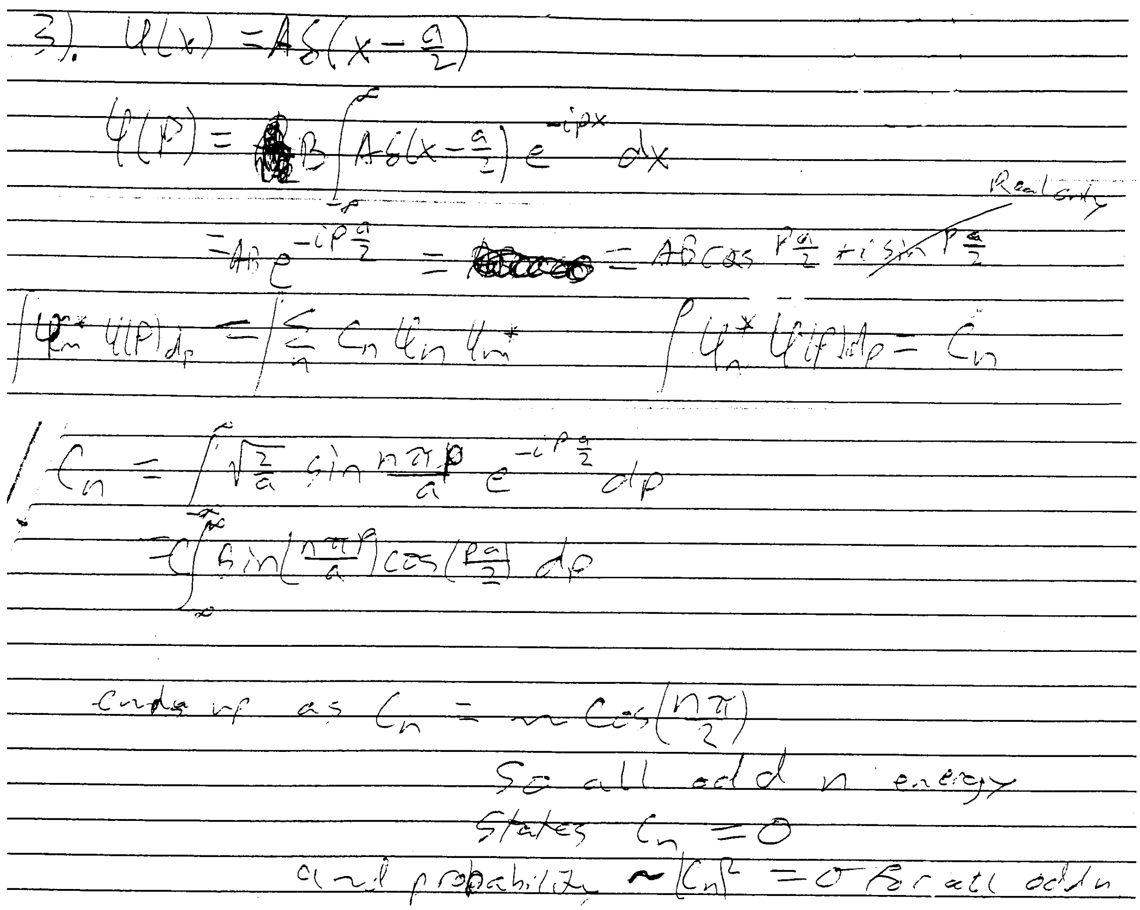


Figure B.2.2b. Sample problem 3 solution of student 2 on the final exam.

The student recognizes the form of the initial state as a delta function in position space. However, then, the student tries to perform a Fourier transform on the delta function in position space to obtain the momentum space wave function and gets lost from there, eventually "fudging" the answer and coming to an incorrect conclusion. This regression from the midterm to the final exam suggests that the student had not internalized the relevant concepts.

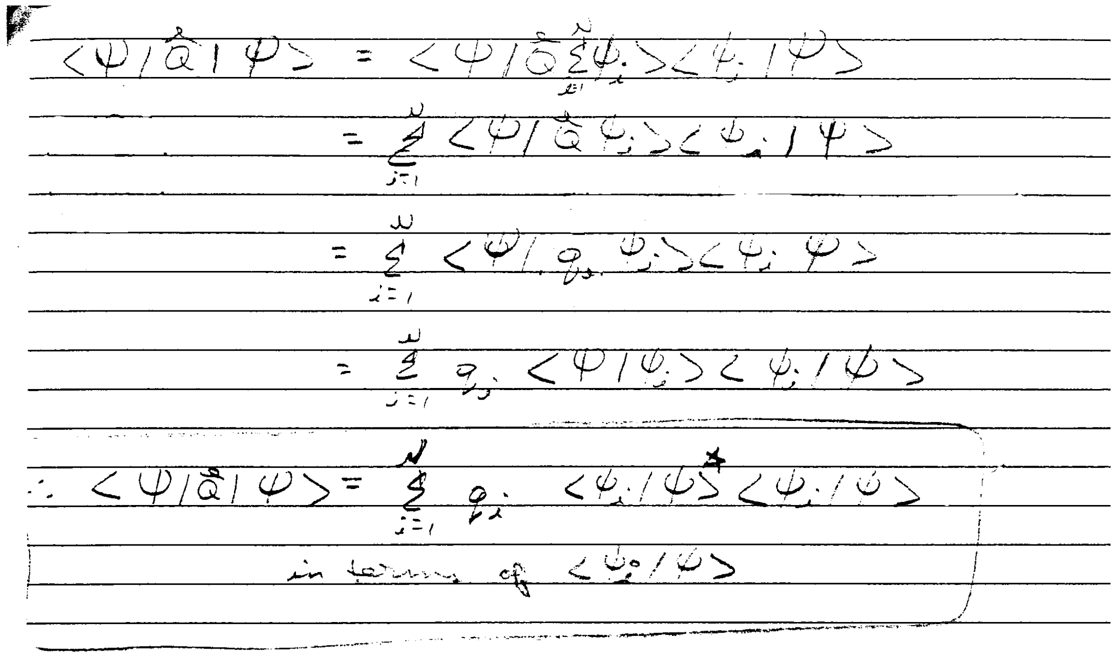


Figure B.2.3a. Sample problem 2 solution of student 3 on the midterm.
The solution is correct.

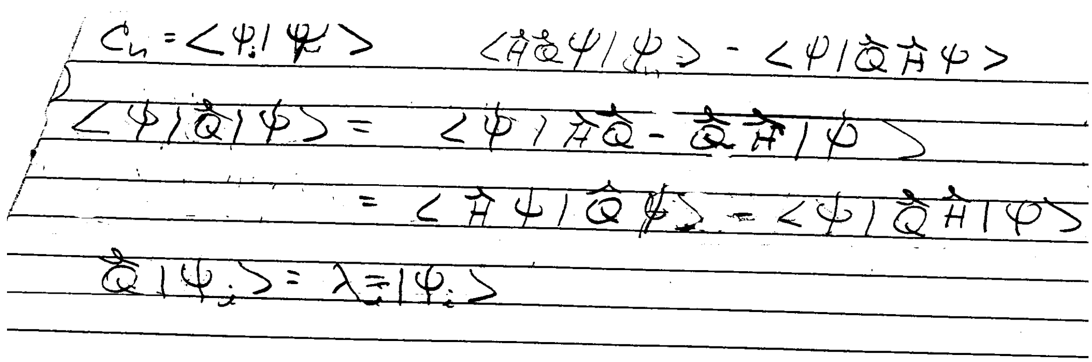


Figure B.2.3b. **S**ample problem 2 solution of student 3 on the final exam.
While this student obtained full credit on this problem in the midterm exam, he performs poorly here. It is apparent that he cannot remember how to do the problem at all.

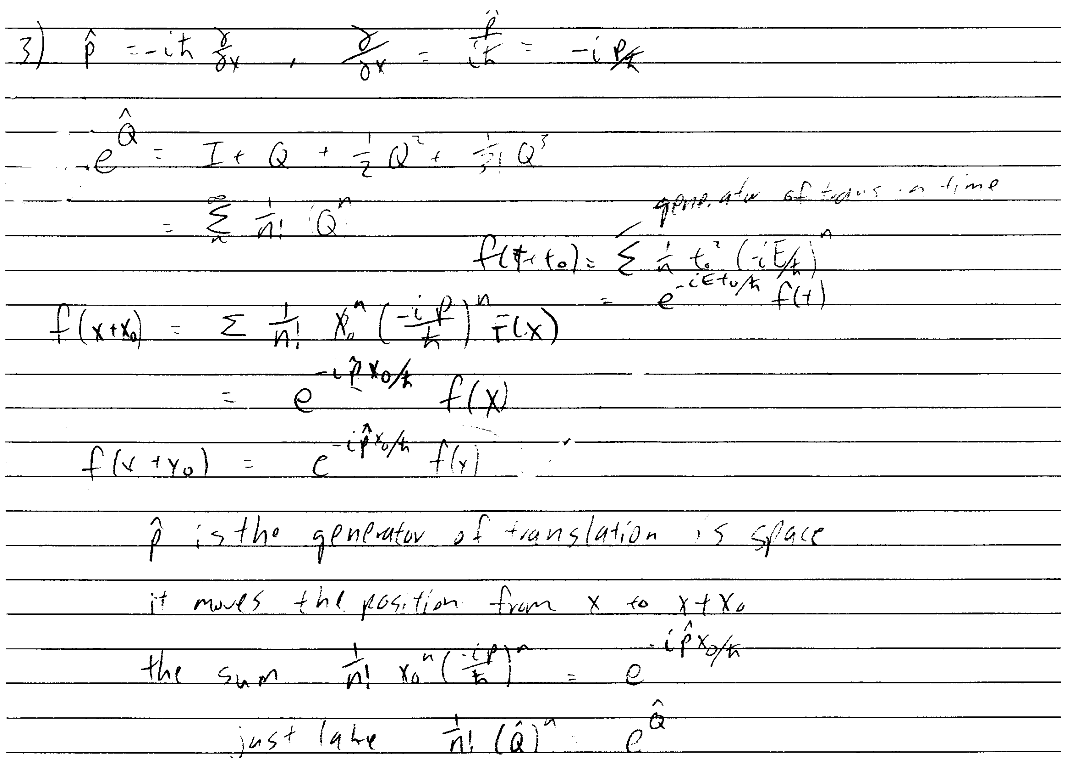


Figure B.2.4a. Sample problem 4 solution of student 6 on the midterm exam.
The student is able to correctly solve the problem except for a minus sign which we ignored in scoring.

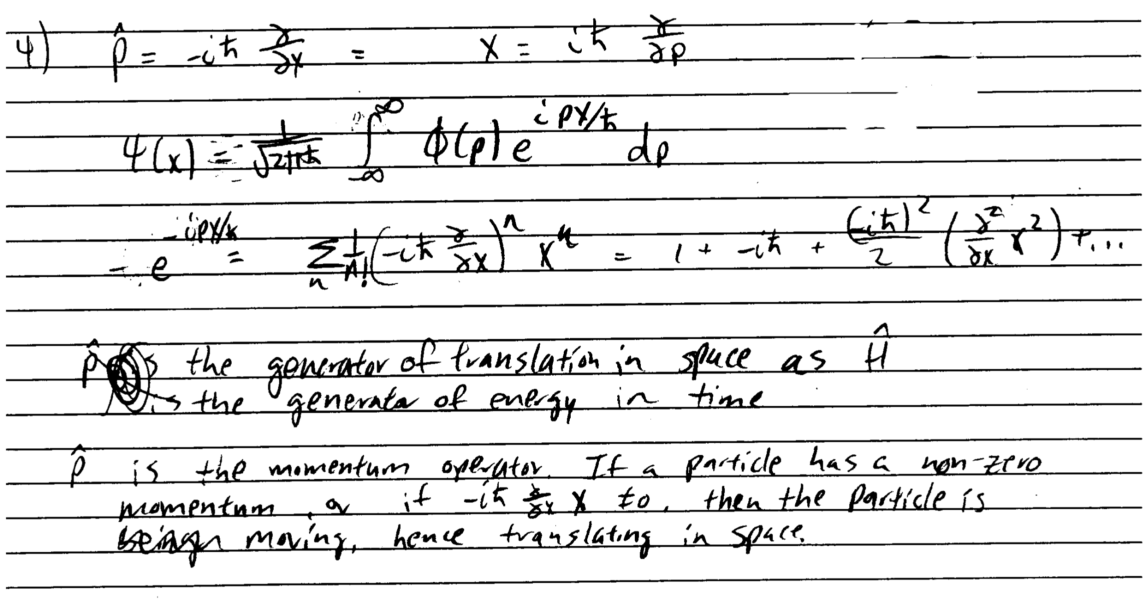
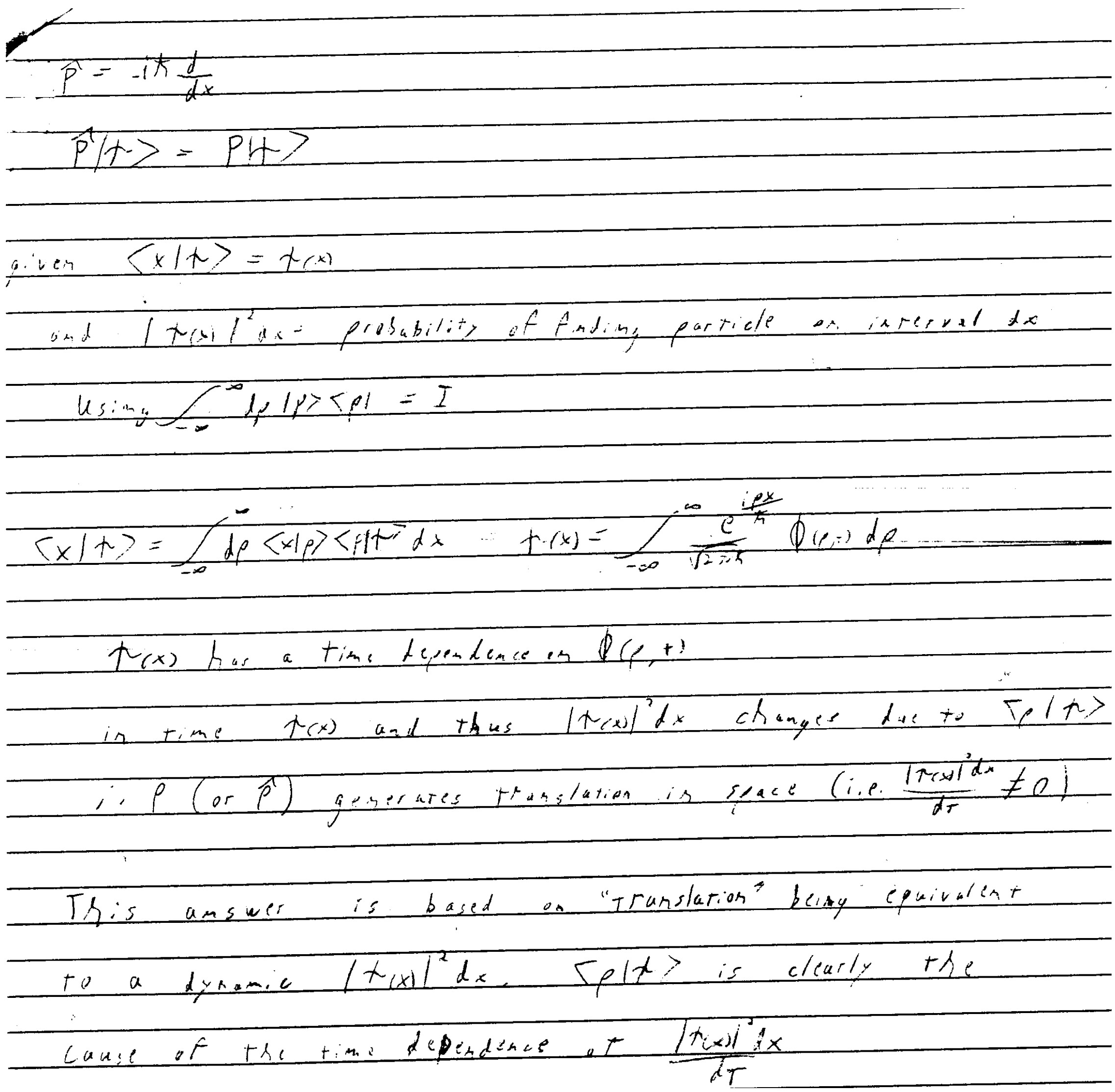


Figure B.2.4b. Sample problem 4 solution of student 6 on the final exam.
This time, in the final exam, the student remembers bits and pieces of correct ideas for the proof but did not complete the proof.



$\hat{p} = -i\hbar \frac{d}{dx}$

$\hat{p}|\psi\rangle = p|\psi\rangle$

given $\langle x|\psi\rangle = \psi(x)$

and $|\psi(x)|^2 dx$ = probability of finding particle on interval dx

Using $\int_{-\infty}^{\infty} dp\, |p\rangle\langle p| = I$

$\langle x|\psi\rangle = \int_{-\infty}^{\infty} dp\, \langle x|p\rangle\langle p|\psi\rangle\, dx \quad \psi(x) = \int_{-\infty}^{\infty} \frac{e^{\frac{ipx}{\hbar}}}{\sqrt{2\pi\hbar}} \phi(p,t)\, dp$

$\psi(x)$ has a time dependence on $\phi(p,t)$

in time $\psi(x)$ and thus $|\psi(x)|^2 dx$ changes due to $\langle p|\psi\rangle$

i.e. p (or $\hat{p}$) generates translation in space (i.e. $\frac{|\psi(x)|^2 dx}{d\tau} \neq 0$)

This answer is based on "translation" being equivalent to a dynamic $|\psi(x)|^2 dx$. $\langle p|\psi\rangle$ is clearly the cause of the time dependence of $\frac{|\psi(x)|^2 dx}{d\tau}$

Figure B.2.5a. Sample problem 4 solution of student 7 on the midterm exam. The student tries to use an irrelevant identity and then a Fourier transform but is unable to prove that the momentum operator is the generator of translation in space.

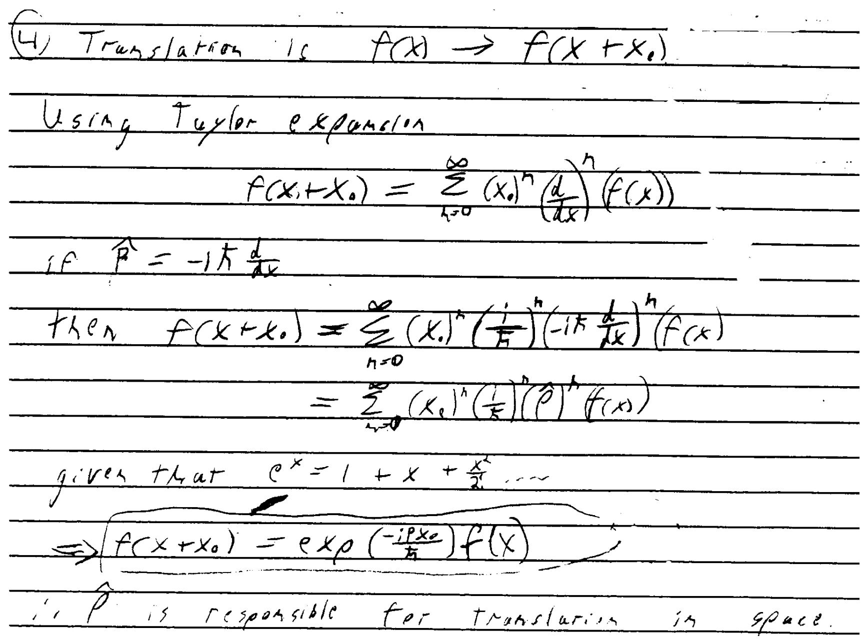


Figure B.2.5b. Sample problem 4 solution of student 7 on the final exam.

Here the student correctly writes the Taylor expansion, and hence completes the proof properly. The student seems to have learned how to do the problem correctly after struggling on this problem on the midterm exam.