Development and Validation of a Physics Problem Difficulty Measure

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Jie Yang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Kenneth J. Heller

August 2019

Acknowledgements

First, I am grateful to my adviser Ken Heller whose direction and constructive criticism not only strengthened this research study but also prepared me for future academic pursuits. This document benefited greatly from the feedback of my thesis reviewing committee: Yuichi Kubota, Cynthia Cattell, and Gillian Roehrig. This dissertation topic is based on previous problem-solving research by the University of Minnesota Physics Education Group, including: Pat Heller, Jennifer Docktor, Leon Hsu, Evan Frodermann, and Qing X. Ryan. Finally, I am also grateful to Mandy Straub, Kaylee Ganser for assistant in validity test of the measure.

Dedication

This dissertation is dedicated to the family, friends, and teachers who persuaded me to attend graduate school and provided encouragement throughout the process.

Abstract

Problem solving is both a primary goal and a standard teaching technique in introductory physics classes at the university level. To assess the utility of various pedagogical materials and techniques, it is necessary to determine student problem solving performance in the authentic situation of the course. However, this performance depends on both the student's problem-solving skill and the problem difficulty. This dissertation proposes a technique for measuring the relative difficulty of the type of physics problems typically used in introductory physics courses for physical science and engineering students. Four categories, the problem context, the physics principles, the mathematical complexity, and the number of words in the problem, were constructed based on current cognitive theories. To test the validity of this measure, 3552 student grades on 20 final examination problems, spanning the full range of topics in a one-year introductory physics course, were compared to each problem difficulty rating. Only two categories, physics principles and mathematical complexity, were needed to account for most of the student problem solving variance. Using the average of those two categories, there was an 88% Pearson correlation between the difficulty score and the average student problem solving grade. The null hypothesis, that the correlation between difficulty score and the average student problem solving grade was not significantly different from zero, had a probability, P value, < 0.001. Three experts used the difficulty measure to test its reliability and had a pairwise Spearman correlation between their difficulty ratings of greater than 94%.

Table of Content

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Introduction

Jobs that require at least a Science, Technology, Engineering and Mathematics (STEM) bachelor's degree have grown to comprise around 20% of the workforce over the past decade [1]. To avoid a 1 million STEM job candidate shortfall over the next decade, the President's Council of Advisors on Science and Technology emphasized the need to improve retention of STEM students in their 2012 report [2]. Currently, only 40% of STEM majors successfully complete their degrees [3].

Studies that have explored factors involved in STEM degree retention [4, 5, 6, 7, 8, 9] have found similar results. Independent of their prior preparation, indicated by high school GPA and ACT/SAT scores, introductory physics along with introductory mathematics and chemistry courses are key early college barriers for many STEM major students.

Changing the pedagogy of introductory physics courses to better support students in overcoming that barrier is a focus of discipline-based physics education research (PER). To have an impact on such courses, any changes must have a noticeable effect on the normal assessment of students that are based to a great extent on their solutions to quantitative examination questions that are usually called problems. How students perform on solving such problems depends on their difficulty as well as their physics knowledge and problem solving skills.

This study evolved from a desire to compare students' problem solving performance in courses using different pedagogies in the authentic environment when they have different instructors who could be at different institutions. To do so requires a simple, reproducible, and quantifiable method to measure the difficulty of problems used in regular student tests.

1.2 Research Motivation

Many physics instructors choose to assess student performance in their introductory classes primarily using open response problems. Their beliefs, born out by research literature, is that free-response problems provide rich, fine-grained information about students' reasoning and such problems emphasize the importance of synthesizing knowledge [10].

Studying how students do on exam problems is complicated by the tradition that every instructor makes up and grades their own problems. These grades are the typical data available from an introductory physics course and are a means for researchers to test the consistency between student performance and the efficacy of classroom learning with a large statistical power. An instrument to test the fidelity of instructor grading of student solutions of these problems has been developed previously [11], and has shown that instructor grading can be a reasonably accurate assessment of students' problem-solving performance. To use this grading to make comparisons among pedagogies, instructors, institutions, or curriculum, it is important to be able to gauge the relative difficulty of those problems in addition to the fidelity of the grading to the quality of student solutions. Although determining the difficulty of instructor generated exam problems has been identified as an important issue in PER [12], there currently exists no established way of doing so.

However, there are theories about students' difficulty in solving physics problems. In addition to difficulties understanding the physics ideas necessary to solve a problem, students' problem-solving is also constrained by their ability to process information to reach a solution [12]. Inspired by those established theories about student difficulty, quantifying the relative difficulty of a problem seems a promising research direction. This dissertation presents such a difficulty measure for physics problems appropriate for those typically used in introductory physics courses. The resulting instrument is based on (1) information processing theory stressing the limited capability of short term and working memory together with (2) cognitive models of mental resources and ontological categories that contribute to the cognitive load.

1.3 Overview of the Dissertation

This dissertation has five chapters that set forth the background of the research questions and give their results. It is important to note that throughout the dissertation the word "problem" refers to what are generally called problems by instructors and students of introductory physics and not to the broader class of real-world problems addressed in cognitive science.

Chapter 1 describes the motivation and need for a quantitative problem difficulty measure in introductory physics, lists the research questions addressed in this dissertation, and gives an overview each chapter of the dissertation.

Chapter 2 provides a brief summary of the current state of research to determine relative problem difficulty for introductory physics students, especially in physics education. It also describes the information processing background used to determine the problem difficulty measure developed for this dissertation. This chapter also includes theories of the origin of misconceptions because these misconceptions can lead to cognitive dissonance when students solve problems. In the framework of information processing theory, this dissonance causes a higher cognitive load.

Chapter 3 begins with a description of the difficulty measure categories, how each of these categories is scored, and how the categories are related to the theories and previous research outlined in Chapter 2. This chapter also illustrates the application of the measure to typical exam problems from introductory physics courses.

Chapter 4 presents the data supporting the validity and reliability of the problem difficulty measuring instrument.

Chapter 5 summarizes the results and limitations of the study and discusses the usability of the measure including some possible future applications.

1.4 Research Questions

This study seeks to answer the following research questions:

RQ1: How does the difficulty of a problem as measured by the instrument described in this dissertation correlate with the average student performance on problems in introductory physics? This addresses the validity of the measure.

RQ2: How well do different raters using this difficulty measure agree on the difficulty level? This addresses the reliability of the measure.

Chapter 2 Literature Review

2.1 Introduction

This chapter summarizes the factors that contribute to the difficulties typically faced by students when solving exam problems in introductory physics and relates them to the theory of information processing . It also presents a summary of previous research designed to determine the relative difficulty of the type of open response exam problems commonly used in introductory physics exams. Although the categories that comprise the difficulty measure are inspired by the theory and empirical research presented in this chapter, they are not derived from them. For the purpose of this dissertation, the success of the difficulty measure will be determined by how well it predicts the performance of students on authentic physics exam problems.

2.2 External and Internal Factors of Problem Difficulty

Jonassen [13] suggests that both external and internal factors contribute to problem difficulty. He relates the internal factors to the personal characteristics of the problem solver, such as prior experience, prior knowledge, strategies used, and epistemological development. These factors are seldom under the control of the instructor. To control these individual variations in determining the difficulty level of problems, I have used a large student population that averages over such internal factors.

External factors are inherent features of the problem, such as abstraction and continuity. Bassok [14] explained these two important external attributes of problems: abstraction refers to the content and context of problem features that either facilitate or impede the student's mapping of one problem to another. Most classroom problems are more abstract than most everyday problems which are embedded in rich contexts. Continuity is the degree to which attributes of problems remain the same or change from problem to problem. High continuity problems are more easily solved and transferred than low continuity problems. Both abstraction and continuity impact how students process the information necessary to solve a problem.

According to Meacham and Emont [15] , problems also vary in terms of complexity. They describe problem complexity as a function of external factors, such as the number of concepts, functions, or variables involved in the problem; the number of interactions among them; and the predictability of their behavior. Complexity has direct implications for working memory requirements. The more complex a problem, the more difficult it will be for the problem solver to actively process the components of the problem.

2.3 Information Processing Theory

Information processing theory deals with the interaction between the internal and external aspects of a problem. A problem, from a traditional, information-processing perspective, consists of sets of initial states, goal states, and path constraints [16]. For a situation to become a problem the interaction of the participant, activities, and context impedes the arrival at the goal state. A solution occurs when the solver finds a path between their initial state of thinking about the problem and their goal state.

A person's problem-solving is constrained by their ability to process the information necessary for them to reach a solution. According to information processing theory [17, 18, 19, 20], human problem solving can be modeled as an iterative process of representation and searching. The solution process proceeds when a student moves from an initial state of acquiring a problem's situation to their interpretation of the problem's goal by using appropriate operations like representing and searching. One such operation is translating the problem into a representation or representations that organizes information in a form they can more easily use. Then a student engages in a search process during which they select a method or methods to achieve the goal of the problem. They attempt to apply their search process either to achieve the problem's goal or a sub-goal that will move them toward their goal. They proceed from sub-goal toward the goal until a satisfactory solution is achieved or the search for a solution is abandoned. If constructing an internal representation of the problem situation's goal or sub-goal activates a single method or schema, its solution is implemented immediately. If no schema is activated, then the solver engages in more general search strategies [21] to locate less inclusive procedures that can be assembled into a solution process.

Carrying out this process of problem solving takes place in three distinct information processing mental environments: short-term memory (STM), "working" memory, and long-term memory. Short-term memory and working memory can only hold a small amount of information for a limited time, whereas long-term memory is essentially unlimited [20] . For this reason, I will combine the functioning of STM and working memory and usually call it STM for the purposes of this dissertation. In order to use information stored in long-term memory, it must be brought into working memory. If problem information and processing methodology exceed the limits of that memory, a solver experiences cognitive overload which makes a problem more difficult. To alleviate this memory overload, problem information is often stored externally by writing it down. Thus, part of processing the information for a difficult problem occurs externally to the brain using tools such as paper and pencil, mathematics or a computer. As a problem solver becomes more proficient, knowledge and procedures become compiled into larger subprocesses, or chunked, taking less space in working memory.

Based on this theory, cognitive overload could be an important cause of problem difficulty. In its simplest application, this means a problem containing more information is relatively more difficult than a problem with less information when other aspects of difficulty are controlled. This is especially true if the solver, like a novice learning a foreign language who must translate every word into their own language before understanding the meaning of a sentence [17], is not quickly able to chunk the problem situation into a larger representation. Likewise, a problem with more complexity such as one with more unknowns, requiring several equations to represent it, would be more difficult because of cognitive overload.

When a student incorporates a representation that conflicts with reality into their problem-solving process, that representation usually conflicts with other representations the student is using, with the information in the problem, or both.   This conflict creates a larger working memory load because a new problem, the conflict, has been generated that must be solved in addition to the original problem. If such representations that conflict with reality are strongly held, they are usually called misconceptions.

According to Docktor and Mestre's extensive review of PER [12], it has become widely accepted among those who follow or participate in science education research that students come to science courses with conceptions about the world that differ from reality and need to be addressed intentionally in instruction. Students' misconceptions of basic physics concepts are a major factor in the difficulty of a physics problem. From this point of view, in addition to difficulties in understanding the physics ideas necessary to solve a problem, students' misunderstanding the physics stresses their ability to process the information to reach a solution. There are two major theoretical constructs of how these misconceptions arise, resource theory [22] and ontology theory [23] . Both have an impact on cognitive load and are summarized in the next sections.

2.4 Resource Theory

Resource theory is also called "knowledge in pieces" theory [22, 24, 25]. According to resource theory, people's knowledge consists of smaller grain-size pieces that are the resources they use to reason in problem solving. Those resources or knowledge pieces are not necessarily compiled into larger concepts. Students activate one or several pieces of knowledge in response to a context and use them to reason on the fly. The misapplication of these pieces of knowledge, often through overgeneralization, is usually interpreted as a misconception. For example, individuals of all ages will state that it is hot during summer because Earth and the Sun are in closer proximity than during winter. They generate this explanation on the spot by searching memory for examples of feeling hotter and coming up with the knowledge piece "closer means stronger". In the resource theory view, they construct a solution of the problem from a general belief which is not proper in this context.

Students have sets of knowledge pieces that are the resources they use to reason about physics. For example, a well-documented misconception of students is that they think the electrons comprising an electric current are used up when lighting a lightbulb. The lack of the piece of knowledge or resource of current conservation causes the student to construct a representation of the lighting a lightbulb process from their other mental resources. The misconception could be explained by students activating the more general belief that every effect, such the light being on, must be caused by using up something, the electrons that flow in a complete circuit. The specific resource that the student might employ is their knowledge that batteries run down, a situation explained if the battery's electrons are used up by the lightbulb. Instead of using the resource of current conservation, the student uses short term and working memory to construct a new representation of the problem situation, using up electrons. More cognitive load might occur because there is no process to determine how many electrons are used up and one might be constructed. This in turn produces a large cognitive load when the student tries to relate the current and voltage in Ohm's law, a resource most students have at this stage. The resulting short-term memory overload can cause the student to use mathematically and logically inconsistent results to reach a solution or abandon the goal of a solution all together.

2.5 Ontology Theory

Ontology theory proposes that people categorize the world they interact with into broad categories called "ontological categories", literally categories of things and phenomena that exist in the real word [23]. According to Chi's work [23], the two most used ontological categories in physics are substance and process. From this point of view, misconceptions are caused by putting knowledge and experiences into inappropriate ontological categories [26, 23, 27, 28]. The substance category is the one that relates most directly to a person's life experience. For this reason, people start by putting every concept in this category. Concepts that fit the substance metaphor are used more correctly than those that are not. For example, energy, because it is a conserved scalar quantity, can be fruitfully treated as substance. It is an abstraction that is taught using concrete metaphors like fluid flow and represented using a pie chart or bar chart that are also used for counting the amount of a substance. On the other hand, if the substance metaphor is used for force, which is not conserved and is a vector, the substance metaphor causes misconceptions. For example, students think of a force as given to an object (a substance property) and being used up as gasoline (a substance) is used up in a car. Thus, they describe throwing a ball as a person's hand transferring a force to the ball and that force being depleted as the ball moves.

In this theory, remediating misconceptions requires people to reorganize their ideas from one ontological category to another one. This requires such a large cognitive load that it usually cannot be accomplished on the fly while solving a problem. Thus, using this misconception in a force problem raises the difficulty of determining the rate that force is dissipated, a concept that does not exist in physics. This in turn produces a large cognitive load as the student must assemble their mental resources to produce this non-physical concept. Again, the student has a short-term memory overload that results in their using mathematically and logically inconsistent results to reach a solution or abandoning the goal of a solution.

2.6 Empirical Studies of Problem Difficulty

There have been few previous studies to determine the difficulty of instructor constructed problems used in introductory physics. These studies addressed two questions: (1) How do students and instructors perceive physics problem difficulty? (2) What are the problem features that could be used to determine difficulty?

In addressing the first question, researchers at Kansas State University found there was a correlation of 84% between instructors' ratings of difficulty and a measure of problem complexity which they defined as the number of equations needed to solve a problem, but no significant correlation between students' ratings and problem complexity [29]. In this study, researchers asked undergraduate students and instructors to rate the difficulty of textbook-style kinematics and work energy problems on a 1 to 10 scale. For the Work-Energy topic, 15 undergraduate students in a calculus-based introductory physics course solved and rated the difficulty level of 16 problems. Fourteen instructors were also asked to judge the difficulty level of those problems for a typical calculus-based introductory physics student. For the Kinematics topic, 21 undergraduate students in a calculus-based introductory physics course and 15 instructors rated the difficulty level for 10 problems. For both Kinematics and Energy, the instructors' and students' difficulty ratings correlated with the students' grades for the problem solution. The instructor problem difficulty rating was more highly correlated with the average student grade on the problems than that of the students.

In another study at the University of Illinois, instructors (4 advanced PER graduate students and 4 PER faculty members) and students (38 undergraduate students enrolled in an introductory calculus-based mechanics course) were asked to pick out the relatively more difficult problems from 79 multi-choice physics problem pairs [30]. All problem pairs were selected from previously administered exams and referred to the same situation. Instructor problem difficulty rankings had a correlation of 96% with the fraction of student correct answers which was better than correlation with the student problem difficulty ranking of 71%. As in the Kansas State study, the instructors were more accurate than students in estimating the difficulty students had in getting the correct answers. In this study, the instructors were asked to write down their reasons for each ranking of problem difficulty. The instructors all had similar criteria for judging problem difficulty. Their common reasons were grouped by the researchers into 3 categories: (1) Question context including question type, distractor, and wording; (2) Content type including more steps, math, direction, and content; and (3) Student characteristics including their familiarity, misconceptions, intuition, and carelessness.

In a study at the University of Minnesota, PER researchers analyzed approximately 2000 student solutions to open-response problems. They found 6 characteristics that they judged contributed to the problem difficulty. These characteristics were: (1) Unfamiliar problem context; (2) Lack of explicit physics principle cues; (3) Extraneous or missing information; (4) Implicit rather than explicit problem target; (5) Number of necessary physics principles; and (6) Number of equations needed. The authors scored each of these six characteristics as 0 (easier) or 1 (more difficult). They found that the sum of the six characteristics accurately predicted the student performance solving the course problems [31]. Both the Illinois and Minnesota studies pointed to problem features such as problem context and number of steps as important features in determining problem difficulty. However, neither of these studies tried to quantify the degree of a problem's difficulty using those features.

A study from The Chinese University of Hong Kong developed a difficulty measure for a specific topic (logarithm problems) in algebra for 9th grade students [32] that is similar to the difficulty measure developed for this dissertation. They found 4 significant factors for determining problem difficulty in this very limited problem regime for a very different set of students: (1) the perceived number of difficult steps (steps where students made non-trivial errors); (2) the number of steps required to finish the problem (the number of steps that an expert would use to solve a problem by the shortest path); (3) the number of operations in the problem expression (an operation was defined as addition, subtraction, multiplication, division, or exponentiation); and (4) students' degree of familiarity with the type of algebra needed to reach the correct answer (problems learned at earlier stages of their education were assumed to be more familiar to students). Their measure using these 4 factors had a higher correlation (81%) with the students' estimate of the problem difficulty than the teachers' (74%) which is different than the results of the Kansas State and Illinois studies. Moreover, the students' estimate of problem difficulty correlated even more strongly (86%) with the percentage of students who answered the problem correctly.

2.7 Summary

In summary, information processing theory supplemented by the resources and ontological category theory of concept formation is a reasonable grounding for the development of a physics problem difficulty measure. The methodology described in the next chapter, uses this theory to construct four major predictors of difficulty: problem context, physics principle, length of problem statement and mathematical complexity. These predictors also reflect the common difficulty factors identified by the empirical studies of problem difficulty described in this chapter.

Chapter 3 Construction of the Difficulty Measure

In this chapter, I describe the design criteria and the development of the difficulty measure with four major categories: Problem Context, Physics Principles, Length of problem statement and Mathematical complexity. I will explain why these four factors are important, specify how to use a 1-5 scale to quantify the difficulty in each category, and give tables for determining the difficulty of problems for a calculus-based introductory physics courses for physical science and engineering majors. This classification of problem difficulty arises from the theories described in Chapter 2 and has some overlap with the difficulty criteria previously proposed in the research summarized in that chapter.

3.1 Design Criteria

My goal was to develop an instrument that would be practical for both researchers and instructors to determine the difficulty of free response problems used in the authentic classroom practice in introductory physics.

The design requirements for the difficulty measure were as follows:

(i) Ease of use. — A difficulty measure is easier to score and interpret if it minimizes the number of categories and the complexity of the scoring.

(ii) Usability in authentic situations. — The difficulty measure focuses on problems written by typical physics faculty members and is applicable to problems spanning the range of problem topics in a typical introductory physics class.

(iii) Evidence for validity, reliability, and utility of the measure. — In particular, the open response problem difficulty measure should quantify the major differences between difficult and easy problems and agree with the performance of the students.

I have tested the difficulty measure for its consistency across several raters and the various topics found in a calculus based introductory physics course. Most importantly, I have taken as an operational definition of problem difficulty the average problem-solving performance of students within a single classroom and pedagogical environment. Problem solving performance in this context means the student’s progress in writing a logical solution built on correct physical and mathematical principles [11], it does not necessarily require arriving at the correct answer. The results of these validity and reliability studies are described in Chapter 4.

3.2 Design Process

According to the theories outlined in Chapter 2, any factors that cause extra cognitive load will increase the difficulty level of a problem. Based on past research about information processing [17, 18, 19, 20], I identified the length of the problem statement and the number of equations used in the solution as two factors that naively would cause a short term memory load and could be used to quantify problem's difficulty. In addition, the connection of resource and ontological category theories with information processing, suggested that the student's familiarity with the problem context and the physics principles needed for the solution were factors that could be used to quantify problem difficulty. Applying the criteria of having the fewest number of categories with the least difficult scoring method, I arrived at four categories for a difficulty measure that I could delineate and reliably implement. These four categories were Problem Context, Physics Principles, Length of Problem statement, and Mathematical Complexity of the solution. Although these categories were inspired by the theory in Chapter 2, I do not claim that they are only way to categorize the difficulty of introductory physics problems. However, as seen in Chapter 4, these categories account for 80% of the variance in student problem solving performance for the sample of the more than 3500 student solutions examined. A more detailed description of these categories and how they are scored is described in next section.

The categories for determining problem difficulty were refined by examining final exam problems and their student solutions to decide what characteristics of each category were easiest to score. These problems were not used in the validity study data of the next Chapter. To make sure that the difficulty scale spanned the problem characteristics typically used in this course, I read through the past 20 years of quizzes and final exam problems from the calculus based introductory physics course at the University of Minnesota. I also compared them to the end of chapter problems found in several of the most common textbooks for this course and found them similar in structure and content.

Then I solved typical open response exam problems from 3 types of introductory courses (algebra-based, calculus-based for engineering and physical science students, and calculus based for biology and premedical students) from the last 5 years at the University of Minnesota. Usually this consisted of 2 problems from each of 3 quizzes, and 5 from each final exam. There were 330 problems in total with a significant amount of overlap. In the interest of scoring simplicity, I have taken the most naïve and mechanistic view of determining the difficulty range of each category by reducing its evaluation to counting where possible. However, evaluating the difficulty level of the physics principles needed to solve a problem was not a straightforward counting process. To arrive at a method to determine this difficulty range, I sorted the problems based on the most fundamental physics principles needed for each solution. I found that most of the problems from the entire first year course relied on one of two basic physics principles, energy or force and motion, for their solution. However, the mathematical concepts used to implement these principles increased in abstractness and subtilty, called mathematical sophistication, as the course progressed. Based on that information, I developed the criteria for scoring each category described below.

3.3 Categories and scoring

As described above, I used four categories to determine problem difficulty that I could reliably distinguish while reading typical introductory physics test problems. For scoring ease, I chose a 1-5 scale for each category, with 1 being least difficulty and 5 being most difficult and averaged all categories to give a final difficulty score. I do not assume that the scale for any category is linear so that a difficulty level of 4 is not necessarily twice as difficult as a level of 2. I did not assign a weight to each category to distinguish its difference in importance because I could not find any theoretical justification for doing so in the research literature. The results, given in chapter 4, show that there is a 90% correlation between the average of the four difficulty scores and students' problem-solving performance. However, after the data analysis described in Chapter 4, some categories proved to be significantly less important than others in predicting the overall problem difficulty and could be eliminated. Below I will explain why each category could be important and give the specific methods of scoring it on a 1-5 scale.

3.3.1 Physics Principles

Physics principles gives the minimal collection of fundamental physics needed to solve a problem. The difficulty that students have using these principles depends both on their tendency to trigger misconceptions and the sophistication of the mathematical ideas needed to use them. In introductory physics, two principles underlie most topics. Almost all problems are solved using extensions of the Force principle or the Energy principle. The extended Force principle includes interaction quantities such as force, torque, 3-dimensional momentum and angular momentum, and vector fields while the extended Energy principle includes conserved scalar quantities such as energy, charge, and one-dimensional momentum and angular momentum.

Based on the theories described in Chapter 2, the “force principle” is more difficult to use than the “energy principle” for several reasons. First, according to the ontological category theory, energy can most often be used by considering it as a substance while force cannot. As summarized in Chapter 2 this leads to the possibility of misconceptions increase cognitive load [33, 34]. According to the resource model, a substance-based ontology is most commonly used by students because of its concreteness and the large number of mental resources from daily life [26]. This means that a problem typically solved at the introductory level by an energy principle should be less difficult than one typically solved using a force principle. In addition, misconceptions that students have about energy would be easier to change within the course because they usually involve modifications within the substance category. Thus, they are less likely to survive to cause difficulty on an exam than force misconceptions which usually involve switching out of the substance category [26].

However, there is a situation where the energy approach becomes more difficult. When an object is bound to another such as in gravitational or Coulomb attraction, their energy is negative compared to its value when the objects are completely separated. Using negative energy has been documented as an area of difficulty for students [35, 36]. The reason could be that a negative substance is not a common experience so no easily accessible mental resources exist, and the students cognitive load increases to produce those resources on the fly.

The use of a specific physics principle to solve a problem in introductory physics is often linked to a mathematical idea to express or implement it. For example, because force is a vector, the use of this principle usually necessitates understanding a more sophisticated mathematics than using the energy principle. The use of vector mathematics increases the amount of conceptual processing required and thus increases cognitive load because these students have not automated the idea of vector addition. This additional difficulty factor also applies to vector conserved quantities such as momentum. An additional mathematical sophistication is usually required in using the force principle to determine an object's motion because force is related to motion through the object's acceleration which is both a vector and a rate of change of a rate of change. This conceptual mathematical sophistication load is in addition to the additional algorithmic steps required to accomplish the vector calculation that I include in another category, mathematical complexity. All those features make the use of the force principle in a problem more difficult than the use of the energy principle in most cases.

In addition, there are physics principles that are most easily expressed using mathematical ideas. In those cases, mathematics is not just a calculational tool, it is a conceptual device that allows chunking which reduces cognitive load. For example, the use of Gauss' Law to understand the relationship between charges and electric field is a direct connection of a mathematically sophisticated idea, surface integrals, to physics principles. Because most students have only experienced the calculational aspect of this mathematics, they have not yet built the mental resources necessary to conceptually apply it to a physics problem without incurring a significant cognitive load. Notice that mathematical sophistication in this case does not necessarily include the need for additional mathematical steps or new mathematical techniques which are addressed in the mathematical complexity category.

Based on the research, cited below, of the use of mathematics by physics students and the experience of instructors, the following is a list of mathematical ideas that increase the difficulty level of a problem. I have listed this mathematics in order of increasing sophistication which I assumed correlates to their increased cognitive load for students. Note that mathematical sophistication is the conceptual representation of an idea and is independent of calculational technique.

• Algebraic or proportional reasoning. Some students have difficulty translating a sentence into a mathematical expression, and in doing so they often place quantities on the wrong side of an equal sign [37, 38].

• Vectors. Studies [39, 40] have documented student difficulties associated with using vectors, particularly for velocity, acceleration, forces, and electric fields.

• Differential calculus. Studies have reported that students often don't understand the idea of differentiation as it applies to real situations [41].

• Integral calculus. Studies [40, 42, 43] show that even upper division students have difficulty dealing conceptually with the types of vector integration used for Gaussian surfaces or Amperian loops. Even the most elementary uses of these abstract geometrical constructs add difficulty to a problem.

In summary, the difficulty of students using physics principles is three-fold. First, they may require switching ontological categories or linking many fine-grained mental resources. Second, they may trigger misconceptions that short circuit or confuse the cognitive process. Third, some require more sophisticated mathematics to more easily represent their conceptual meaning than others which adds additional levels of difficulty.

The Physics Principle Difficulty Scale

Based on the discussion above, Tables 1 and 2 summarize the relative problem difficulty within the two primary physics principles (force and energy) in a calculus-based introductory physics course. I took the principles that directly use the substance ontology, such as conservation of energy or one-dimensional conservation of momentum, to be level 1. I then assumed an incremental model of difficulty, adding 1 for every type of problem difficulty feature that occurs. However, I set the highest difficulty level to 5 no matter how many difficult features are involved in a problem solution. A more detailed explanation of the difficulty levels is given below.

Conservation Principles / Energy:

Physics Principles	Difficulty level	Reason
Conservation of Energy for a closed system	1	Direct use of substance ontology
One dimensional conservation of momentum	1	Direct use of substance ontology
Conservation of momentum in 2 or 3 dimensions	2	The use of vectors.
Energy transfer, work, caused by a constant force	2	Extra cognitive load of a continuous transfer process.
Negative Potential Energy with a constant field	2	Violates the naïve substance ontology
Energy transfer, work, caused by a non-constant force	3	The transfer process with the sophisticated mathematical process of integration over a path.
Negative Potential Energy with a non-constant field	3	Violates the naïve substance ontology and requires integration over a path.

Table 1: Conservation Principle difficulty scale.

Note that the cognitive load of using conservation of energy is increased by adding sophisticated mathematical ideas or violating the substance analogy. For example, including work in conservation of energy remains within the substance ontology but adds the mathematical idea of accumulating the effect of a force over a path, conceptually an integration, that increases its difficulty score by 1 to a level of 2. Likewise, using conservation of energy with a negative potential energy violates the substance ontology resulting in an extra cognitive load that increases its difficulty to level 2. When a non-constant force is used for work or non-constant field for potential energy, the accumulation process represented by the integration of a variable adds an additional degree of difficulty level to level 3.

Interaction Principles / Force:

Physics Principles/Laws	Difficulty level	Reason
Newton's 2nd law in one dimension	2	Force violates substance ontology
Two-dimensional Force or Motion	3	Adds the additional mathematical sophistication of vectors.
Varying Force and Oscillations in one dimension	3	Adds the mathematical sophistication of a varying force.
Force with circular motion	4	Adds the mathematical sophistication of varying acceleration.
Gauss's Law for a static electric field Ampere's Law with a static magnetic field	5	Few if any mental resources exist for representing the required imaginary surface or path adding mathematical sophistication.
Faraday's Law linking a changing magnetic field to an electric potential	5	Few if any mental resources exist for representing flux which is an unfamiliar application of the substance ontology. Mathematical sophistication includes the idea of a vector integral over an abstract surface. Also requires the mathematically sophisticated concept of rates.

Table 2: Interaction Principle difficulty scale.

Note that when a force changes with time as, for example, in circular motion or oscillations, this adds another element of mathematical sophistication needed to describe a motion increasing the difficulty to level 4. Physics principles that require the concept of integrating fields over imaginary surfaces, as in Gauss's law, or imaginary paths, as in Ampere's law, add yet another level of difficulty. Equally sophisticated is the concept of a changing flux, as in Faraday's law, which combines using both the idea of integrals and derivatives. Even in its most simple calculational form, this requires mental resources that have not been chunked by most introductory students resulting in a cognitive load that that takes it to difficulty level 5.

3.3.2 Mathematical complexity

The mathematical complexity is distinct from mathematical sophistication in that it is only concerned with the length of the logical, usually mathematical, chain needed to reach an answer. Because introductory students typically view mathematics solely as a method for calculating an answer, they often do not use a written mathematical process to help them determine a path to a solution. For that reason, they must rely on their working and short-term memory to guide their mathematical calculations. This means that a problem with more interrelated unknown quantities requires more information in short term memory and more processing in working memory. For example, a circuit with more branches and components has a higher difficulty level than a simpler circuit with fewer. To quantify the difficulty caused by mathematical complexity, I chose to count the number of unknown quantities in the problem. Typically, the number of interrelated equations necessary to solve a problem is equal to the number of unknowns.

The mathematical complexity difficulty scale

Table 3 gives a difficulty scale for the mathematical complexity of a problem in a calculus-based introductory physics course where the students are majors in physical science or engineering. Of course, the level of difficulty for this category depends on the student population, especially their ability to organize long calculations on paper to reduce the load on their short term and working memory. Based on an analysis of test problems over the last twenty years of this course, the number of unknowns in a problem typically ranges from 1 to 5. For simplicity, I use 1 extra unknown as the increment for the difficulty level.

Difficulty level	1	2	3	4	5
Mathematical Complexity	1 unknown	2 unknowns	3 unknowns	4 unknowns	5 or more unknowns

Table 3: Mathematical Complexity difficulty scale.

3.3.3 Problem Context

Problem context is composed of a storyline or situation and the objects in that situation. According to resource theory, students possess a variety of mental resources that are activated differentially in specific situations. The more familiar the situation, the more likely it is that students have a collection of appropriately chunked mental resources to build a representation or series of representations that facilitate a solution. If the situation is unfamiliar, the student must assemble many discrete mental resources which adds to the cognitive load. For example, the principle of conservation of energy can be easier to use in the context of a skateboard going up a ramp than for the excitation of an atom because most students are more familiar with the behavior of skateboards than atoms. For any particular student population, the rater needs to know their experiences to judge the degree to which a situation is familiar, or the objects involved are concrete.

The concreteness of the objects in a situation is another factor in determining the problem difficulty in the context category. For example, the situation of "a collision between two blocks" refers to abstract objects, blocks, that are not usually linked to collisions in the minds of typical college students. This means that using a collision with blocks to solve a problem requires more mental resources than the more familiar situation of a car collision. For example, in contemplating block collisions students tend to equate the initial and final kinetic energy of the system because abstract colliding blocks do not easily link to the resource of change. On the other hand, the problem situation "two cars collide" is likely to trigger the change resource that allows the final kinetic energy to be less than the initial kinetic energy. This is because most students have a visual car collision experience where the cars are different before and after the collision. This makes the difficulty level of a context using abstract objects higher than using concrete and familiar objects.

The Context Difficulty Scale

Based on a context analysis using information processing theory that incorporates resource theory, Table 4 gives a difficulty scale for students majoring in physical science or engineering in a calculus-based introductory physics course. The least difficult context is one for which the situation is familiar, and the objects are both familiar and concrete while the most difficult is an unfamiliar situation with abstract objects. An explanation of Table 4 is given below.

Difficulty level	Problem context	Examples
1	Concrete and familiar objects in a familiar situation.	A car goes down a slope. Circuits with resistors, batteries and switch.
2	Concrete and familiar objects in an unfamiliar situation. Concrete but less familiar objects in a familiar situation.	A car going around a banked curve. A circuit with capacitors or inductors.
3	Concrete but unfamiliar objects in an unfamiliar situation.	Two identical stars orbit around each other.
4	Abstract objects in a familiar situation.	An infinite long straight current carrying wire placed near a small wire loop.
5	Abstract objects in an unfamiliar situation.	A uniform time dependent magnetic field along the axis of a solenoid.

Table 4: Problem Context Difficulty Scale.

For example, all students in the classes tested have experience with cars, either while riding in them, observing them, or seeing them in movies. To them cars are very familiar and concrete objects. Likewise, most students majoring in engineering or physical sciences are familiar with the basic electrical circuit consisting of batteries, switches, and lightbulbs or resistors either from personal experience or previous class work. Note that being familiar with a situation and the objects in that situation does not mean that the students have the correct idea about their behavior. Difficulties with problem solving caused by the common misconceptions that often arise in such contexts are accounted for in the category of physics principles.

By way of contrast, the problem context of a car going around a banked curve, has concrete and familiar objects, a car and a road, in an unfamiliar situation. Although students have experienced going around curves, they have generally not attended to the banked curve part of the experience. This example of familiar and concrete objects in an unfamiliar situation has the same difficulty, level 2, as does a familiar situation, such as an electric circuit, with a concrete but an unfamiliar object such as a capacitor. It is also possible to have concrete but unfamiliar objects interacting in an unfamiliar situation such as binary star orbits. Stars are concrete objects to these students, but they have no experience, either from real life or schooling, about how they interact. This makes them concrete but unfamiliar objects. A binary star system, where both objects have similar mass, is also an unfamiliar situation compared a planet orbiting a much heavier star, raising the difficulty level to 3.

When the problem context deals with abstract objects such as an infinite long straight wire or a uniform magnetic or electric field, the difficulty level increases to a level 4 even if the situation is familiar. Finally, there are problems that have abstract objects in an unfamiliar situation such as a uniform time dependent magnetic field in a solenoid. This has the highest level of difficulty, level 5.

3.3.4 Length of the problem statement

The number of words in a problem statement could also be important in determining its difficulty. In a physics problem, introductory students often attend to every word of a problem statement because they have not built the mental resources to quickly identify meaningful information [19]. As reviewed in Chapter 2, short-term memory can only hold a small amount of information for a limited time [20]. A student in introductory physics reading a physics problem is much like a person learning a foreign language who reads by translating every word into their native language and then processing those words. If students read and process every word in a physics problem individually without automatically assembling them into larger concepts, a longer problem statement puts more information in short term memory and increases the load on working memory. For this measure, I chose word count as a proxy of the amount of problem information. Word count averages over the other semantic features such as word familiarity, word length, and syntax.

The Problem Statement Length Scale

Table 5 gives a problem difficulty score based on the length of the problem statement. After reviewing test problems for the past twenty years for the calculus-based introductory physics course for physical science and engineering majors at the University of Minnesota, I found that the length of the problem statement typically falls in the range of 25 to 250 words. I use 50 words as a convenient increment for a difficulty level change.

Difficulty level	1	2	3	4	5
Length of problem statement	50 words (25-74)	100 words (75-124)	150 words (125-174)	200 words (175-224)	>250 words (225-)

Table 5: Length of Problem Statement difficulty scale.

In summary, Tables 1 to 5 each give the scale by which the four different aspects of problem difficulty for students was measured. In the next section, I will give examples of applying the difficulty measure to real introductory physics exam problems.

3.3.5 Examples of applying the difficulty measure

Below I described how I arrived at the difficulty score for a subset of the final exam problems that I used in this study. Although I have kept these problems very close to their original format, I have made some changes to facilitate easier reading here. I have also outlined a solution for each problem.

Problems 1 and 2 below are from the first semester of an introductory physics course (mechanics) final exam. Problem 1 is relatively easy while problem 2 is more difficult. Problems 3 and 4 are from the second semester of an introductory physics course (electricity & magnetism) final exam. Problem 3 is relatively easy while problem 4 is more difficult.

Problem 1:

You are helping your friend prepare for a skateboard exhibition by determining if the planned program will work. Your friend will take a running start and jump onto a heavy-duty 7.0 kg stationary skateboard. The skateboard will glide in a straight line along a short, level section of track, then up a sloped wall. The goal is to reach a height of at least 8.0 m above the ground before coming back down the wall. Your friend's maximum running speed is 7.0 m/s, and your friend has a mass of 68 kg. The wall has a slope of 53.1° with the ground. Can your friend perform this trick? Note you must show your work to get credit.

Solution outline:

Conservation of momentum comparing the skateboard + person momentum before and after the person jumps on the skateboard.

$$mv_1 = (m + M)v_2$$

Conservation of energy comparing the skateboard + person energy from just after the person jumps on the skateboard to when it reaches its highest point on the wall.

$$\frac{1}{2}(m + M)v_2^2 = (m + M)gh$$

Difficulty calculation

Length of Problem Statement: 2 (117 words)

Problem Context: 1 (Familiar situation and familiar and concrete objects for these students)

Physics Principles: 1 (Conservation of energy, one-dimension conservation of momentum)

Mathematical Complexity: 2 (Two equations to solve for 2 unknowns)

Average: 1.5

Difficulty explanation

Riding a skateboard up a sloped ramp is a familiar situation with concrete and familiar objects for most students, since most of them have participated in or seen this situation.

This problem requires the use of both conservation of energy and conservation of momentum in one dimension. The substance ontology applies to both principles, and there is no extra mathematical sophistication. So, the difficulty of the physics principles is level 1 for each. The two principles do not add. The difficulty of having two principles rather than 1 is taken into account in the increased mathematical complexity.

Problem 2:

(a) Two identical stars of mass M are in circular orbits around their CM. Show that

$$T^2 = 2\pi^2 R^3/GM$$

where R is the distance between the stars and T is the period of rotation.

(b) Now consider a star and a satellite with unequal masses m and M and show, in the case of circular orbits, that

$$T^2 = 4\pi^2 R^3/G(M+m)$$

(c) Starting from the circular orbit you found in part (b), in the case where m << M and the distance between star and satellite is R, find the escape speed for the satellite.

Solution outline:

Part (a):

Apply Newton's second law using the gravitational force law and the acceleration for uniform circular motion. Each star goes around the center of mass of the system which is equidistant from the two stars.

$$\frac{GMM}{R^2} = \frac{Mv^2}{\frac{R}{2}}$$

Use the definition of speed in the case of a constant speed as the star goes around a complete orbit.

$$T = \frac{\pi R}{v}$$

Part (b):

Determine the position of the center of mass of the star satellite system. Use the definition of center of mass.

$$mr_1 = Mr_2$$
$$r_1 + r_2 = R$$

Apply Newton's second law and the definition of constant speed as in part (a)

$$\frac{GMm}{R^2} = \frac{mv_1^2}{r_1}$$
$$T = \frac{2\pi r_1}{v_1}$$

Part (c):

When m << M, $r_1 = R$

Apply conservation of energy to the star + satellite system taking the initial position of the satellite as its orbit around the star and the final position as infinitely far away from the star. Use conservation of energy and assume that the kinetic energy of the satellite at infinity is 0.

$$\frac{1}{2}mv_{escape}^2 - \frac{GMm}{R} = 0$$

Difficulty calculation:

Length of Problem Statement: 2 (103 words)

Problem Context: 3 (Unfamiliar situation, concrete but unfamiliar objects)

Physics Principles: 4 (Force with circular motion, Negative energy)

Mathematics Complexity: 4 (4 equations for part b)

Average: 3.25

Difficulty explanation:

The stars are concrete objects for students although not as familiar as the skateboard mentioned in previous problem. The gravitational force between stars cannot be considered constant as is the gravitational force on earth. That makes using the gravitational force in this situation unfamiliar for these students giving the context a rating of 3. The physics principles for parts (a) and (b) requires force and circular motion, rated as 4 on Table 2. The negative energy with a non-constant field required in part (d) was rated as 3 according to Table 1. Assuming that a student treats each part independently, the final rating for physics principles is 4, the highest of the parts.

Problem 3:

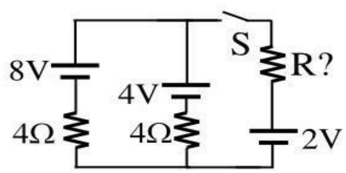
In the circuit shown below, after the switch “S” is closed, the current through the 8.0 V battery increases by 0.20 A as compared to its value when the switch had been open.

(a) What is the unknown resistance R?

After the switch is closed.

(b) How much power is dissipated in the resistor with resistance R?

(c) What is the total net power supplied to the circuit by the batteries (note batteries that produce power supply positive power, and batteries that consume power supply negative power)?



Solution outline:

Part (a):

Before the switch is closed:

Apply conservation of energy around the complete circuit.

$$8V - 4V = (4\Omega + 4\Omega)I_{before}$$

After the switch is closed:

The current through the 8V battery is now

$$I_1 = I_{before} + 0.2A$$

Apply conservation of current

$$I_1 = I_2 + I_3$$

Apply conservation of energy around each of the two circuit loops.

$$8V - 4V - (4\Omega)I_2 - (4\Omega)I_2 = 0$$

$$4V - RI_3 - 2V - (4\Omega)I_2 = 0$$

Part (b):

Use the definition of power for an ohmic object.

$$P_{in\ R} = I_3^2 R$$

Part (c):

Use the definition of power:

$$P = (8V)I_1 - (4V)I_2 - (2V)I_3 = 3.6W$$

Difficulty calculation:

Length of Problem Statement: 2 (89 words)

Problem Context: 1 (circuits with battery and resistors, familiar situation, concrete and familiar objects)

Physics Principles: 1 (conservation of energy in a closed system, conservation of charge in a closed system)

Mathematics Complexity: 5 (5 equations needed in part (a))

Average: 2.25

Difficulty explanation:

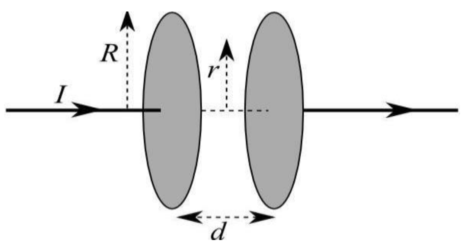
This problem involves direct current circuits with batteries and resistors, the most familiar situation with the most concrete and familiar objects in E&M. There are no unfamiliar objects such as inductors or capacitors and no abstract entities such as electric or magnetic fields or infinitely long wires. The physics principles are conservation of energy and charge in a closed system both of which make direct use of the substance ontology. No sophisticated mathematical ideas are needed.

Problem 4:

A parallel-plate capacitor has circular plates of radius R=0.30 m. Its plates are separated by a distance d = 0.10 mm. The capacitor is being charged with a constant current I =7.0 A.

a) What is the magnitude of the magnetic field between the plates at a distance r = 0.20 m from the central axis of the capacitor?

b) If you are looking down the axis of the capacitor with the positive plate closer to you and the negative plate further from you, in which direction will the B-field loop around (clockwise or counterclockwise)?



Solution outline:
Part (a):

Use Ampere's law for a displacement current

$$2\pi rB = \mu_0\varepsilon_0\pi r^2\frac{dE}{dt}$$

Use the relationship of voltage to electric field in a capacitor:

$$Ed=V$$

Use the definition of capacitance

$$C=Q/V$$

Use the definition of current

$$I = \frac{dQ}{dt}$$

Use the relationship of capacitance to the geometry of the capacitor

$$C=\varepsilon_o\pi R^2/d$$

get

$$B = \frac{\mu_o Ir}{2\pi R^2}$$

Alternative Solution outline:

Part (a):

Use Ampere's law for a displacement current

$$2\pi rB = \mu_o \varepsilon_o \pi r^2 \frac{dE}{dt}$$

Use Gauss' law for electric field

$$E = \frac{q}{\varepsilon_o \pi r^2}$$

Use the definition of current

$$I = \frac{dq}{dt}$$

get

$$B = \frac{\mu_o I r}{2\pi R^2}$$

A third solution is less straight-forward but is much shorter. It relies on assuming that there is an effective current between the parallel plates of the capacitor. This solution relies on a logical argument about why one can treat the current this way even though no such physical current exists. This solution is not expected for introductory students but is given for completeness.

Another alternative solution outline:

Part (a):

Use Ampere's law for a real current that is assumed to be the same as the displacement current

$$2\pi rB = \mu_o I \pi r^2 / \pi R^2$$

get

$$B = \frac{\mu_o I r}{2\pi R^2}$$

Part (b):

Use the Ampere's law right hand rule and conventional current that goes from + to - B is clockwise

Difficulty calculation:

Length of Problem Statement: 2 (98 words)

Problem Context: 5 (An unfamiliar situation of a circuit with no source and a non-physical current in a circuit gap between two capacitor plates, a capacitor is an unfamiliar object, and abstract entities of electric and magnetic fields.)

Physics Principles: 5 (Time varying electric field causes a magnetic field)

Mathematics Complexity: 5 (5 equations for solution 1) or 4 (4 equations for solution 2)

Average: 4.25 (solution 1) or 4.0 (solution 2)

Difficulty explanation:

For this problem, the context is the unfamiliar situation of a circuit with no source and a displacement current in the gap of a capacitor. There is also an unfamiliar object, the capacitor, and abstract electric and magnetic fields. The physics principles require the mathematical concept of an integral over a Gaussian surface, an integral along an Amperian path, and a time varying quantity, the electric field. Each of the three solutions requires a different number of equations for the solutions. Because solution 2 doesn't use the distance between the two plates, and solution 3 requires an unusual interpretation of current, I chose solution 1 as the most likely path to a solution that would be attempted by a student.

3.4 Summary

In this chapter I have given a short description of the process by which I arrived at the categories for the difficulty measurement. These categories were based on the application of the theories summarized in Chapter 2 and an attempt to make reproducible categories for the problems used on physics tests at the University of Minnesota. None of those problems were used in the validity analysis in the next chapter. The final categories have an overlap with those reported in previous studies but are not the same. I then gave the criteria for scoring the problem difficulty in each category and gave examples to illustrate how I applied those criteria to five of the problems used in the validity analysis.

Chapter 4 Validity and Reliability

After developing the four-category difficulty measure for physics problems, I gathered evidence to determine its validity and reliability. Validity is the extent to which the scores from a measure represent the intended quantity. Here I consider three basic types of validity: construct validity, criterion validity and internal structure of the measure [44]. The reliability is the extent to which the scoring is independent of the person doing it.

4.1 Construct Validity

Construct validity is the extent to which a measure has a theoretical backing [45]. This validity is based on the construction of the four categories of difficulty described in Chapter 3 that reflect the theories described in Chapter 2.

4.2 Criterion Validity

Criterion validity is the extent to which the scores on a measure are correlated with other variables (known as criteria) that one would expect to be related [44]. The criterion validity evidence for the difficulty measuring instrument is the extent to which its score agrees with student's performance when solving the physics problems typically used for grading purposes. Here one assumes that the more difficult the problem, the lower the student performance will be on that problem. As shown in Appendix A, the distributions of student numerical grades for the problems used in this study are not simple which makes it difficult to characterize them with a single number. Nevertheless, how well the student population does on a problem can be thought of as the total number of grading points that all of the students receive for their individual problem solutions. Since the average grade gives the same information as the sum of all the grading points students received on that problem, I will take the average numerical grade as the measure of student performance on that problem. Using the average to characterize student performance is also used in all previous studies reviewed in chapter 2.

4.2.1 Experimental Environment

I compared the difficulty scores of final exam problems from four semesters of classes of calculus based introductory physics for engineering and physical science students at the University of Minnesota to the grades for the student solutions of those same problems assigned by the graduate teaching assistants (TAs) in those courses. The problem grades are a good indicator of student performance according to previous work that compared TA grading of problems to an evaluation how well the student performance approached expert behavior [11]. Two of the problem samples, fall 2016 and fall 2014, were from multiple lecture sections teaching the first semester of the introductory physics course and addressed Classical Mechanics. The other two problem samples, spring 2016 and spring 2012, were from multiple lecture sections teaching the second semester of the course and addressed Electricity and Magnetism. These problem samples were chosen because the classes were not sequential and therefore had fewer continuing instructors, and thus more likely to have a larger diversity of problem styles.

These semesters also had a full set of data, individual grades for each problem on the final examination, available. In all cases, the final exam had 5 open response problems that were the same for every lecture section in that semester.

In fall 2016, the first semester of introductory physics (Mechanics) had 3 lecture sections with 900 students in total. In fall 2014, the first semester of the course had 5 lecture sections with a total of 858 students. Each lecture section within a semester had approximately the same number of students. In each semester, all the sections used the same problems and took the exam at the same time. The difficulty measure was applied to all 5 open response final exam problems from each of the two semesters. The problems for each semester were different. These problems are given in Appendix B. They addressed topics that included kinematics, Newton's laws, conservation of energy, and conservation of momentum. Each of the 10 problems was graded independently for each lecture section by a different TA from that section on a scale from 0 to 25.

Exactly the same process was followed for the second semester of the introductory physics course (Electricity and Magnetism). In spring 2016, the course had 4 lecture sections with 963 students in total. In spring 2012, the second semester had 5 lecture sections with 831 students. These problems, given in Appendix B, addressed topics including electric circuits, Coulomb's law, Ampere's law, and Faraday's Law.

4.2.2 Results for Criterion Validity

An overall score for each problem's difficulty was calculated by assigning a score between 1 and 5 for each of the four categories of the difficulty measure and then averaging the scores from all the categories. Averaging the scores from 4 categories gives the same weight to each category in determining the problem difficulty. This assumes that each category has an equivalent effect on cognitive processing in the absence of information to the contrary. The single overall score was compared with the students' average grade for that problem.

As shown in Figure 1, the overall problem difficulty scores and the average student grade for the 20 final examination problems from two offerings of each semester of the introductory physics course are highly correlated. In this figure, each point represents a problem and its error bars is dominated by the systematic uncertainty of the average problem grade determined in the manner described later in this section. The gray area shows the 95% confidence level for a linear fit to the data which is the dashed line. The Pearson correlation is 90%.

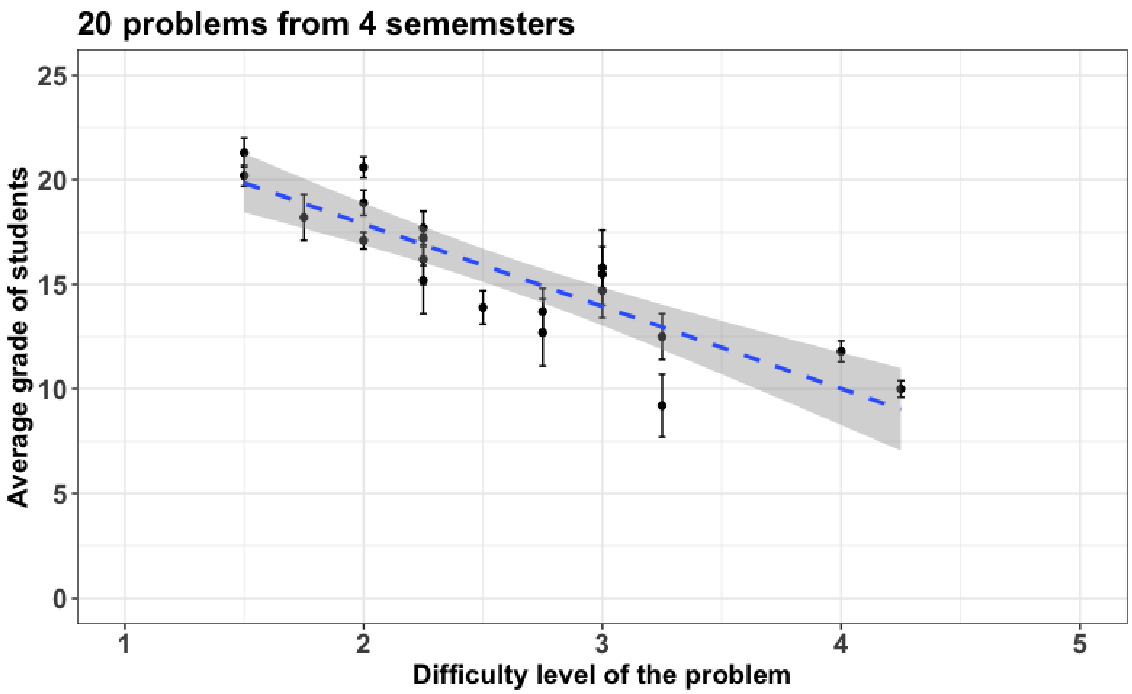


Figure 1: The relationship between the average student grade and the problem's difficulty score for 20 problems from 4 semesters. The dashed line is a linear fit to the data.

This extremely high correlation between difficulty scores and the average performance of students on each problem provides evidence of the validity of the difficulty measure. Details of the measurement for each semester are given below where the mechanics semesters of introductory physics are examined separately from the electromagnetism semesters. It is these separate measurements that are combined to give the results of the 17 large lecture sections displayed in Figure 1.

4.2.3 Studies with Introductory Mechanics

Fall 2016

The following tables and graphs show the relationship of the problem difficulty score and the student performance on each problem as determined by their numerical grade. Tables 6 – 8 give the measurements displayed on the graphs in Figures 2 and 3.

Table 6 gives the difficulty scores for each of the 5 mechanics problems on the final examination in Fall 2016 in each of the 4 categories of the measurement. These difficulty scores were determined by the method given in Chapter 3. The table also gives an overall difficulty score found by averaging the scores of the 4 categories.

	Problem 1	Problem 2	Problem 3	Problem 4	Problem 5
Length of Problem Statement	2	2	1	2	2
Problem Context	1	1	1	3	1
Physics Principle	1	3	3	4	3
Mathematical Complexity	2	2	2	4	2
Average Difficulty Level	1.5	2.25	1.75	3.25	2.0

Table 6: The problem difficulty measure on the fall 2016 final examination.

Table 7 gives the average grade for each problem out of 25 possible points for each lecture section. The uncertainties given are statistical, the standard error of the mean

(SEM). Note that these statistical uncertainties are similar for all the sections and all the problems.

Difficulty score	Average grade section 1 (297 students)	Average grade section 2 (304 students)	Average grade section 3 (299 students)
1.5 (Problem 1)	20.3±0.4	21.1±0.3	22.6±0.3
2.25 (Problem 2)	12.4±0.3	18.1±0.4	15.3±0.5
1.75 (Problem 3)	16.6±0.4	17.7±0.4	20.3±0.3
3.25 (Problem 4)	8.2±0.4	7.1±0.4	12.2±0.4
2.00 (Problem 5)	17.6±0.4	16.4±0.4	17.5±0.3

Table 7: The average student grade for each of the final examination problems in fall 2016 by course section. The uncertainties shown are statistical only.

Table 8 shows the combined student grade for all of the 3 sections for each of the final examination problems. Each average over the grading of 3 different TAs. If there were a similar grading standard among the graders, the averaging process should enhance the signal and reduce the noise caused by individual grading differences. Deviations of the graded student performance outside of that predicted by statistics among the sections could come from differences in grading policies, teaching, or student populations. This systematic uncertainty was estimated by determining the difference between the average grade of each section and average grade of all students divided by the square root of number of sections minus 1. The systematic uncertainty, given in Table 8 and illustrated by the spread in Figure 2, is much larger than the statistical uncertainty and dominates the measurement when they are added in quadrature.

Difficulty score	Average grade all 3 sections (900 students)	Systematic uncertainty
1.5 (Problem 1)	21.3±0.2	0.7
2.25 (Problem 2)	15.2±0.3	1.6
1.75 (Problem 3)	18.2±0.2	1.1
3.25 (Problem 4)	9.2±0.3	1.5
2.00 (Problem 5)	17.1±0.2	0.4

Table 8: The average student grade for all students from 3 sections for each of the final examination problems in fall 2016. The uncertainties shown with the average grade are statistical only.

Figure 2 illustrates the extent of the systematic error for all sections on a single graph. The color of each data point indicates the section and the error bars are statistical. The spread of grades for the same problem outside of that predicted by statistics is evident. This spread is caused by systematic uncertainties due to each section having different problem graders, different instructors, and, possibly, different student populations.

As discussed previously, averaging each problem's grade over the three sections should reduce the effect of the differences of TA grading. These results are shown in Figure 3 along with the linear fit to the points and a shaded region representing the 95%

confidence level for the fit. The result of the linear regression is a Pearson correlation of 99% with the null hypothesis probability of 0.001. The error bar shown for each point on this graph is the square root of the sum of the statistical uncertainty in quadrature with the systematic uncertainty.

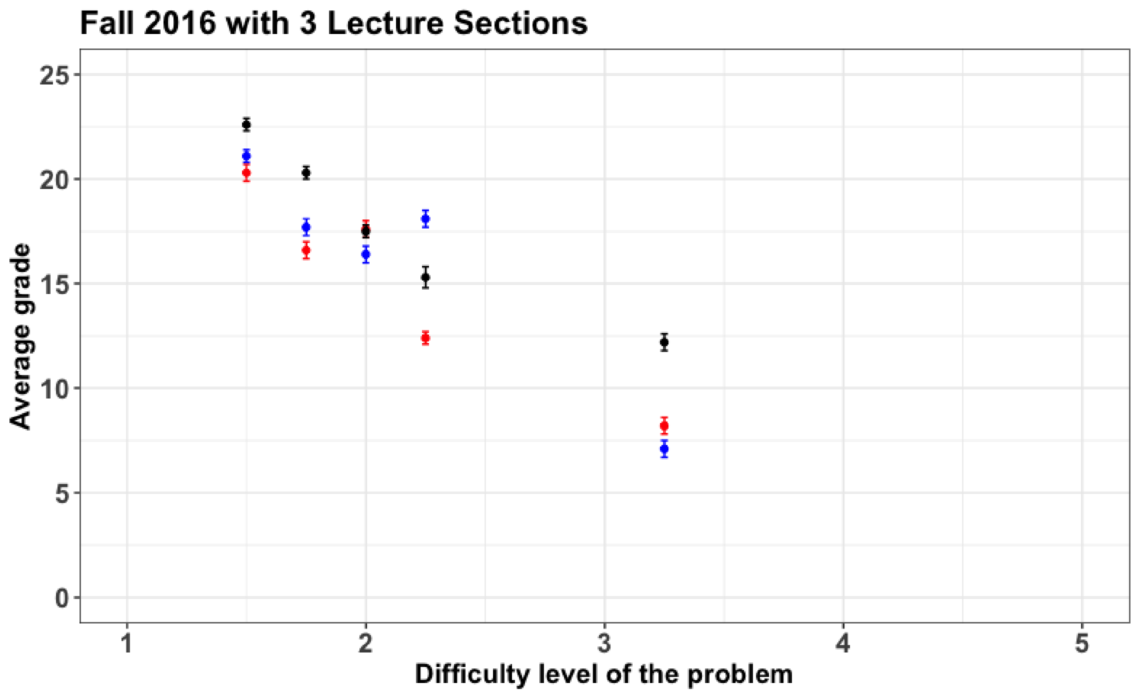


Figure 2: The relationship between the average grade for each final exam problem and its difficulty level for all lecture sections in Fall 2016. The colors designate the different lecture sections. The error bars represent the statistical uncertainty.

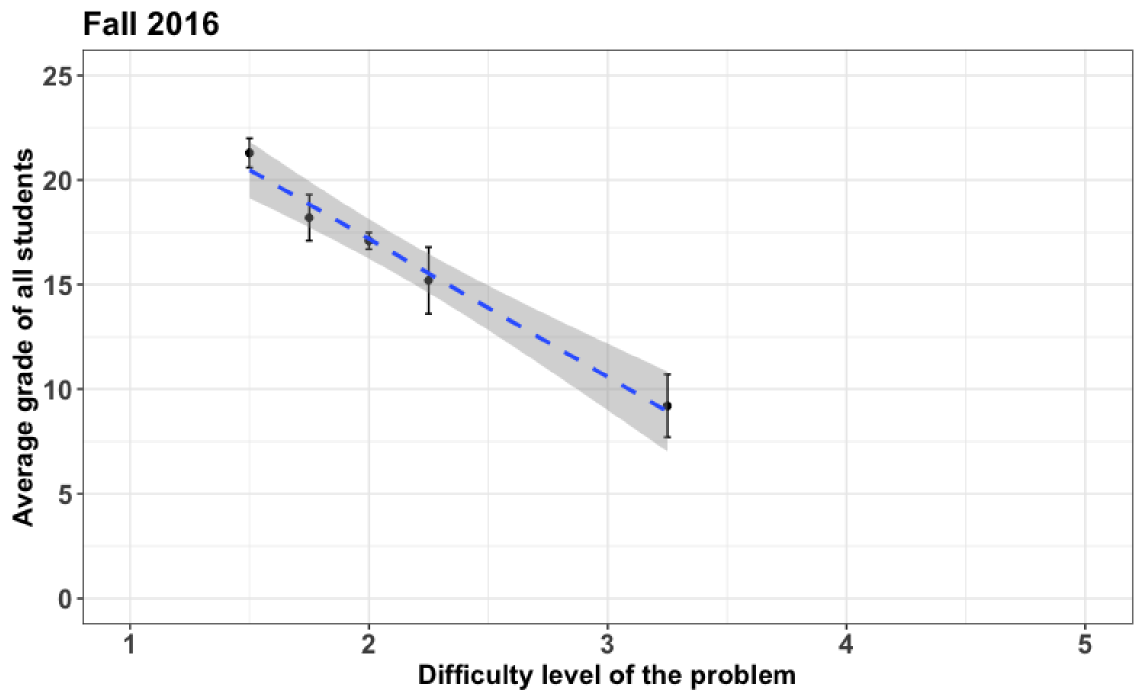


Figure 3: The relationship between the average grade for each final exam problem and its difficulty level for all lecture sections combined in Fall 2016. The error bars are dominated by the systematic uncertainty of the student grade illustrated by Figure 2 and calculated in Table 8.

The study was repeated for the problems given on the final examination of the first semester of the introductory physics course in Fall 2014. In this sample there are 5 different lecture sections and thus 5 different professors each with a different set of TAs. The result was similar and is given below.

Fall 2014

Below are the data for the 5 different open response problems used by all the lecture sections in Fall, 2014.

Tables 9 – 11 and Figures 4 and 5 give the same information as the corresponding Tables 6 - 8 and Figures 2 and 3 for the Fall 2016 sections.

	Problem 1	Problem 2	Problem 3	Problem 4	Problem 5
Length of Problem Statement	3	2	3	2	3
Problem Context	1	2	1	2	1
Physics Principle	2	3	4	4	3
Mathematical Complexity	2	2	3	2	4
Average Difficulty Level	2	2.25	2.75	2.5	2.75

Table 9: The difficulty scores of problems on the final examination in fall 2014.

Difficulty score	Average grade section 1 (177 students)	Average grade section 2 (198 students)	Average grade section 3 (207 students)	Average grade section 4 (171 students)	Average grade section 5 (111 students)
2 (Problem 1)	22.1±0.2	21.0±0.3	19.1±0.4	20.4±0.4	20.5±0.5
2.25 (Problem 2)	19.9±0.4	17.3±0.4	15.4±0.3	18.3±0.5	19.0±0.5
2.75 (Problem 3)	12.7±0.5	12.0±0.5	14.3±0.4	16.0±0.5	6.5±0.8
2.5 (Problem 4)	15.7±0.5	12.6±0.6	14.0±0.6	12.3±0.6	16.0±0.9
2.75 (Problem 5)	12.2±0.6	16.2±0.5	12.9±0.6	15.8±0.6	10.2±0.9

Table 10: The average student grade for each of the final examination problems in fall 2014 by course lecture section. The uncertainties shown are statistical only.

Difficulty score	Average grade all 5 sections (864 students)	Systematic uncertainty
2 (Problem 1)	20.6±0.2	0.5
2.25 (Problem 2)	17.7±0.2	0.8
2.75 (Problem 3)	12.7±0.2	1.6
2.5 (Problem 4)	13.9±0.3	0.8
2.75 (Problem 5)	13.7±0.3	1.1

Table 11: The average student grade for all students from 5 sections for each of the final examination problems in fall 2014. The uncertainties shown in the column with the average grades are statistical only.

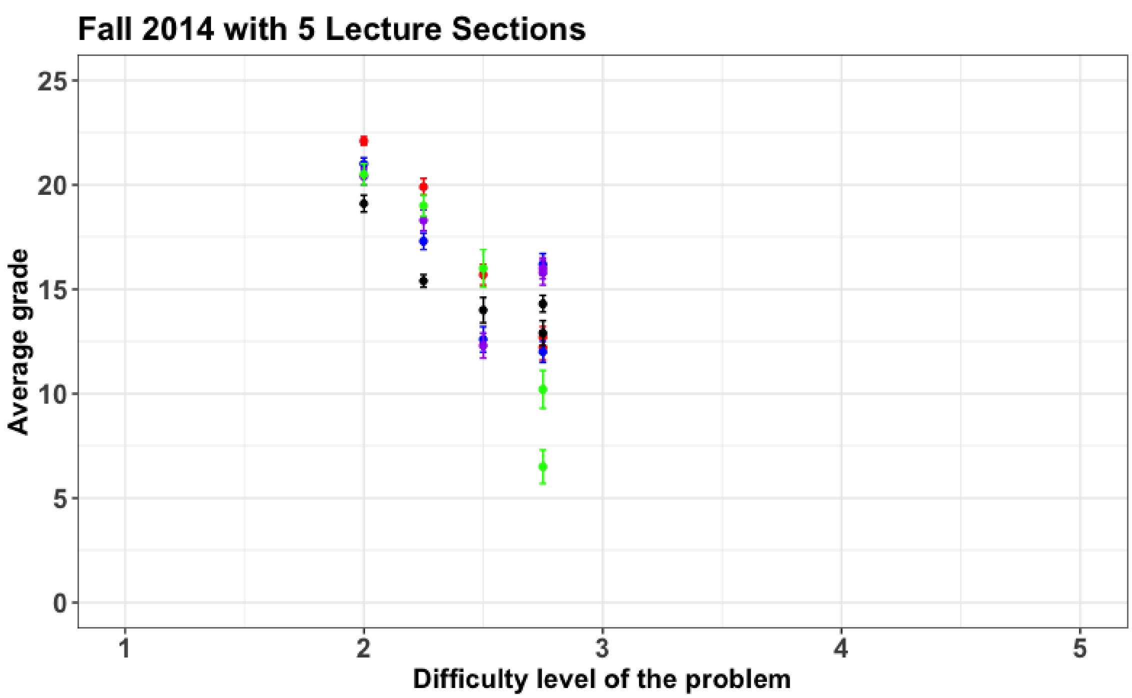


Figure 4: The relationship between the average grade for each final exam problem and its difficulty level for all lecture sections in Fall 2014. Note that there are two problems with the highest difficulty level. The colors designate the lecture sections. The error bars represent the statistical uncertainty.

As before, the effect of systematic differences in grading, given in Table 11 and shown by the spread in Figure 4, is reduced by averaging the scores of the five sections. The result of this averaging is shown in Figure 5 along with the linear fit to the points and the shaded region representing the 95% confidence level for the fit. The result of the linear regression is a Pearson correlation of 97% with the null hypothesis probability of 0.007. The error bars shown are a combination of the statistical and systematic uncertainty which is dominated by the systematic uncertainty. These combined results for Fall 2014 show the same behavior as those for Fall 2016.

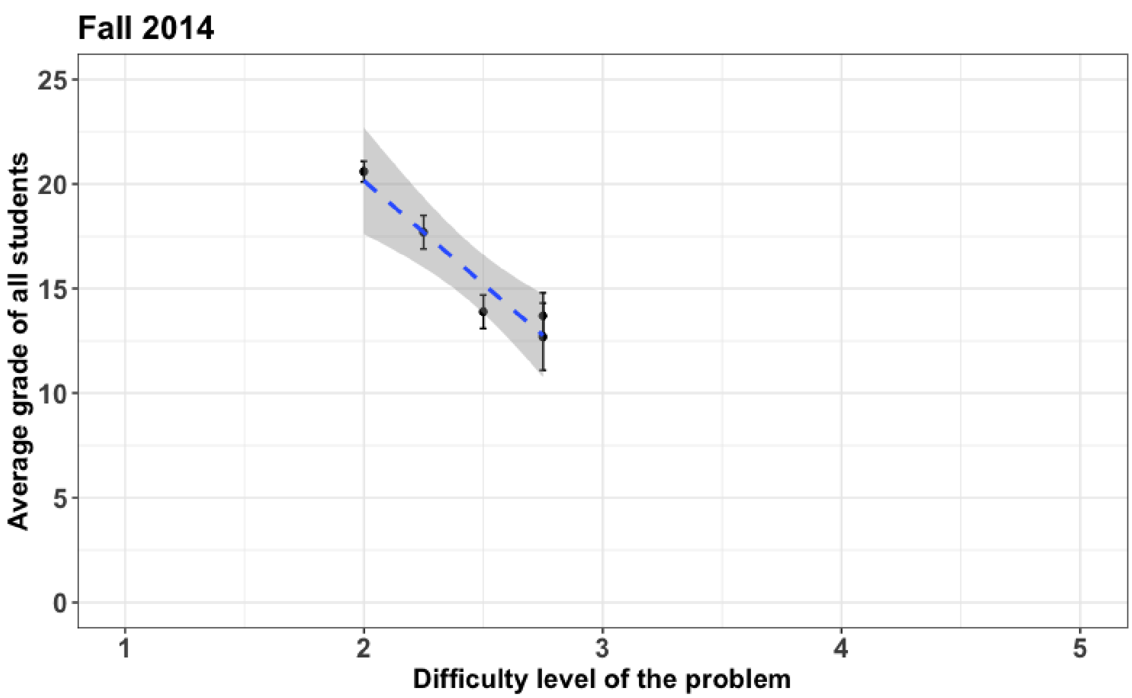


Figure 5: The relationship between the average grade for each final exam problem and its difficulty level for all lecture sections combined in Fall 2014. The error bars are dominated by the systematic uncertainty of the student grade illustrated by Figure 4 and calculated in Table 11.

Having determined that the difficulty measurement for 10 different introductory physics mechanics problems constructed by 8 different professors is highly correlated with student performance on those problems, the same test was applied to two different electromagnetism semesters. All test and analysis procedures were the same as described previously.

4.2.4 Studies with introductory Electricity and Magnetism

The analysis of the correlation between the student performance on final exam problems and their difficulty level for two instances of the semester of introductory physics teaching electricity and magnetism, is given below. The difficulty level was determined by the method given in Chapter 3. These difficulty scores tend to be higher than the mechanics problems as seen by a comparison of Table 12 with Table 9. This is not surprising because the electromagnetism semester is considered more difficult than mechanics by both students and instructors.

Spring 2016

Below is the analysis of the 5 open response problems from the 2016 Spring Final Exam. Tables 12 – 14 give the data displayed on the graphs in Figures 6 and 7.

	Problem 1	Problem 2	Problem 3	Problem 4	Problem 5
Length of Problem Statement	3	2	4	5	2
Problem Context	3	1	4	3	5
Physics Principle	3	1	4	3	5
Mathematical Complexity	3	5	4	1	5
Average Difficulty Level	3	2.25	4	3	4.25

Table 12: The difficulty measure scores for each of the 5 problems on the final examination in spring 2016.

Difficulty score	Average grade section 1 (277 students)	Average grade section 2 (259 students)	Average grade section 3 (243 students)	Average grade section 4 (184 students)
3 (Problem 1)	21.2±0.2	11.4±0.4	13.9±0.3	15±0.5
2.25 (Problem 2)	15.6±0.4	15.8±0.4	19.4±0.3	13.7±0.6
4 (Problem 3)	10.6±0.4	11.8±0.4	13±0.4	11.8±0.5
3 (Problem 4)	16.7±0.4	13.7±0.4	14.4±0.4	13.8±0.5
4.25 (Problem 5)	9.6±0.4	10±0.6	9.7±0.4	11.2±0.5

Table 13 :The average student grade for each of the final examination problems by course lecture section in Spring 2016. The uncertainties shown are statistical only.

Difficulty score	Average grade of all 4 sections (963 students)	Systematic uncertainties
3 (Problem 1)	15.5±0.2	2.1
2.25 (Problem 2)	16.2±0.2	1.2
4 (Problem 3)	11.8±0.2	0.5
3 (Problem 4)	14.7±0.2	0.7
4.25 (Problem 5)	10±0.2	0.4

Table 14: The average student grade for each of the final examination problems for all students from the 4 sections in spring 2016. The uncertainties shown in the column with the average grades are statistical only.

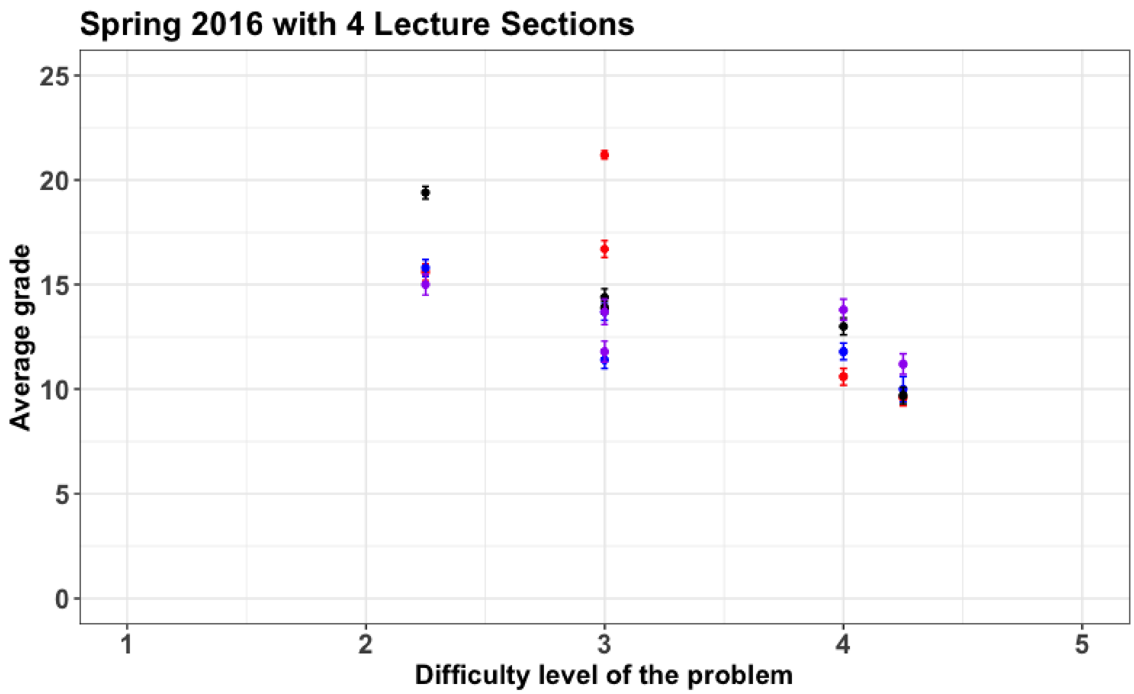


Figure 6: The relationship between the average grade for each final exam problem and its difficulty level for all lecture sections in Spring 2016. The colors designate the different lecture sections. The error bars represent the statistical uncertainty.

Averaging over the five different sections reduces the effect of the grading systematics. The result of this process is shown in Figure 7 along with the linear fit to the points and the shaded region representing the 95% confidence level for the fit. A linear regression gives a Pearson correlation of 96% with the null hypothesis probability of 0.008. As before, the error bars shown are a combination of the statistical and systematic uncertainty which is dominated by the systematic uncertainty. This behavior is similar to that of the mechanics semesters but covers a wider range of problem difficulty.

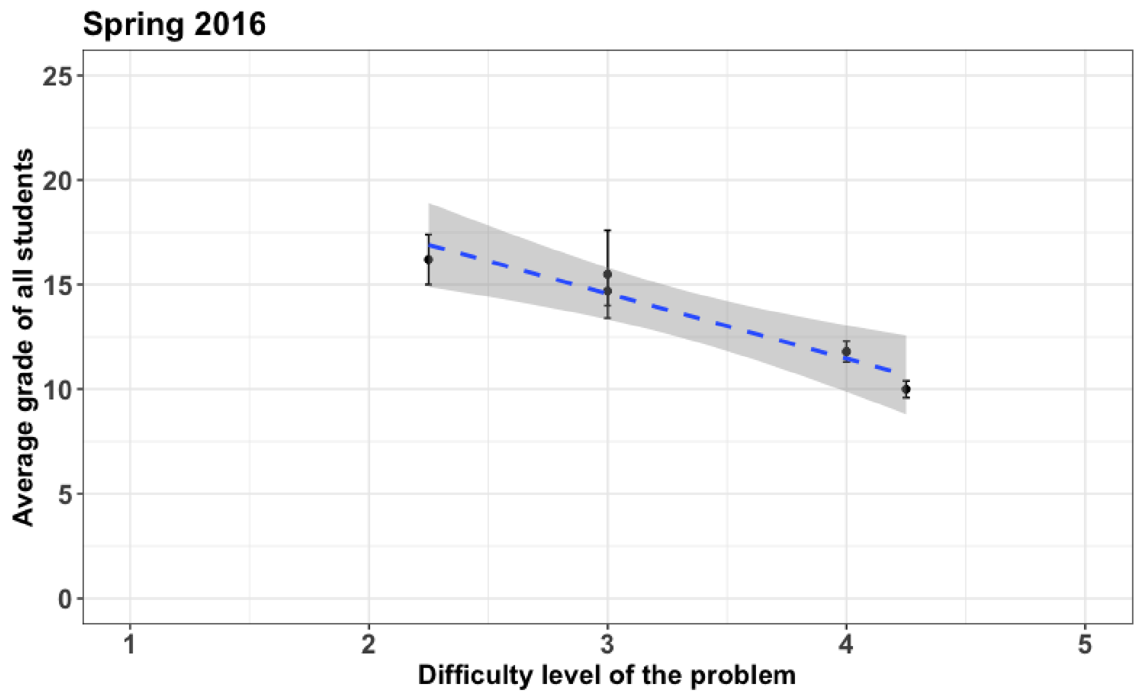


Figure 7: The relationship between the average grade for each final exam problem and its difficulty level for all lecture sections combined in Spring 2016. The error bars are dominated by the systematic uncertainty of the student grade illustrated by Figure 6 and calculated in Table 14.

Spring 2012

Below is the analysis of the 5 open response problems from the 2012 Spring Final Exam. Tables 15 – 17 give the data for the Spring 2012 sections that is shown in Figures 8 and 9.

	Problem 1	Problem 2	Problem 3	Problem 4	Problem 5
Length of Problem Statement	3	2	3	2	3
Problem Context	1	2	2	3	3
Physics Principle	3	1	1	4	3
Mathematical Complexity	1	1	3	3	4
Average Difficulty Level	2	1.5	2.25	3	3.25

Table 15: The difficulty measure scores for each of the 5 problems on the final examination in spring 2012.

Difficulty score	Average grade section 1 (154 students)	Average grade section 2 (152 students)	Average grade section 3 (190 students)	Average grade section 4 (165 students)	Average grade section 5 (170 students)
2 (Problem 1)	17.8±0.5	19.0±0.4	18.4±0.4	18.8±0.4	20.5±0.3
1.5 (Problem 2)	21.7±0.5	21.8±0.5	19.9±0.5	19.4±0.6	18.7±0.5
2.25 (Problem 3)	17.8±0.5	18.2±0.5	17.7±0.4	12.2±0.4	20.1±0.4
3 (Problem 4)	14.2±0.5	14.6±0.5	19.6±0.4	14.2±0.5	15.5±0.6
3.25 (Problem 5)	12.0±0.6	11.2±0.5	13±0.4	9.7±0.4	16.4±0.4

Table 16: The average student grade for each of the final examination problems by lecture section in Spring 2012. The uncertainties shown are statistical only.

Difficulty score	Average grade of all 5 sections (831 students)	Systematic uncertainties
2 (Problem 1)	18.9±0.2	0.6
1.5 (Problem 2)	20.2±0.3	0.5
2.25 (Problem 3)	17.2±0.3	1.3
3 (Problem 4)	15.8±0.3	1
3.25 (Problem 5)	12.5±0.3	1.1

Table 17: The average student grade for all students from the 5 sections for each of the final examination problems in Spring 2012. The uncertainties shown in the column with the average grades are statistical only.

Collecting all the data on a single graph in Figure 8 again shows the extent of the systematic uncertainties. The color of each data point indicates the section and the error bar represents the statistical uncertainty.

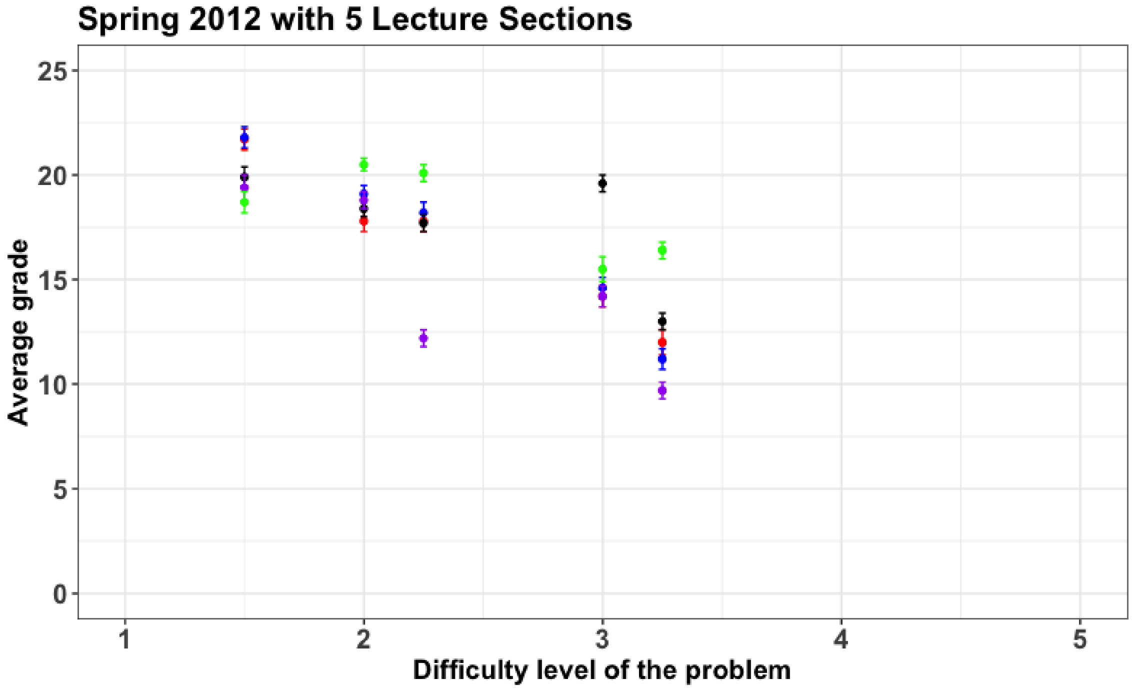


Figure 8: The relationship between the average grade for each final exam problem and its difficulty level for all lecture sections in Spring 2012. The colors designate the different lecture sections. The error bars represent the statistical uncertainty.

Again, assuming that the systematic uncertainties are caused by random differences in the students, pedagogies, and grading of the sections, their effect is reduced by averaging the student grades from all the sections for each problem. The result of this averaging process is shown in Figure 9. The shaded region represents the 95% confidence level for a linear fit of the dependence of the average student grade on the problem difficulty score. The result of the linear regression is a Pearson correlation of 97% with the null hypothesis probability of 0.01.

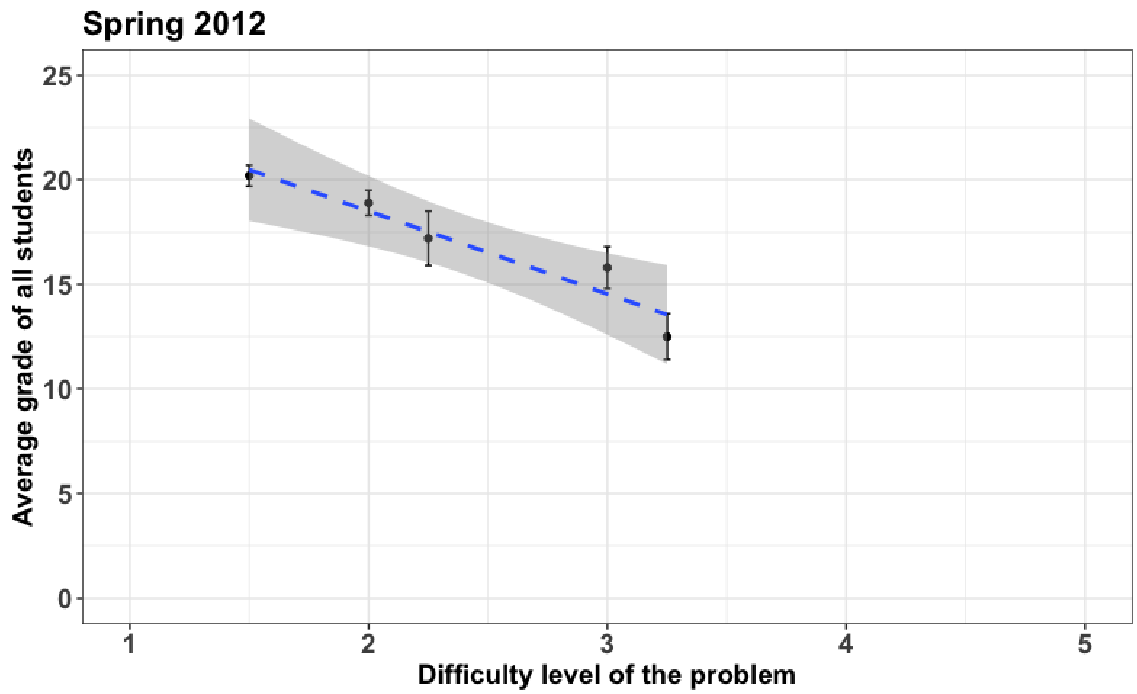


Figure 9: The relationship between the average grade for each final exam problem and its difficulty level for all lecture sections combined in Spring 2012. The error bars are dominated by the systematic uncertainty of the student grade illustrated by Figure 8 and calculated in Table 17.

4.2.5 Comparison of the problem difficulty measure of mechanics topics and E&M topics

A summary of the average student final exam problem scores for all semesters of the introductory physics course is shown in Figure 10. This graph is the same as that shown in Figure 1 except that the mechanics semesters are shown in green and the E&M semesters in red. The gray area shows the 95% confidence level for a linear fit to the data, the dashed line. The Pearson correlation is 90% with a probability that there is no correlation of less than 0.001.

This analysis shows that the problem difficulty measure is highly correlated with student performance on those problems as graded by standard classroom procedures when the TA have extensive preparation to teach the class [46]. The high correlation between the problem difficulty measure and student problem grade is consistent for both Mechanics problems and E&M problems demonstrating that the measure has construct validity

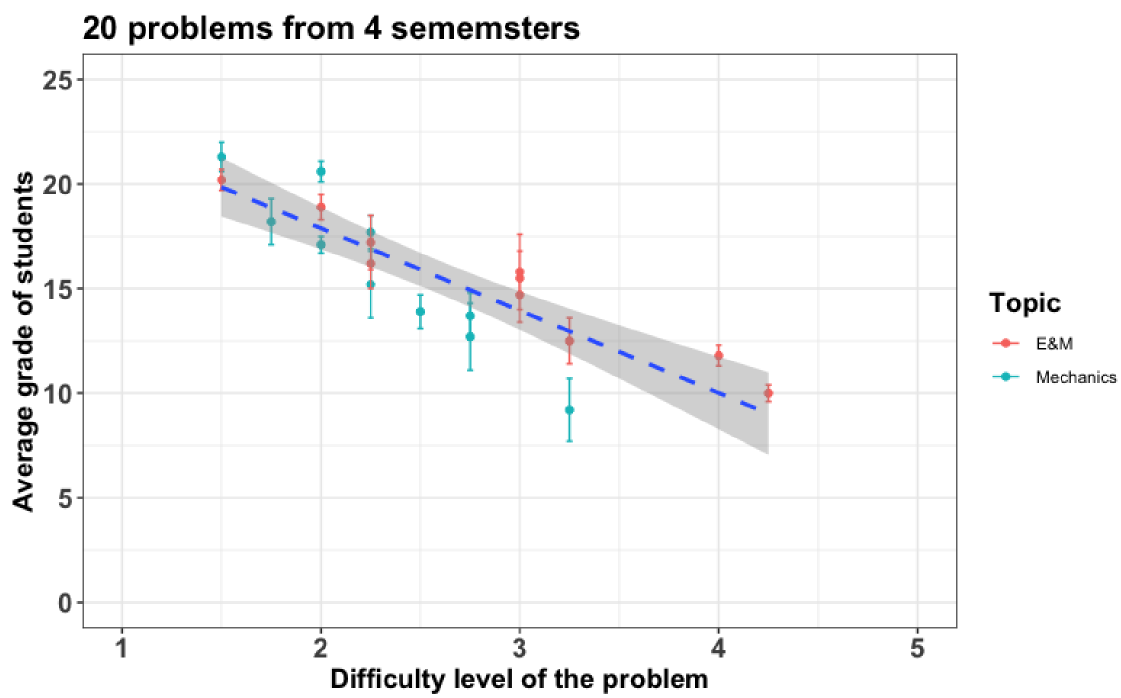


Figure 10: The relationship between the average student grade and the problem's difficulty score for 20 problems from 4 semesters. Mechanics problems are shown in green and E&M problems in red. The error bars shown are dominated by systematic uncertainty.

4.2.6 Strength of category correlations

To determine the internal structure of the four-category problem difficulty measure, I first looked at the correlation of each with student problem solving performance. Table 18 gives the correlation coefficient and probability of the null hypothesis for each individual category. As can be seen, there is no significant correlation between the problem length and the student grade for each problem.

	Length of Problem Statement	Problem Context	Physics Principle	Mathematical complexity
Correlation	-0.20	-0.63	-0.74	-0.67
Probability of null hypothesis	0.40	0.003	<0.001	0.001

Table 18： The correlation coefficient between each category and students’ grade

As a confirmation that the Problem Length category has little predictive power, I left it out of the problem difficulty calculation and found a correlation of 88% with a null hypothesis probability of <0.001. This is very close to the 90% correlation with all 4 categories. Figure 11 illustrates the success of the 3 categories difficulty measure.

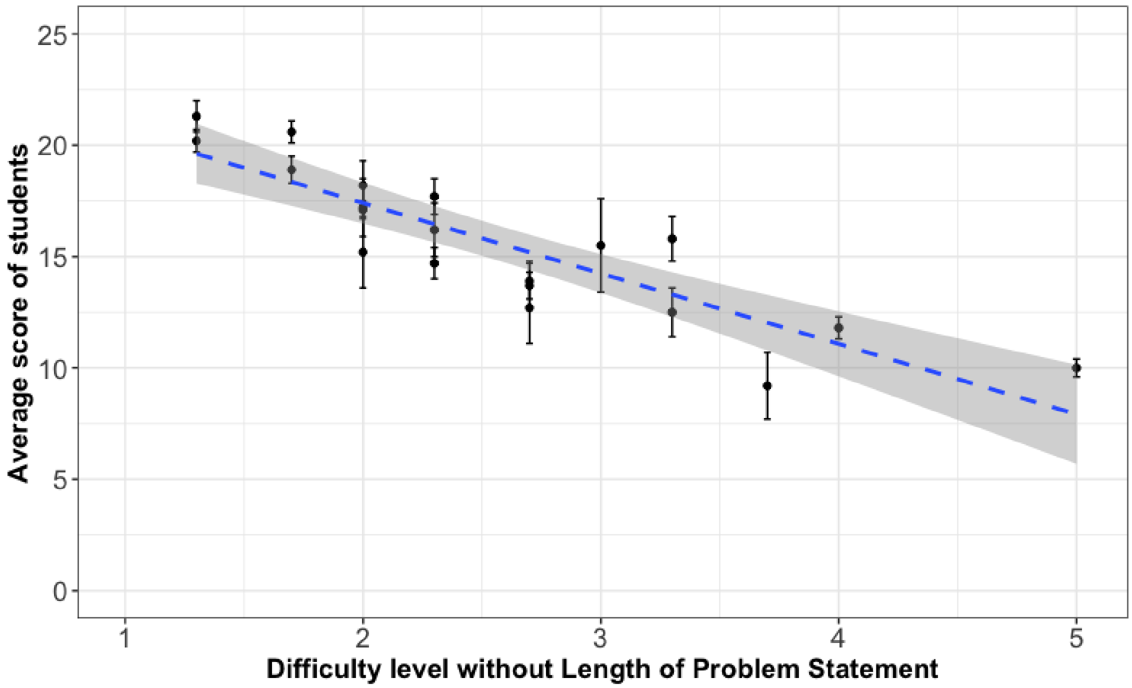


Figure 11: The relationship between the average TA grade and the problem's difficulty score with average of 3 categories without Length of Problem Statement. The error bars are dominated by systematic uncertainty.

4.3 Internal Structure

Another test of the validity of the determination of the problem difficulty is the investigation of the internal structure of the measure, in particular, the extent to which correlations between scores of the four categories agree with the expectation that the categories should be almost independent. The inter-category correlation matrix for the difficulty of all 20 problems used in the previous section is shown in Table 19. This correlation matrix is separated into the matrices for the 10 mechanics problems and the 10 E&M problems in Tables 20 and 21. All correlations reject the null hypothesis at the < 0.01 level.

These tables show that the categories are not independent. Problem Context is highly correlated with both Physics Principles and Mathematical Complexity. The connection between these two categories could be that instructors tend to write problems with difficult Physics Principles using unfamiliar and abstract contexts or it could mean that difficult Physics Principles arise primarily in those contexts. This seems to be especially true in E&M (0.79 correlation coefficient).

	Problem Context	Physics Principles	Mathematical Complexity
Problem Context	1	**0.53**	**0.43**
Physics Principles		1	**0.27**
Mathematical Complexity			1

Table 1819: Inter-category correlation coefficients between difficulty category scores. for both mechanics and E&M.

	Problem Context	Physics Principles	Mathematical Complexity
Problem Context	1	**0.51**	**0.37**
Physics Principles		1	**0.42**
Mathematical Complexity			1

Table 20: Inter-category correlation coefficients between difficulty category scores for mechanics.

	Problem Context	Physics Principles	Mathematical Complexity
Problem Context	1	**0.79**	**0.40**
Physics Principles		1	**0.25**
Mathematical Complexity			1

Table 21: Inter-category correlation coefficients between difficulty category scores for E&M.

Based on the strong cross correlations between Problem Context and Physics Principles, I eliminated the Problem Context category and used only Physics Principles and Mathematical Complexity to determine the problem difficulty. The result of the average of these two category problem difficulty measures gave a correlation of 88% with the average student problem solving grade with a null hypothesis probability of <0.001, essentially the same as using all 4 categories. Figure 12 is the graph showing this correlation.

For completeness, I also took the other combinations of two categories to correlate with student problem solving grade: Problem Context and Physics Principles gave a correlation of 78% with a null hypothesis probability of <0.001; and Problem Context and Mathematical Complexity gave correlation of 77% with a null hypothesis probability of <0.001. Both are lower than the 88% correlation using the average of Physics Principles and Mathematical Complexity to determine problem difficulty. The low correlation between these two categories of 0.25 also shows that they are nearly independent problem difficulty measures.

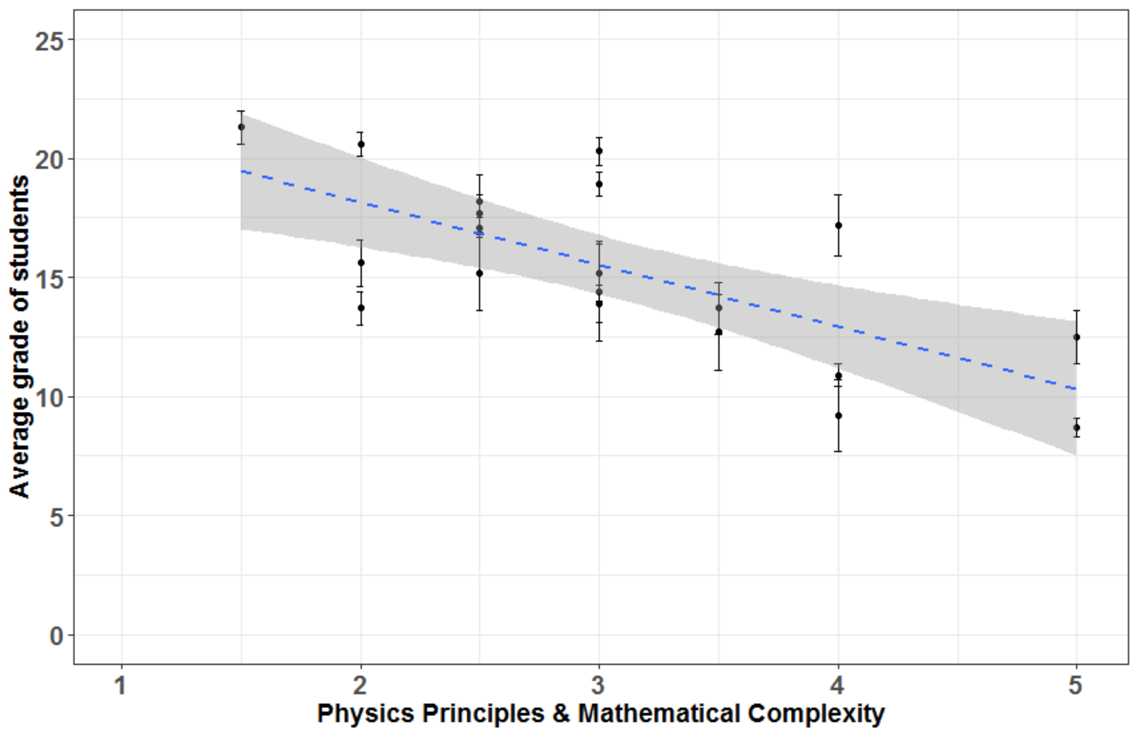


Figure 12: The relationship between the average student grade and the problem's difficulty score with average of physics principle and mathematical complexity.

In summary, I have described the evidence of construct validity, criterion validity, and validity of the internal structure of the problem difficulty measure. Based on these data, this measure simplifies to only 2 categories, Physics Principles and Mathematical Complexity which explain 77% of the student problem solving performance variance. However, this narrowing to 2 categories may be an artifact of the small range of instructors at one institution that constructed the final exam problems.

4.4 Inter-rater Reliability

Another important test of the problem difficulty measure is its reliability. Reliability refers to the consistency of different applications of a measure. Inter-rater reliability is the extent to which different users of the measure get consistent results. In order to obtain good agreement, raters must agree on the meaning of the scoring categories as well as the levels within each category. To assess reliability, I conducted a study with 3 raters who had considerable experience in physics education research, one of whom was myself, and the other two raters were both graduate students in physics education research and experienced teaching assistants. One of these was also an experienced high school teacher.

Since this test was done before the analysis of the student problem solving data, the raters used all four categories. They used the categories to determine the difficulty of 16 different free-response final exam problems, 8 from the calculus-based introductory mechanics course and 8 from the calculus-based introductory electricity and magnetism course, which they scored over a period of one month. Each rater independently scored the 16 problems and recorded their individual scores. After rating the difficulty of the first four problems, the justifications for the individual ratings were discussed and, after this discussion, each rater could adjust their difficulty rating. Then the raters met one week later and discussed their ratings for the next four problems and so on until the sixteen problems were complete. Adjustments to the difficulty ratings after discussion were both small and rare. The results of the independent scorings after discussion for all 16 problems were then analyzed to determine reliability.

Figures 13 and 14 show the comparison the problem difficulty score of each of the other raters with my score (rater 1). The Spearman correlation for one pairwise comparison was 94%, Figure 13, and 95% for the other, Figure 14. Figure 15 shows the comparison between the two other raters. These ratings have a Spearman correlation of 95%. In the Figures, each point is a different problem and the dashed line is a linear fit to the data. The 95% confidence level of that fit shown as the gray region. In all cases, the probability of no correlation, the null hypothesis, was < 0.001.

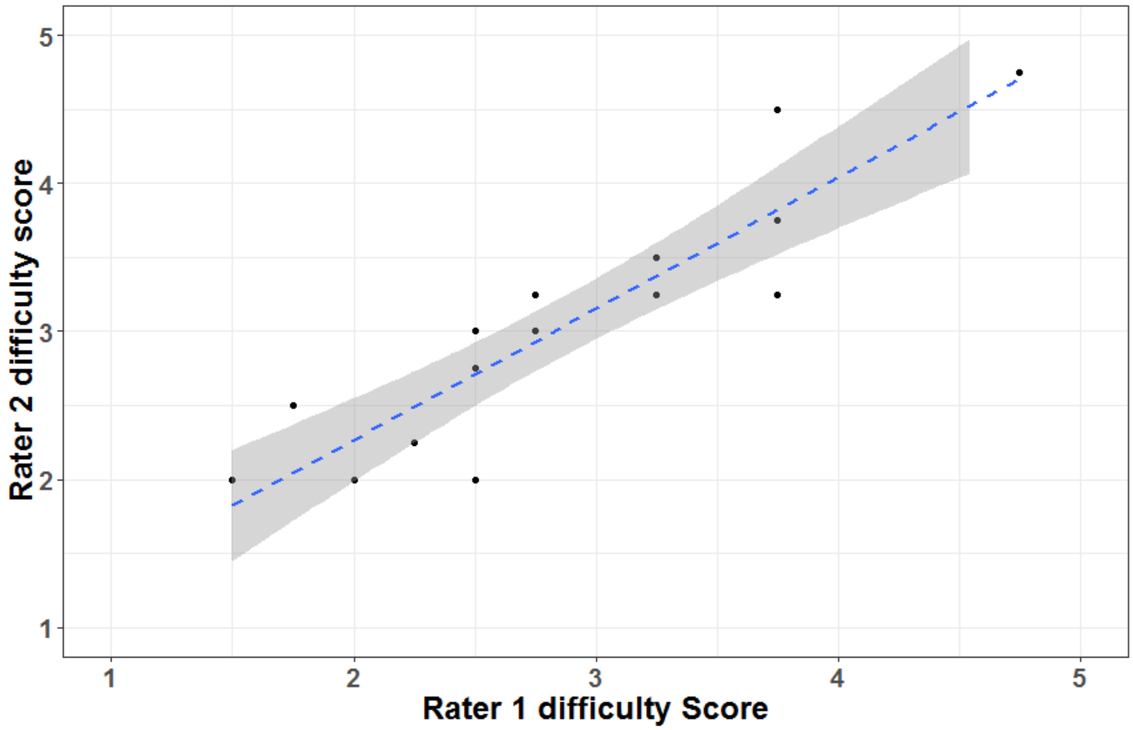


Figure 13: The relationship between the assigned difficulty score by expert rater 1 and expert rater 2. Each point represents a different problem.

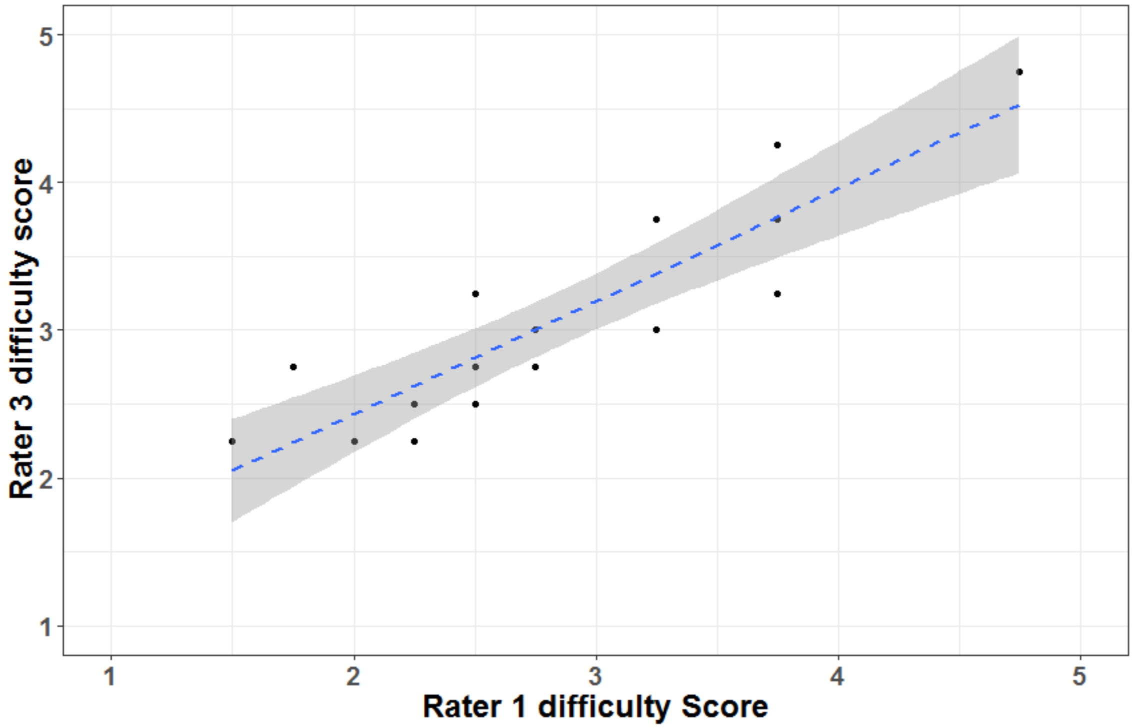


Figure 14: The relationship between the assigned difficulty score by expert rater 1 and expert rater 3. Each point represents a different problem.

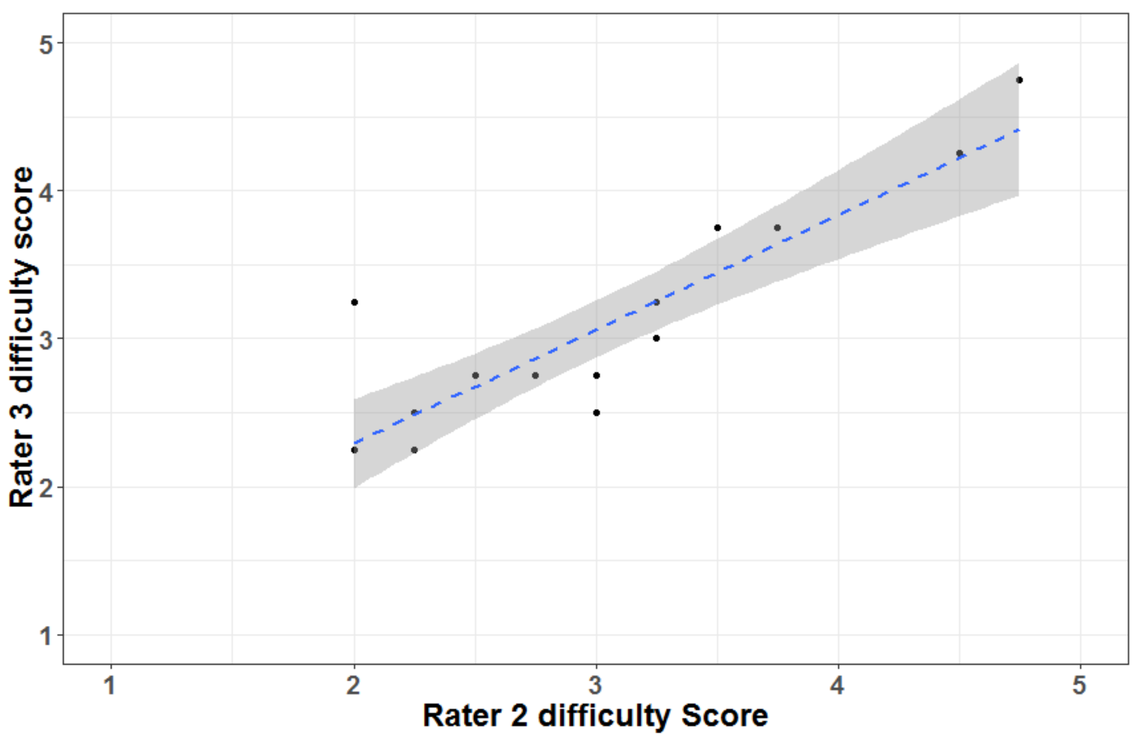


Figure 15: The relationship between the assigned difficulty score by expert rater 2 and expert rater 3. Each point represents a different problem.

The high pairwise correlations between the difficulty ratings of the PER graduate students and the rubric developer indicates a good agreement when these experts apply the problem difficulty measure. This level of reliability shows that the problem difficulty measure can be of research quality when used by PER researchers who have a substantial background in the study student difficulties in problem solving.

4.5 Summary

In this chapter, I have described the evidence of construct validity, criterion validity, and validity of the internal structure of the measure. The data used in this study were the grades of five problems from each of 3552 students in classes taught by 17 faculty members using 85 TA graders. I then used the categories of the difficulty measure on each of the 20 different exam problems graded. As a result of the analysis describe above, the difficulty measure was been reduced to only two categories that explain 77% of the variance of student performance on typical exam problems for calculus based introductory physics. I have also shown evidence for the reliability of the problem difficulty measure.

Despite the large sample size, this study has limitations. The validity has only been tested at one institution, and the reliability data of 3 raters are from one PER group.

In next chapter, I will discuss these limitations, possible applications for the instrument to measure the difficulty of physics problems, and future directions for research.

Chapter 5 Summary and Possible Applications of the Problem Difficulty Measure

5.1 Summary of the results of this study

In summary, I have developed an instrument for measuring the difficulty of the types of physics problems typically used in introductory physics courses by analyzing the wording and physics content of problems along with the mathematical processes required to arrive at a solution. This measure is based on the presumed cognitive load placed on students trying to solve that problem. Judging that cognitive load is based on information processing theory stressing the limited capability of short term and working memory together with the cognitive models of mental resources and ontological categories. This study used the performance of a large student population on final examination problems in introductory physics courses to test the validity of this problem difficulty measure. The study assumes that student performance in the authentic situation of solving quantitative multi-step physics questions made up by their instructors on a final examination is an indicator of question difficulty. These questions were found to be similar to those given in standard textbooks and are typical of instructor questions used around the country. A detailed examination of a sample of the written student solutions showed that the questions were indeed problems for those students.

For ease of use, I chose to measure four identifiable categories that I hypothesized impact the cognitive load placed on students attempting to solve a problem: Length of Problem Statement, Problem Context, Physics Principles, and Mathematical Complexity. Each category was scored on a scale from 1(least difficult) to 5 (most difficult) using a rubric developed for that purpose. These categories are similar to those proposed by other researchers based on their analysis of student problem solutions, the opinions of physics instructors, and the opinions of students [47, 31, 29, 30]. Some of the difficulty categories found in previous research [31], such as no topic cueing, extraneous or missing information, or lack of explicitness of the question, did not arise since these features did not occur in the introductory physics exam problems used by this large sample of instructors. If they do arise, they might be subsumed into an expanded definition of problem context.

Based on the student performance on final exam problems, the difficulty measure collapsed into only 2 categories: Physics Principles and Mathematical Complexity, with a Spearman correlation coefficient 0.88 for the 2 categories compared to 0.90 for all 4 categories. Using just these two categories of difficulty predicted 77% of the variance of student performance on their final exam problems.

In this dissertation, I have described the importance of building a problem difficulty measure that can be determined directly from the problem statement (Chapter 1), the existing research that forms the empirical and theoretical foundation of its construction (Chapter 2), the categories and procedures for applying this measure (Chapter 3), and the evidence for the validity and reliability of this measure (Chapter 4). In this section (Chapter 5), I will discuss the limitations of the study as well as its implications, and possible applications for the instrument outside the scope of this study.

5.2 Limitations

This study was conducted at a single large midwestern research university which may not be indicative of the diversity of students and institutions around the country. For example, the students in this population tend to be from the top 5% of their high school classes in a state with one of the best public-school education systems in the country. To test the generality of these results, additional studies are necessary at a broader range of institutions encompassing different student populations, faculty, and pedagogy used in this course.

In addition, the student sample was analyzed in aggregate as appropriate for a study of this size. It is possible that the results could mask a problem difficulty diversity experienced by specific subpopulations of the students in this study. In a larger study, additional analysis could determine if the results apply to those different student sub-groups.

It is important to note that this university provides a support program for the graduate teaching assistants who grade student problem solutions that emphasizes evaluating student problem solving. Using student problem solving grades to indicate student problem solving performance at institutions where this type of support does not exist for problem graders may not be appropriate. This can be tested by determining the correlation of problem grades with the results of using a rubric developed to measure the extent to which student problem solving is expert-like [11].

The reliability data were from raters in a single physics education research group. This group has a history of emphasizing physics education issues connected to student problem solving. Additional data are needed to determine if the results apply to a broader group of experts.

5.3 Implications

It is well known that physics instructors have different styles of writing problems that they use for their course. This makes comparing student problem solving performance with different instructors or over time with the same instructor problematic. In its simplest use, the problem difficulty measure could determine if instructors are making consistent assessments of their students while using different final exam problems. This would lead to a more consistent assessment of students taking a course than requiring instructors to have similar student grade distributions.

Making sure that problems used for student assessment have a consistent range of difficulty is especially important if an institutional goal is to encourage the improvement of instructional techniques. Determining a problem's difficulty in a manner that is independent of a particular problem format, or even its subject matter, allows researchers to test pedagogical interventions across different instructors, courses, and institutions. This allows the comparison of an instructional change with the historical data from that course even though the instructors and examination problems are not identical. The use of historical data decreases the statistical uncertainty of an analysis of the data because, if a course has been taught for many years, as is true for most introductory physics courses, there is a large amount of student performance data available. The historical data comparison can be made more robust by using propensity analysis techniques [48].

This work shows that written problems spanning the types typically used on exams in a calculus based introductory physics classes can be analyzed for their difficulty from the problem statement. Because two of the original four categories tested were not necessary to predict student performance, the results call into question the problem difficulty measure's inspiration of information processing theory. Whether this result has uncovered a different theoretical framework that underlies problem difficulty or it is caused of the limited range of problem types used by these instructors is an open question. Nevertheless, the two category measure predicts 77% of student problem solving performance. Whatever the correct underlying theoretical construct of problem difficulty, it is likely universal, meaning that a similar measure for quantifying problem difficulty might be constructed for other quantitative courses. Being able to determine difficulty of problems means that the different techniques used to teach students problem solving can be directly compared. This could lead to a more standard technique of teaching problem solving that would reinforce student learning across departmental boundaries.

5.4 Application to classroom teaching

Previously I have shown that an expert rater can use the instrument in a manner precise enough for research purposes. However, this raises the question if the measure can be used by classroom teachers to help them guide the alignment of their student assessment with their course goals.

To be of direct use to a single instructor in the classroom, they must be able to apply it without directly consulting an expert rater. In a small pilot study, two physics graduate students who were experienced TAs in introductory physics and in at least their third year of study, scored the difficulty of the same 20 problems given in Appendix B. These students had no affiliation or contact with the PER group. The graduate students were provided with a written instruction sheet, brief definitions of each difficulty category, tables of how the 1-5 scale was applied for each category as in Chapter 4, the problem statement, a correct solution to the problem, and a blank scoring template table. If these were instructors using the instrument to determine the difficulty of their own problems, they would already be familiar with the problem statement and a correct solution to the problem. There was no other contact with the researcher and no organized contact among these graduate students. After one week they handed in their difficulty ratings of the problems.

Based on this small-scale study with minimal training, the results were not conclusive but encouraging enough to deserve future study. The assigned difficulty score by the two inexperienced raters showed correlations with the students' performance although not as strongly as an expert rater. The Spearman correlation coefficient between each inexperienced rater and the developer was 0.48 and 0.51 with a probably of no correlation of 0.04 and 0.02. The data for this pilot study is in Appendix C.

5.5 Conclusion

This work reports on the development of an instrument to measure the difficulty of instructor written physics problems based on problem characteristics related to the presumed student cognitive load. Among the 4 identified difficulty factors tested, two account for almost all of the variance with student problem solving performance: Physics Principles and Mathematical Complexity. The validity of the measure was based, in part, on its ability to account for 77% of the variance of the performance of 3552 students on final exam problems in a one-year introductory physics spanning mechanics and electricity & magnetism. These students were in multiple classes taught by 17 different professors. The reliability of the measure was demonstrated by the agreement of three expert raters with a pairwise correlation of greater than 94%.

The measure described in this dissertation quantifies the difficulty of authentic introductory physics problems based on an evaluation of an expert's solution to the problem and the presumed student cognitive load engendered by the concepts required for that solution. It improved on previous research in PER that only identified perceived difficulty factors [31] or ranked the relative difficulty of similar problems [29] [30]. Compared to a more limited measure developed for algebra [32], this instrument was able to score the difficulty of problems spanning all the topics in a one-year introductory physics. The two categories of difficulty found by this study, Physics Principles and Mathematical Complexity, agree with similar categories of content type including more steps, math, direction, and content found in previous study in the University of Illinois [30] and the number of necessary physics principles found in the Minnesota study [31].

Having a standard technique for measuring problem difficulty, such as the one developed for this dissertation, will be a useful tool in ongoing research investigating effective modes of instruction, assessing the efficacy of instructional materials, and determining the consistency of classroom assessment of students. This dissertation gives an existence proof for such a measure. However, its generality needs to be the subject of further research involving students and instructors at institutions of higher learning that differ from the University of Minnesota where it was developed and tested.

Bibliography

[1] Board, National Science, Revisiting the STEM workforce: A companion to science and engineering indicators 2014, Arlington.

[2] Olson, S. , Riordan, D. G., Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President., ERIC, 2012.

[3] Chen, X. and Soldner, M., "STEM attrition: college students' paths into and out of STEM fields. Statistical Analysis Report," *National Center for Education Statistics*, 2013.

[4] Shaw, E. J. and Barbuti, S., "Patterns of persistence in intended college major with a focus on STEM majors," *NACADA Journal*, 2010.

[5] Maltese, A. V. and Tai, R. H., "Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students," *Science education*, 2011.

[6] Zhang, G.,Anderson, T. J., Ohland, M. W. and Thorndyke, B. R., "Identifying factors influencing engineering student graduation: A longitudinal and cross-institutional study," *Journal of Engineering education*, 2004.

[7] French, B. F., Immekus, J. C. and Oakes, W.C., "An examination of indicators of engineering students' success and persistence," *Journal of Engineering Education*, 2005.

[8] Marra, R. M.,Rodgers, K. A.,Shen, D. and Bogue, B., "Leaving engineering: A multi-year single institution study," *Journal of Engineering Education,* 2012.

[9] Hall, C.W., Kauffmann, P.J., Wuensch, K. L., Swart, W. E., DeUrquidi, K. A., Griffin, O. H., and Duncan, C.S., "Aptitude and personality traits in retention of engineering students," *Journal of Engineering Education,* 2015.

[10] Etkina, E. ,Heuvelen,A. V., White-Brahmia, S., Brookes, D. T.,Gentile,M. , Murthy,S. , Rosengrant, D. and Warren,A., "Scientific abilities and their assessment," *Physical Review special topics-physics education research,* 2006.

[11] Docktor, J. L., Dornfeld, J., Frodermann, E., Heller, K., Hsu, L., Jackson, K. A. and Yang, J., "Assessing student written problem solutions: A problem-solving rubric with application to introductory physics," *Physical Review Physics Education Research, 12(1), 1–18.,* 2016.

[12] Docktor, J. L .and Mestre, J., "Synthesis of discipline-based education research in physics," *Physical Review Special Topics-Physics Education Research,* 2014.

[13] Jonassen, David H, "What makes scientific problems difficult?," in *Learning to solve complex scientific problems,* 2017.

[14] Bassok, M., "Analogical transfer in problem solving".

[15] Meacham, J. A. and Emont, N. C., The interpersonal basis of everyday problem solving., 1989.

[16] Wood, Phillip K, "Inquiring systems and problem structure: Implications for cognitive development," *Human Development,* 1983.

[17] Newell,A., Simon,H. A., Human Problem Solving, Englewood Cliffs, NJ: Prentice-Hall, 1972.

[18] Newell, A. ,Shaw,J. D. and Simon,H. A., "Elements of a theory of human problem solving," *Psychological review,* 1958.

[19] Sweller, J., "Cognitive load during problem solving: Effects on learning," *Cognitive science,* 1988.

[20] Simon, H. A., "Information-processing theory of human problem solving," *Handbook of learning and cognitive processes,* 1978.

[21] Gick, M. L., "Problem-solving strategies," *Educational psychologist,* 1986.

[22] diSessa, A. A., "Toward an epistemology of physics," *Cognition and instruction,* 1993.

[23] Chi,M. T. H. and Slotta,J. D., "The ontological coherence of intuitive physics," *Cognition and instruction,* 1993.

[24] diSessa, A. A.,Sherin,B. L., "What changes in conceptual change?," *International journal of science education,* 1998.

[25] Hammer, D., "Misconceptions or p-prims: How might alternative perspectives on cognitive structure influence instructional perceptions and intentions," *The Journal of the Learning Sciences,* 1996.

[26] Chi, M. T. H. ,Slotta, J. D. and de Leeuw,N., "From things to processes: A theory of conceptual change for learning science concepts," *Learning and instruction,* 1994.

[27] Chi, M.T.H., Feltovich, P. J. and Glaser, R., "Categorization and representation of physics problems by experts and novices," *Cognitive science,* 1981.

[28] Slotta,J. D. , Chi, M. T. H. and Joram,E., "Assessing students' misclassifications of physics concepts: An ontological basis for conceptual change," *Cognition and instruction,* 1995.

[29] Gire ,E. and Rebello,N. S., "Investigating the perceived difficulty of introductory physics problems," *AIP conference proceedings,* 2010.

[30] Fakcharoenphol, W.,Morphew, J.W.and Mestre, J.P., "Judgments of physics problem difficulty among experts and novices," *Physical Review Special Topics-Physics Education Research,* 2015.

[31] Heller, P., Keith, R., and Anderson, S., "Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving.," *American Journal of Physics,* 1992.

[32] Lee, F., Heyworth, R., "Problem complexity: A measure of problem difficulty in algebra by using computer," *Education Journal-Hong Kong-Chinese University of Hong Kong,* 2000.

[33] Brewe, E., "Energy as a substancelike quantity that flows: Theoretical considerations and pedagogical consequences," *Physical Review Special Topics-Physics Education Research,* 2011.

[34] Duit, R., "Should energy be illustrated as something quasi- material?," *International Journal of Science Education*, 1987.

[35] Stephanik ,B. M. and Shaffer,P. S., "Examining student ability to interpret and use potential energy diagrams for classical systems," in *AIP Conference Proceedings*, 2012.

[36] Lindsey, B., "Student reasoning about electrostatic and gravitational potential energy: An exploratory study with interdisciplinary consequences," *Physical Review Special Topics-Physics Education Research*, 2014.

[37] Heller, P., Post, T., Behr, M., & Lesh, R, "Qualitative and Numerical Reasoning About Fractions and Rates by Seventh and Eighth Grade Students," *Journal for Research in Mathematics Education*, 1990.

[38] Cohen ,E. and Kanim,S. E., "Factors influencing the algebra “reversal error”," *American Journal of Physics*, 2005.

[39] Flores,S. , Kanim, S. E.and Kautz,C. H., "Student use of vectors in introductory mechanics," *American Journal of Physics*, 2004.

[40] Bollen, L., Kampen, P. V. and Cock, M. D., "Students ’ difficulties with vector calculus in electrodynamics," *Physical Review Special Topics-Physics Education Research*, 2015.

[41] Hu, D. and Rebello, N. S., "Understanding student use of differentials in physics integration problems," *Physical Review Special Topics-Physics Education Research*, 2013.

[42] Wallace, C. S. and Chasteen, S. V., "Upper-division students' difficulties with Ampere's law," *Physical Review Special Topics-Physics Education Research,* 2010.

[43] Pepper, R.E., Chasteen, S. V., Pollock, S. J. , Perkins, K. K., "Our best juniors still struggle with Gauss's Law: Characterizing their difficulties," *AIP Conference Proceedings,* 2010.

[44] American Educational Research Association, Psychological Association, & National Council on Measurement in Education, Standards for Educational and Psychological Testing., Washington, DC: American Educational Research Association, 1999.

[45] Cronbach, L. J. , Meehl, P.E., Construct Validity in Psychological Tests, Psychological Bulletin, 1955.

[46] Lawrenz, F. and others, "Training the Teaching Assistant.," *Journal of College Science Teaching,* 1992.

[47] McDaniel, M. A.,Stoen, S. M., Frey, R.F.,Markow, Z.E.,Hynes, K.M.,Zhao, J. and Cahill, M. J., "Dissociative conceptual and quantitative problem solving outcomes across interactive engagement and traditional format introductory physics," *Physical Review Physics Education Research,* 2016.

[48] Rosenbaum, P. R., Rubin, D. B., "The central role of the propensity score in observational studies for causal effects," *Biometrika,* 1983.

[49] Beatty, I. D., Gerace, W.J., "Probing physics students' conceptual knowledge structures through term association," *American Journal of Physics,* 2002.

[50] Eylon, B.,Reif, F, "Effects of knowledge organization on task performance," *Cognition and Instruction,* 1984.

[51] Schoenfeld, A. H., Herrmann, D. J., "Problem perception and knowledge structure in expert and novice mathematical problem solvers.," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 1982.

[52] Gliner, G. S., "College Students' Organization of Mathematics Word Problems in Relation to Success in Problem Solving," *School Science and Mathematics,* 1989.

[53] Gliner, G. S., "College students' organization of mathematics word problems in terms of mathematical structure vs. surface structure," *School Science and Mathematics,* 1991.

[54] Reif, F.,Heller, J. I., "Knowledge structure and problem solving in physics," *Educational psychologist,* 1982.

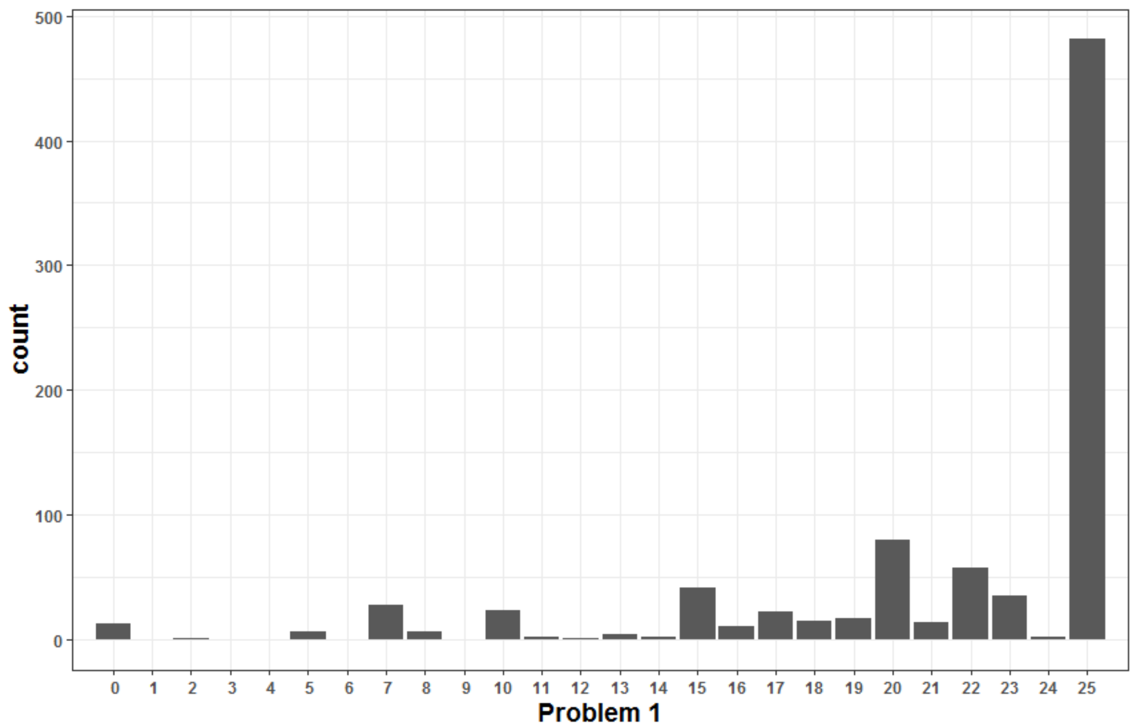
[55] Miller, G. A., "The magical number seven, plus or minus two: Some limits on our capacity for processing information.," *Psychological review,* 1956.

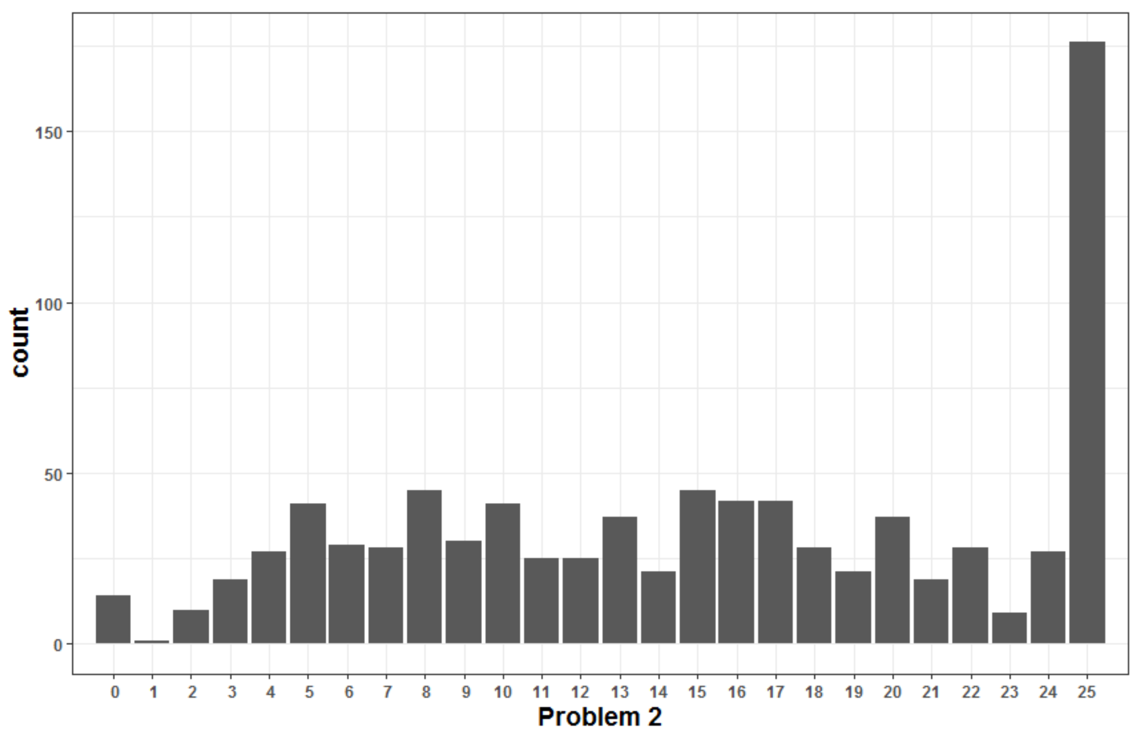
[56] McDermott, L. C., "Research on conceptual understanding in mechanics," *Physics Today,* 1984.

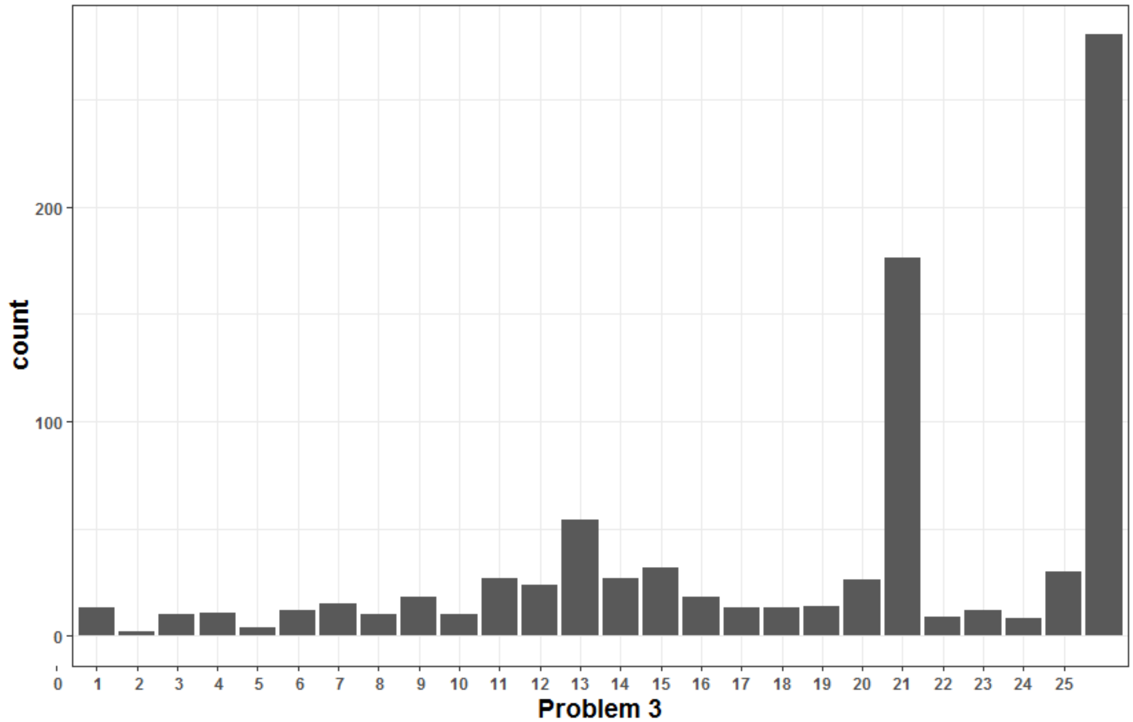
[57] Clement, J. J., "Students' preconceptions in introductory mechanics," *American Journal of physics,* 1982.

Appendix A: Distribution of Students' Scores on the Final Exam Problems Used in this Study.

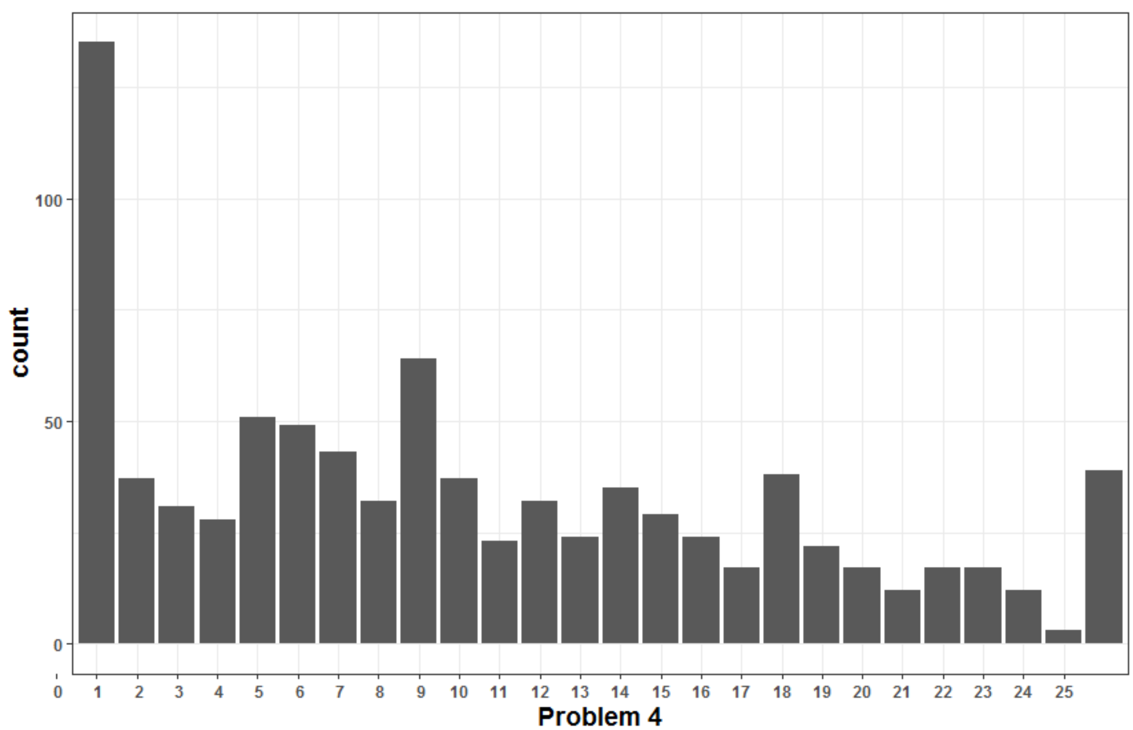
Fall 2016


count
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 1

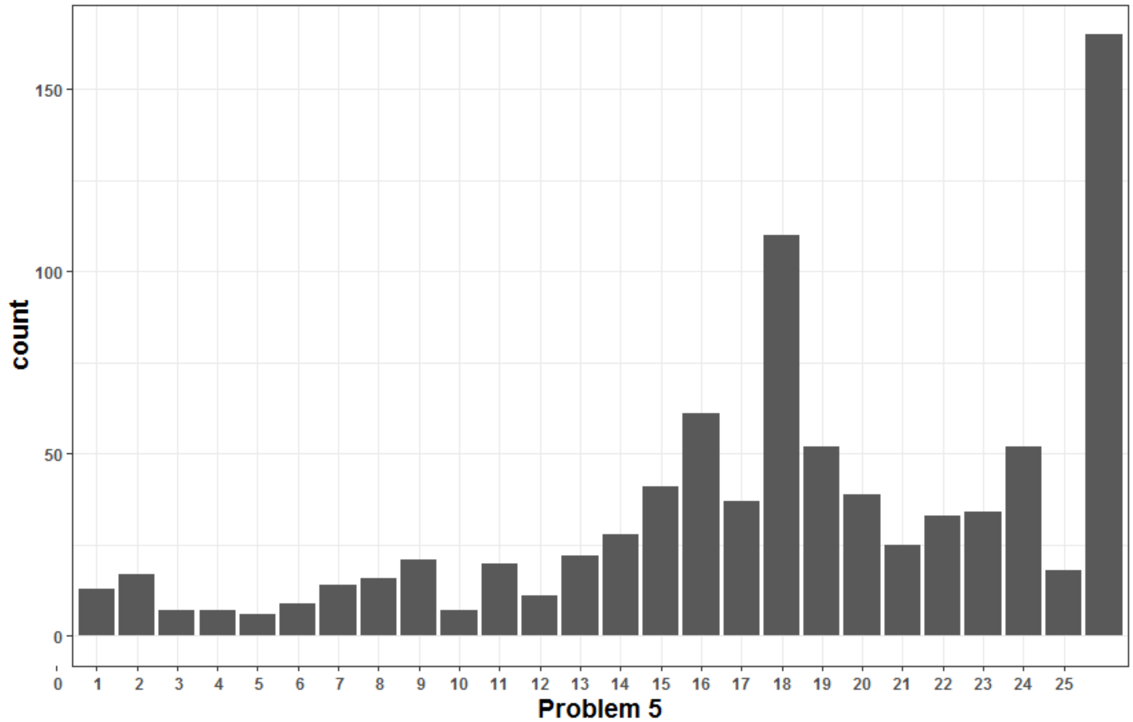

count
150
100
50
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 2



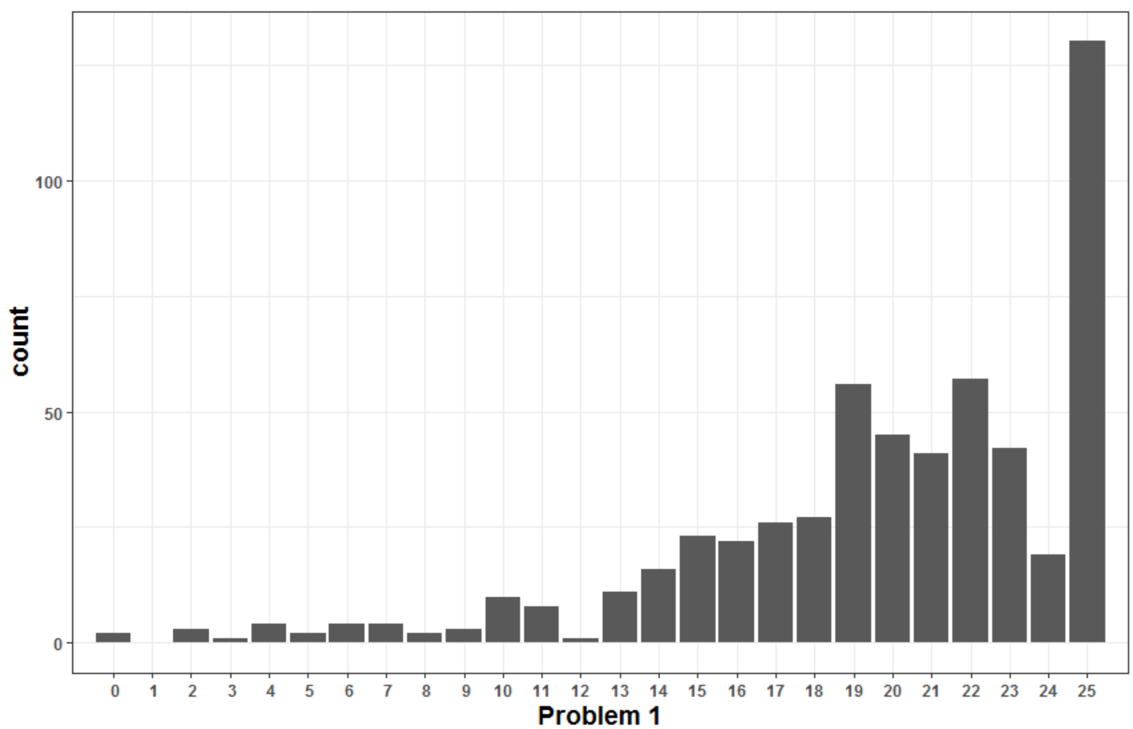
count
200
100
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 3

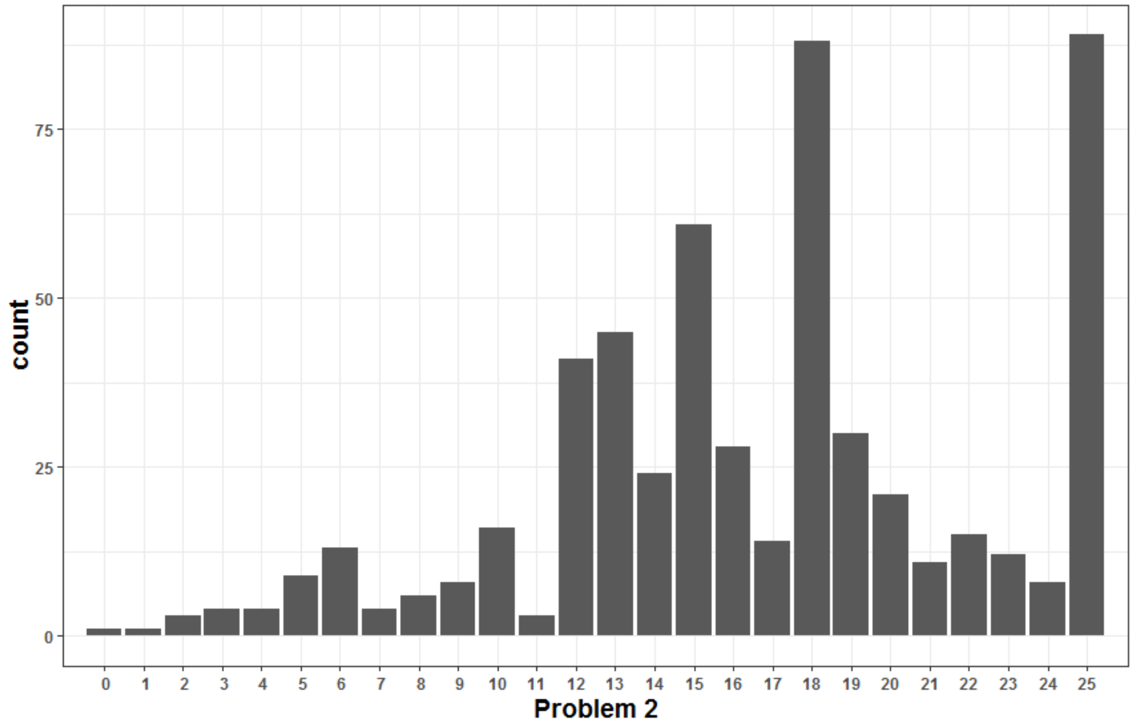


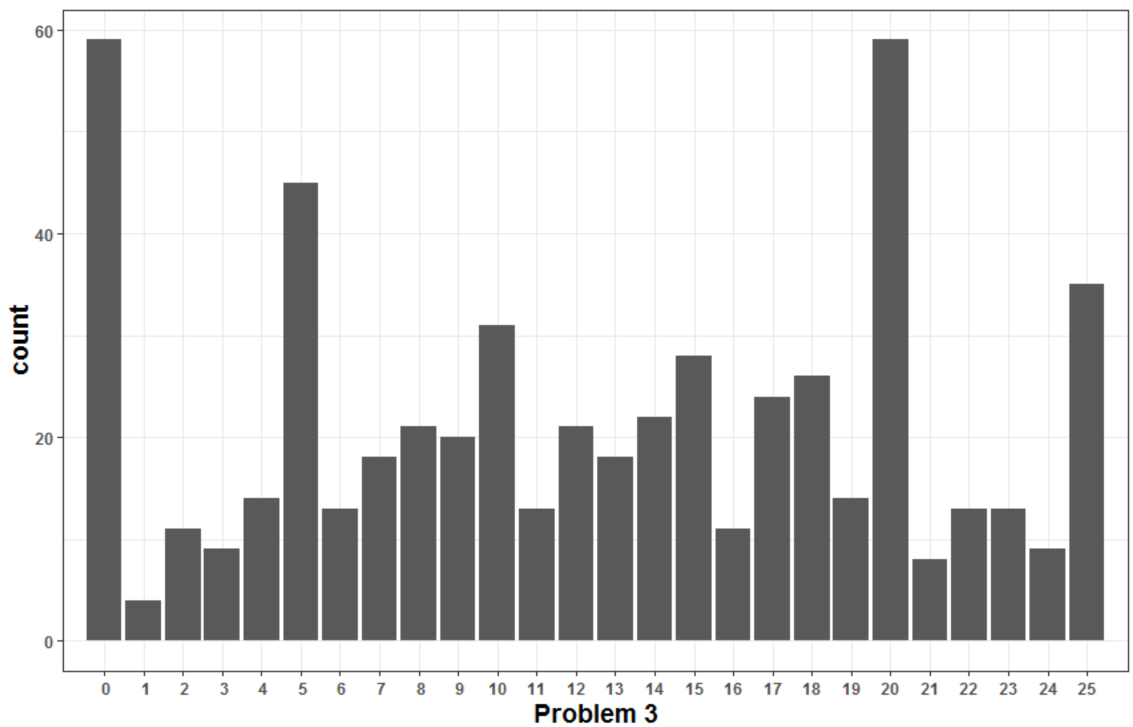
count
100
50
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 4

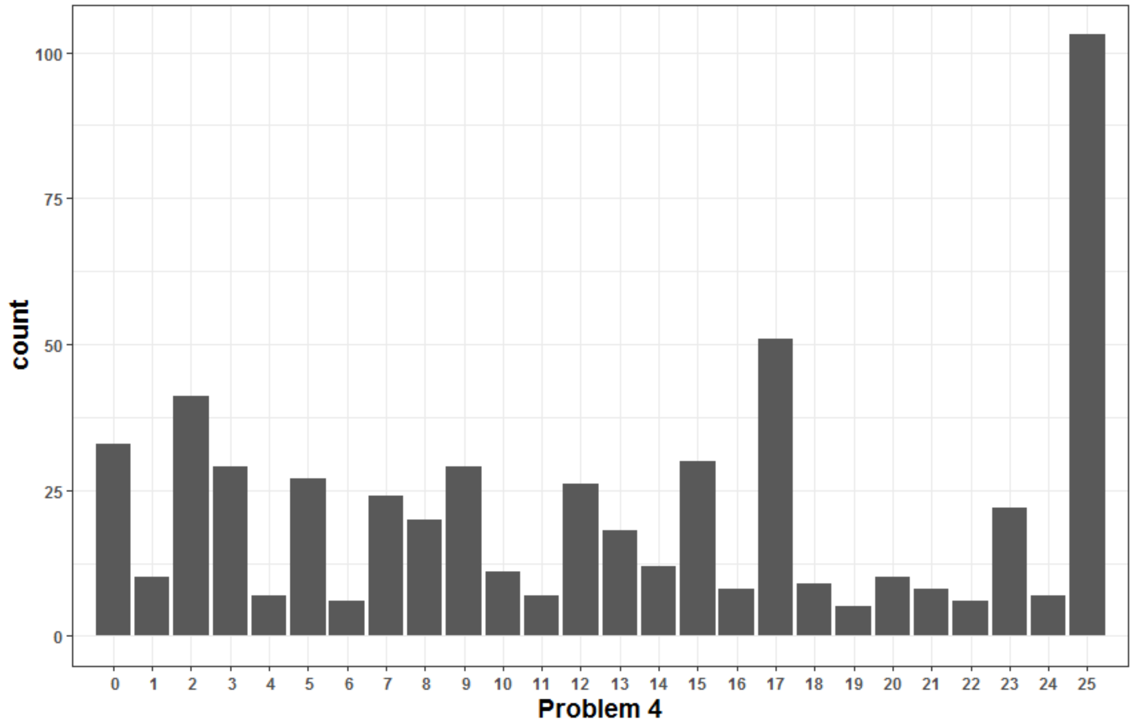

count
150
100
50
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
Problem 5

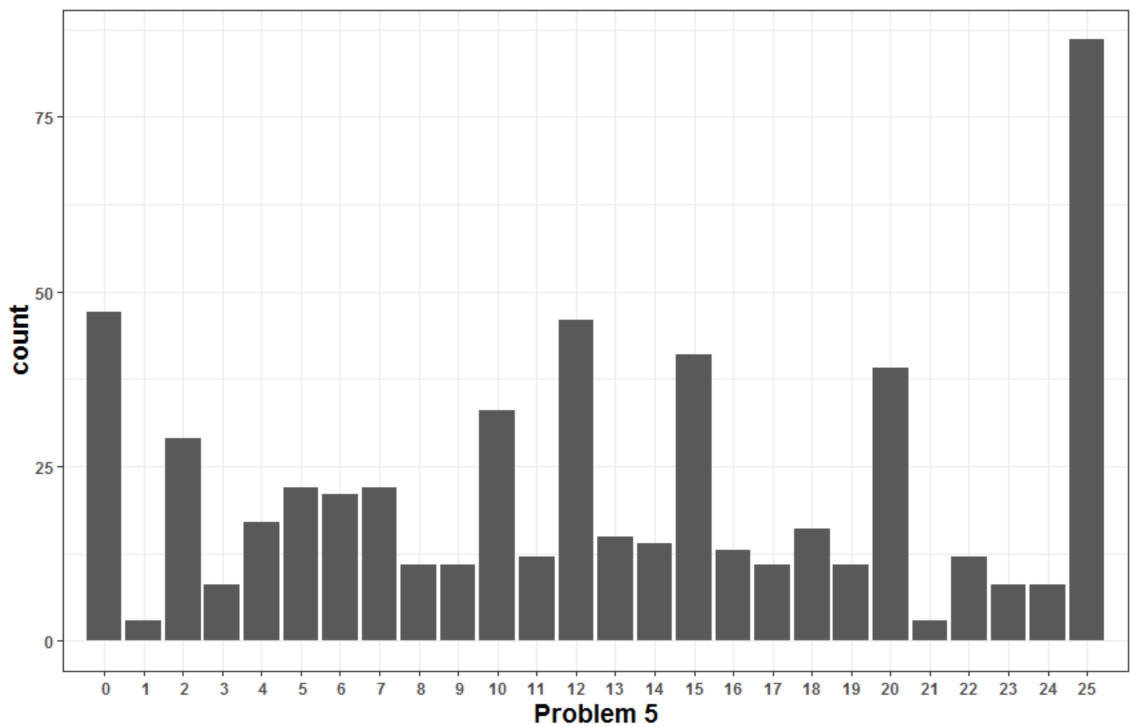
Fall 2014


count
100
50
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
Problem 1

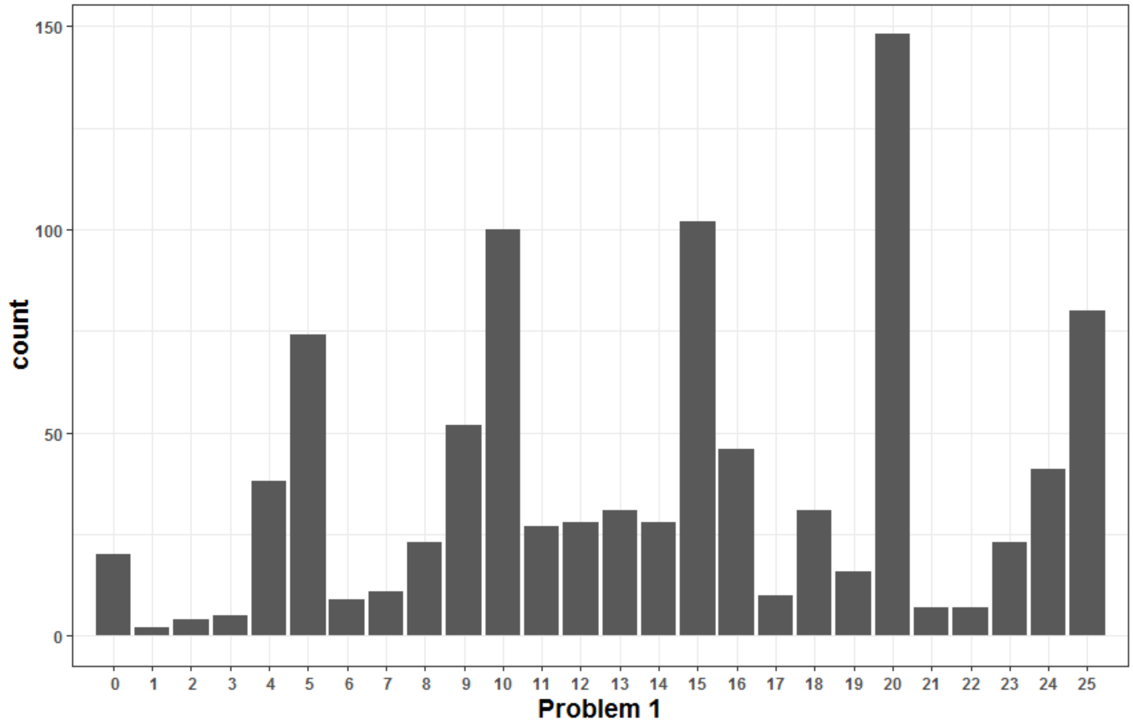

count
75
50
25
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 2

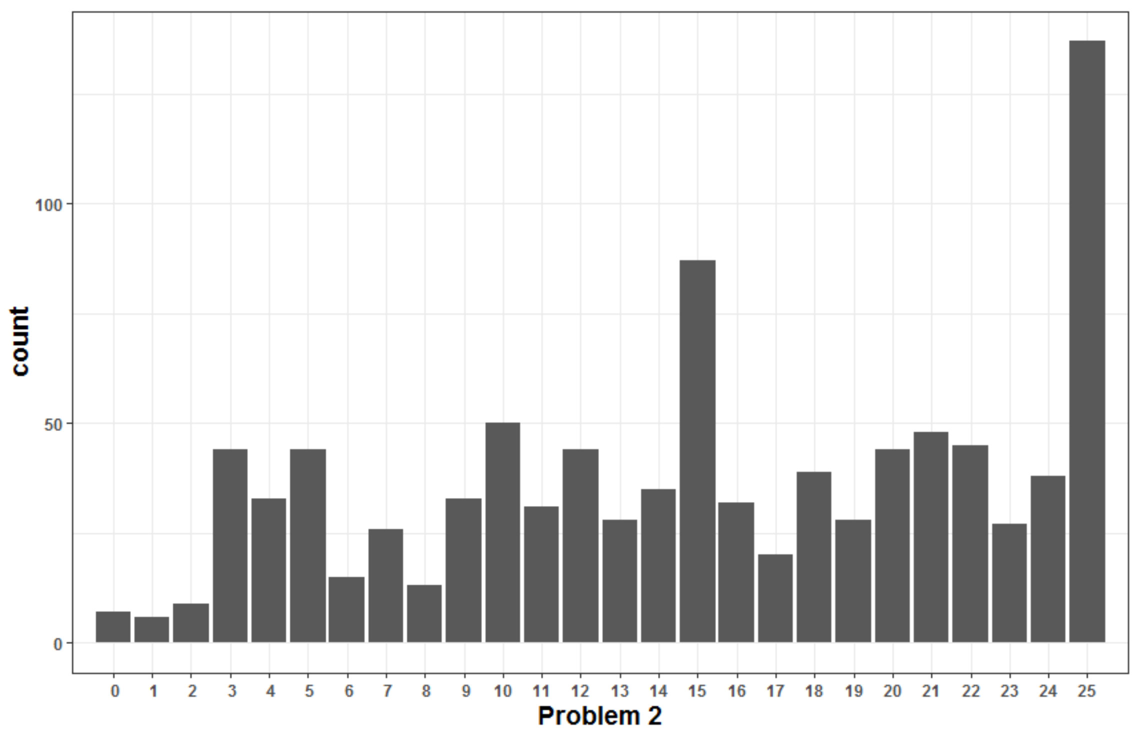

count
60
40
20
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 3

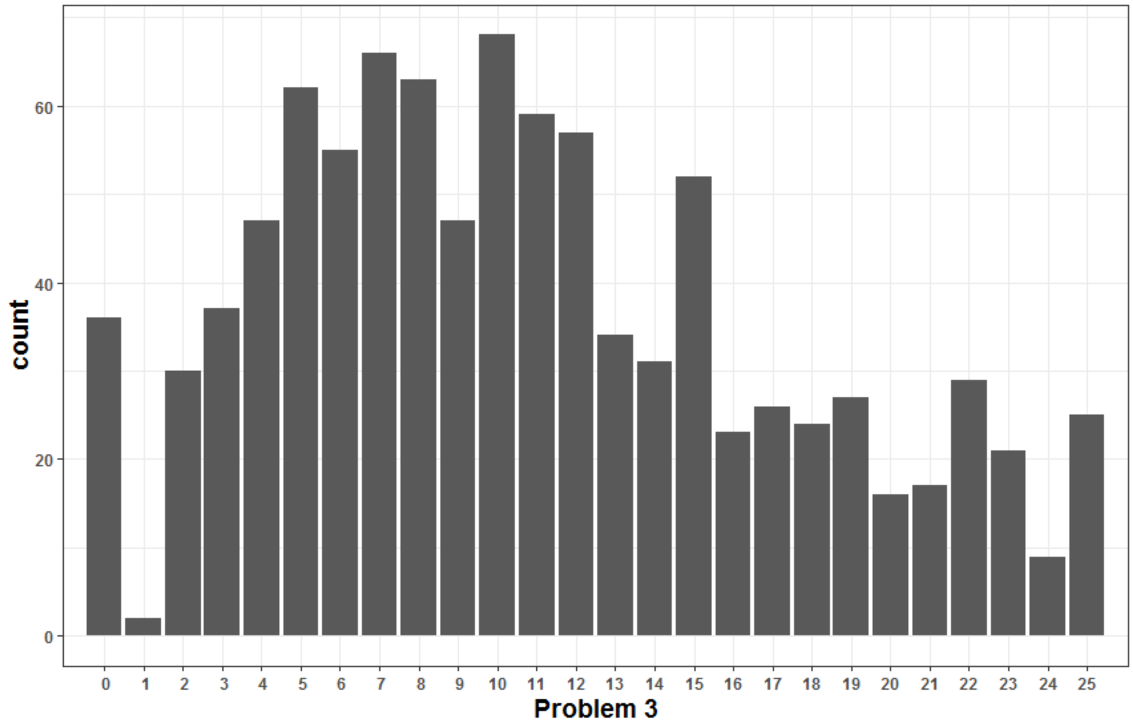

count
100
75
50
25
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 4

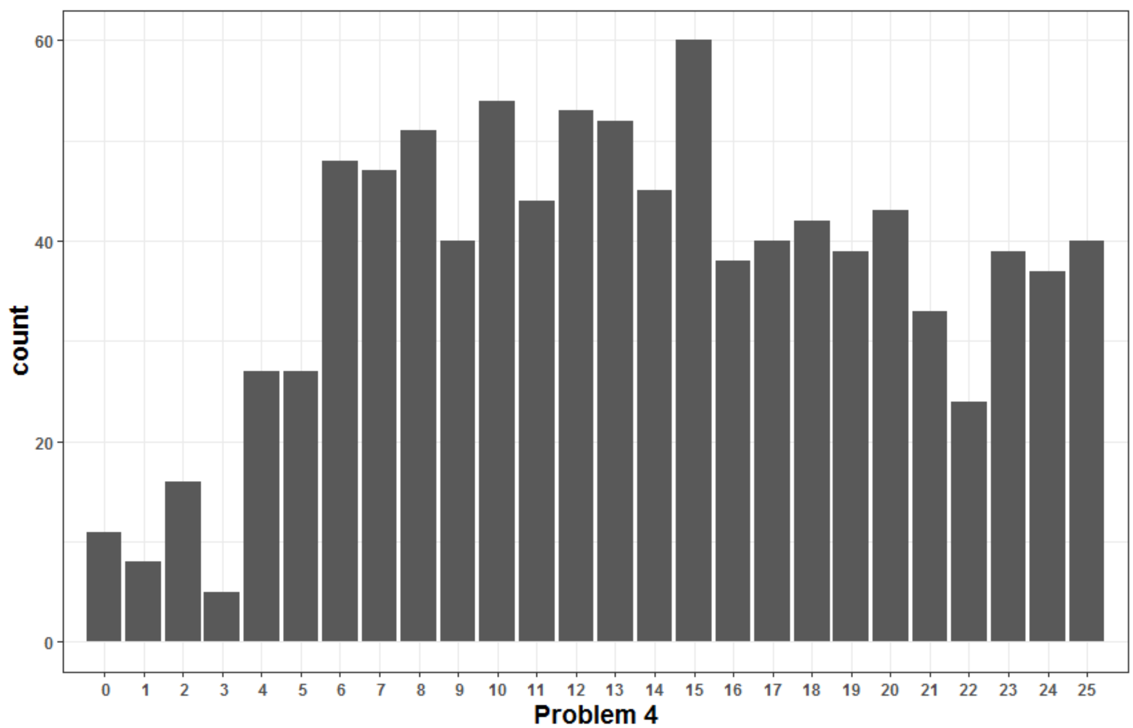

count
75
50
25
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 5

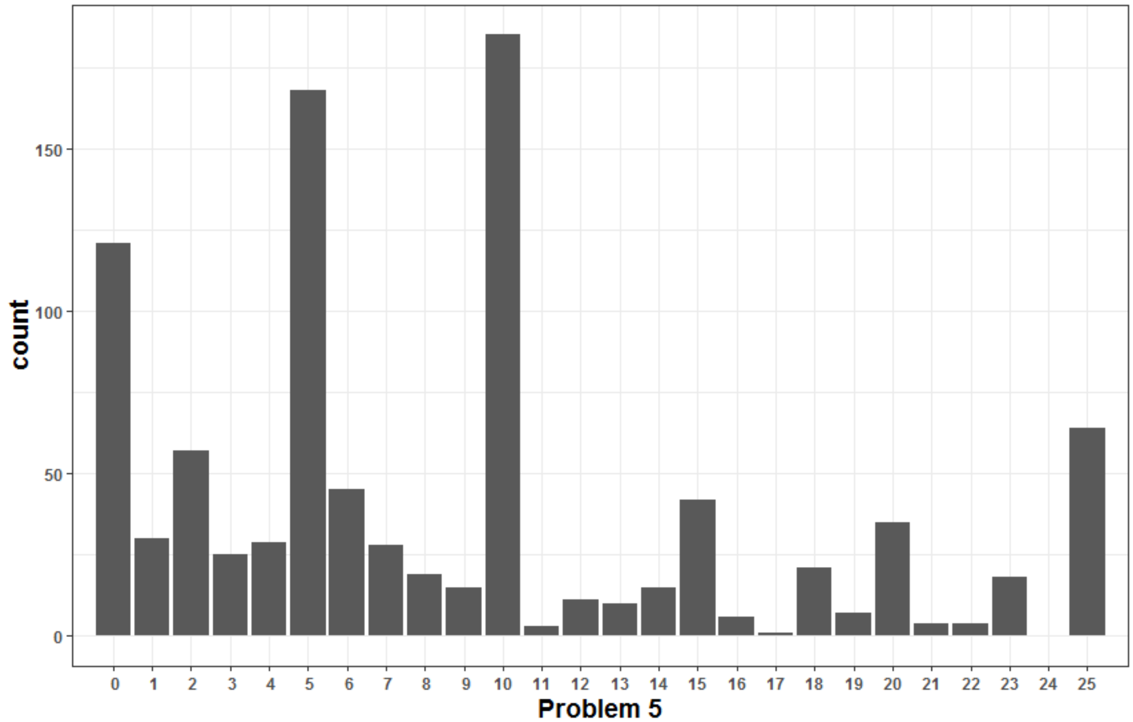
Spring 2016


150
100
50
0
count
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 1

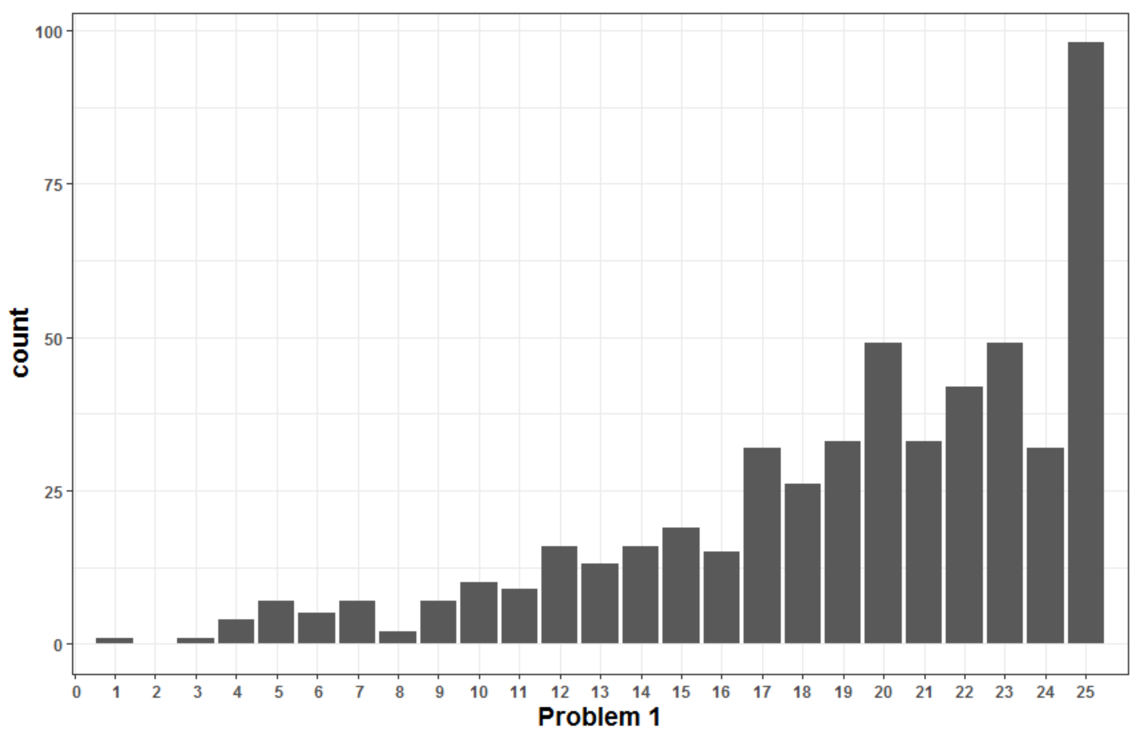

100
50
0
count
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 2

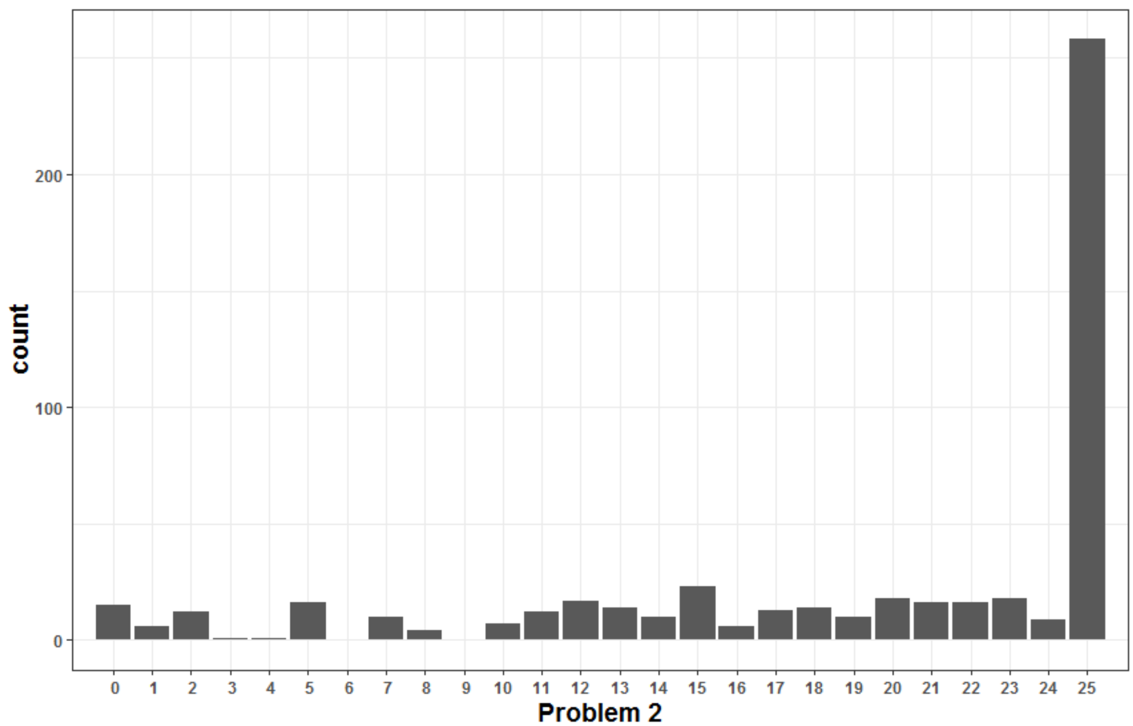

count
Problem 3

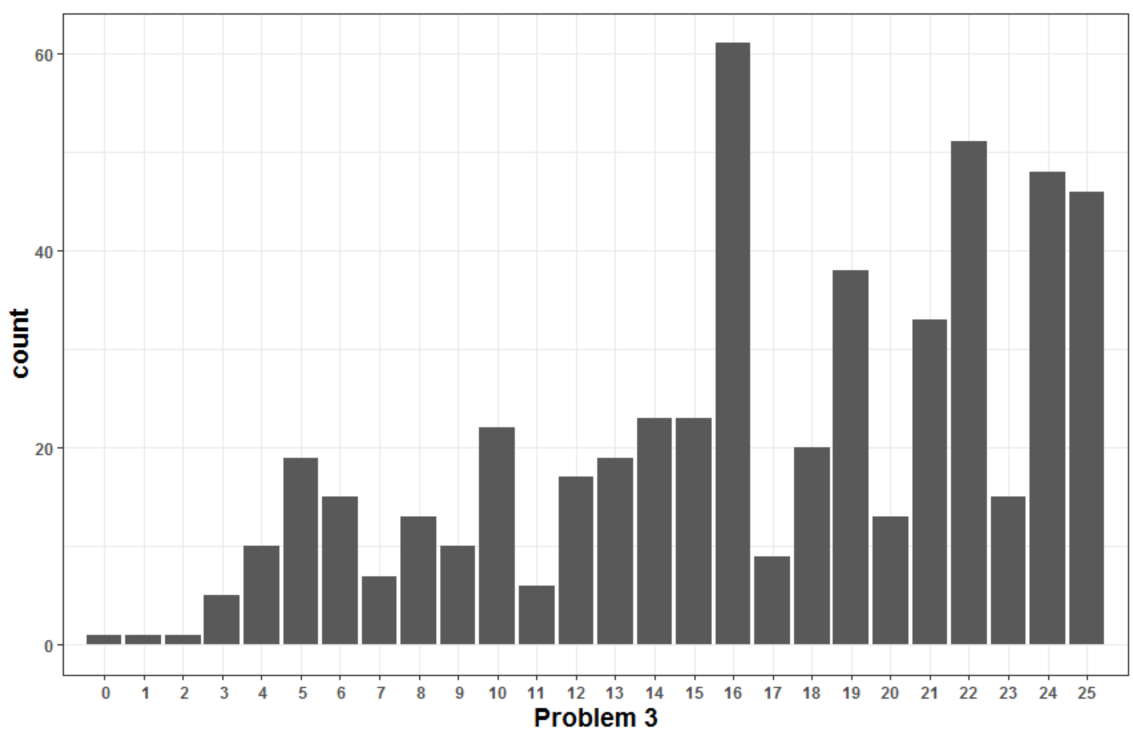

count
Problem 4

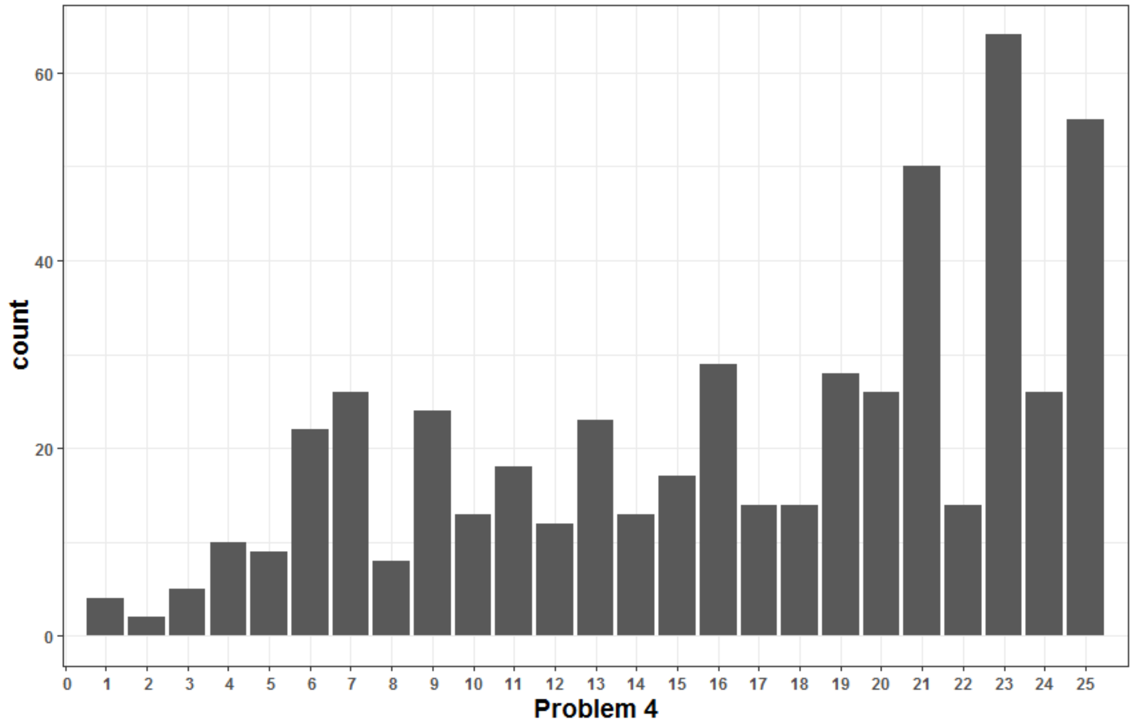

count
150
100
50
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 5

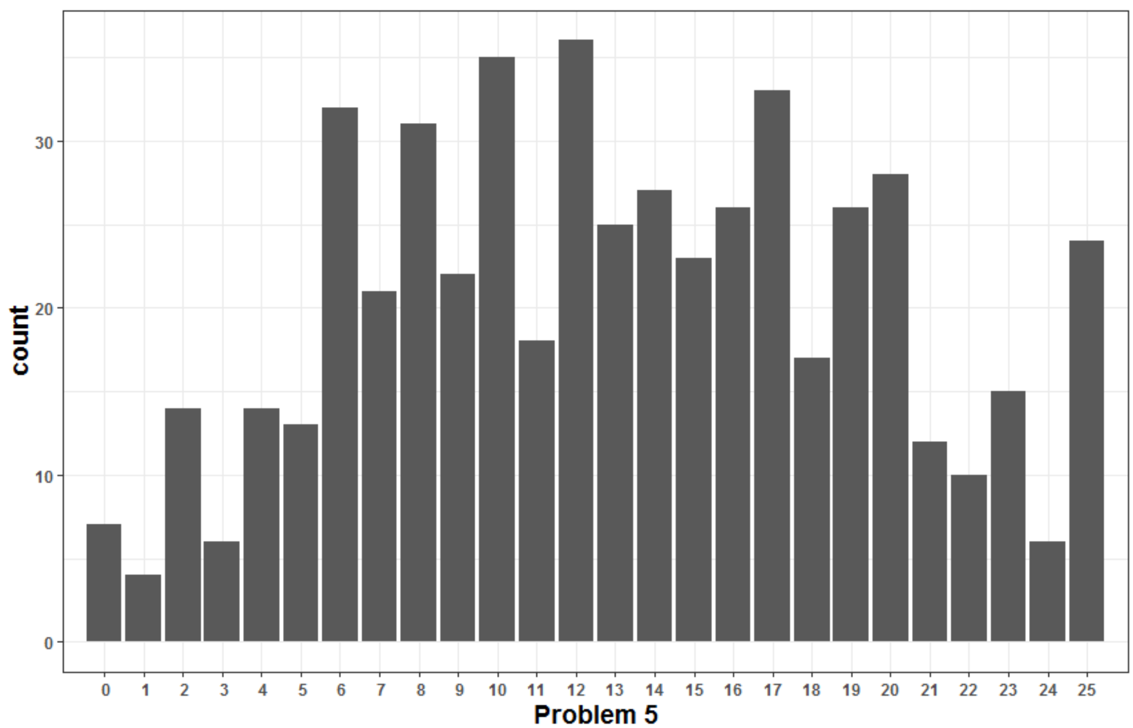
Spring 2012


count
100
75
50
25
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 1


count
200
100
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 2


count
60
40
20
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 3


60
40
20
0
count
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 4


30
20
10
0
count
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem 5

Appendix B: Test Problems

Fall 2016

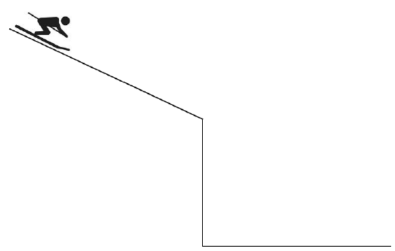
Final test of fall 2016 was given with a formula sheet which is shown below.

Problem 1 (25 points)

You are helping your friend prepare for a skateboard exhibition by determining if the planned program will work. Your friend will take a running start and jumps onto a heavy-duty 7.0 kg stationary skateboard. The skateboard will glide in a straight line along a short, level section of track, then up a sloped wall. The goal is to reach a height of at least 8.0 m above the ground before coming back down the wall. Your friend's maximum running speed is 7.0 m/s, and your friend has a mass of 68 kg. The wall has a slope of 53.1° with the ground. Can your friend perform this trick? Note you must show your work to get credit.

Problem 2 (25 points)

Immediately after your final exams, you fly to Jackson Hole for some winter skiing. You start from rest at the top of a $L = 100$ meter slope, angled at 10° above horizontal. At the end of the slope is a cliff $h = 20$ meters above ground level. Ignore friction.
(A) What is your speed at the end of the slope?
(B) How long does it take you to reach the ground after leaving the slope?
(C) How far from the base of the slope do you land?
(D) What is your velocity $\vec{v}$ just before you hit the ground?



Problem 3 (25 points)

Your lawnmower is quite old and all the wheels have rusted, so you have to push to mow the lawn. The handle of your 22-kg lawnmower makes a 35° angle with the horizontal. If the coefficient of friction between lawnmower and ground is 0.68, what magnitude of force is required to push the mower at constant velocity? Assume the force is applied in the direction of the handle. Compare with the mower's weight.

Problem 4 (25 points)

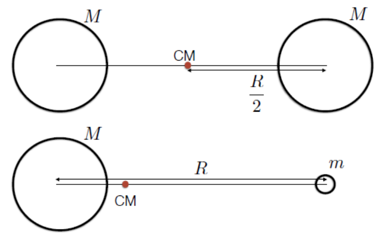
(a) Two identical stars of mass M are in circular orbits around their CM. Show that

$$T^2 = \frac{2\pi^2 R^3}{GM},$$

where R is the distance between the stars and T is the period of rotation.

(b) Now consider a star and a satellite with unequal masses m and M and show, in the case of circular orbits, that

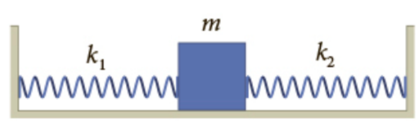
$$T^2 = \frac{4\pi^2 R^3}{G(M+m)}.$$



(c) Starting from the circular orbit you found in part (b), in the case where $m << M$ and the distance between star and satellite is R, find the escape speed for the satellite.

Problem 5 (25 points)

A block with a mass of m = 6.0 kg on a horizontal surface is anchored to two facing walls by springs. Both springs are initially at their relaxed length and have spring constants of $k_1 = 100$ N/m and $k_2 = 200$ N/m . Then the block is displaced 20 mm to the right and released. Ignore friction.
(A) What is the effective spring constant of the system?
(B) Find the equation for the position of the block as a function of time, assuming $+x$ to be to the right.
(C) What is the maximum speed of the block?
(D) What is the period of oscillation?



Fall 2014

The final test of fall 2014 was given without a formula sheet. The 5 open response problems are shown in the original format below.

Problem #1

Computer aided special effects have enabled the production of more and more fantastic movies, populated with goblins and monsters and other mythical characters. In a particular movie, the hero stands on a castle wall 21.5 m above the ground below, dropping rocks on presumed enemies below. Exactly one second after the hero drops a rock, an arrow shooting monster shoots an arrow straight upward at the rock from a point 1.5 m above the ground at a speed of 50.0 m/s.

(a) How far above the ground is the rock at the instant the arrow is launched?

(b) How long after the rock is dropped, does the arrow hit it?

(c) How far above the ground does the arrow hit the rock?

(d) The monster notices immediately that the arrow has no effect on the rock's motion, which continues to fall. Just as the arrow hits the rock, then, the monster starts running away at a speed of 3.0 m/s. How far does the monster run before the rock hits the ground?

Problem #2

A 50 kg crate rests on the bed of a truck. The coefficient of static friction between the crate and the bed of the truck is 0.40 and the coefficient of kinetic friction is 0.30.

(a) What is the fastest rate at which the truck can accelerate such that the crate does not slide across the bed of the truck?

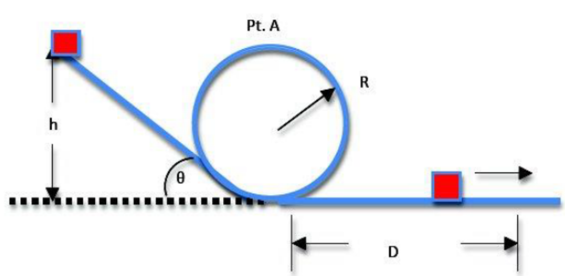
(b) At what rate must the truck be accelerating if the crate slides 1.0 m across the bed as the truck begins accelerating for 1.0 seconds? Note the distance the crate travels across the bed of the truck is different than the distance it travels relative to the ground.

Problem #3

A block of mass 2.0 kg slides along a track, as shown in the figure, that makes an angle $\theta = 30°$ with the horizontal. It starts at rest a height h above the ground, and then slides down to the base, at which point it enters a circular section of the track with radius R = 0.5 meters. The track, both the inclined plane and the horizontal section, has a coefficient of kinetic friction of $\mu_k = 0.3$, while the track in the circular section is frictionless.

(a) What is the minimum height h so that, when the block is released, it will remain on the track at the highest point (pt. A)?

(b) Find, instead, the value of the vertical height h' at which the block must be released on the incline such that the block will continue a distance D = 6.0 meters on the horizontal section of the track before coming to rest.



Problem #4

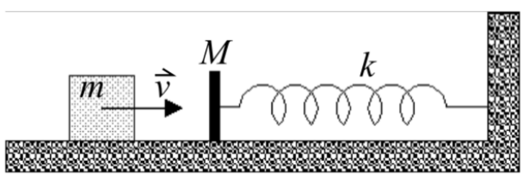
A 30-kg child stands on a merry-go-round, 2.5 m away from its center. The merry-go-round and child are initially at rest. The child then starts running on the merry-go-round with a constant speed 0.50 m/s with respect to the ground, maintaining a constant distance of 2.5 m from the center. If the merry-go-round has a moment of inertia of 320 kg-m^2 and can spin without friction around its axis:

(a) What is the final angular velocity of the merry-go-round?
(b) What is the final speed of the child **with respect to the merry-go-round at the location of the child**?

Problem #5

As shown in the figure below, a block of mass m is propelled at a speed v toward a stationary mass M that is attached to a spring with a spring constant k. The spring, in turn, is mounted to a rigid wall at its other end. After the collision, the block sticks to the mass M, and the spring begins oscillating. The surface on which the block slides is frictionless.

(a) Determine the amplitude of oscillation of the system after the block has stuck to mass M. Write your answer in terms of the given quantities.
(b) What is the ratio m/M if the oscillation frequency for the system after the block has stuck to mass M is 1/2 of that for mass M alone oscillating on the spring?



Spring 2016

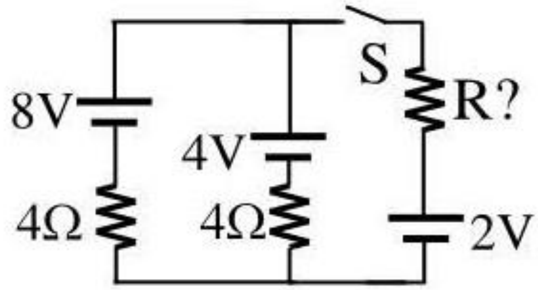
The final test of spring 2016 was given without a formula sheet. The 5 open response problems are shown in the original format below.

Question 16 of 20: Metal sphere "A", with a radius R_A = 2.00 mm, is fixed at the origin. A smaller metal sphere "B" (m_B = 10.0 grams, R_B = 1.00 mm) is tethered to sphere "A" by a thin, taut conducting string that is 1.00 m long. A total charge of 9.00 μC is placed on the connected spheres.

(a) Assuming no charge remains on the thin string, determine how the 9.00 μC charge is distributed on the spheres, i.e., how much of this charge lies on sphere "A" and how much lies on sphere "B"?
(b) Find the resulting electrical force between these spheres.
(c) The string between them is then cut, so sphere "B" begins accelerating away from sphere "A" which is still fixed at the origin. Determine the speed of sphere "B" when it is very far from sphere "A". Assume that only the electrical force between the spheres is acting on sphere "B" as it accelerates away from sphere "A".

Question 17 of 20: In the circuit shown below, after the switch "S" is closed, the current through the 8.0 V battery increases by 0.20 A as compared to its value when the switch had been open.

(a) What is the unknown resistance R?
After the switch is closed
(b) How much power is dissipated in the resistor with resistance R?
(c) What is the total net power supplied to the circuit by the batteries (note batteries that produce power supply positive power, and batteries that consume power supply negative power)?

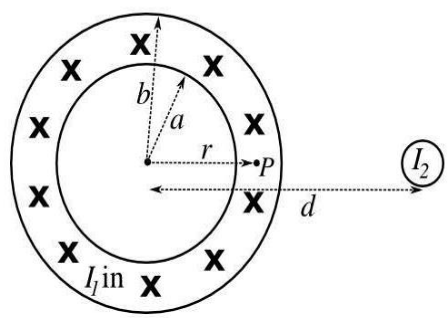


Question 18 of 20: A long cylindrical conducting shell (inner radius a = 2.0 mm, outer radius b = 3.0 mm) carries a uniform current, i.e., the current is distributed evenly throughout the cross section of the shell, of I_1= 1.0 A that is directed into the page. Another thin long wire, that lies a distance d = 6.0 mm to the right of the central axis of the conducting shell, carries some unknown current I_2 parallel to the shell that is directed either into or out of the page. (A cross section of this system is shown in the diagram.)

(a) What must the magnitude of the current I_2 be such that the total magnetic field due to the conducting shell and this unknown current is zero at point P, which lies a distance r = 2.5 mm to the right of the central axis of the conducting shell?

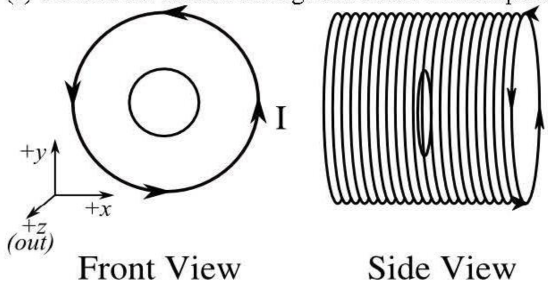
(b) What is the direction, into the page or out of the page, of I_2?

(c) For these values of the currents, what is the magnitude and direction of the force per unit length on the wire due to the shell's current?



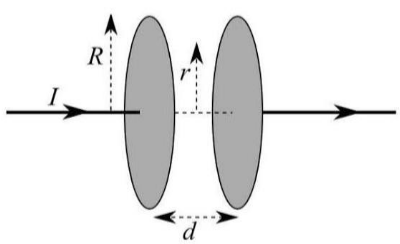
Question 19 of 20: A solenoid is made by wrapping a wire 200 times around to form a cylinder. The solenoid's length is 0.20 m and its radius is 0.050 m. A small wire loop with a radius of 0.020 m and a resistance of 0.10 Ω is placed in the center of the solenoid with its plane perpendicular to the long axis of the solenoid. At $t = 0$, the current through the solenoid is zero. Between $t = 0$ and $t = 2.0$ seconds, the current in the solenoid increases as $I(t) = ct$ (in the counterclockwise direction when looking end on as in the figure below), where $c = 3.0$ amps/s. For $t > 2.0$ s, the current is a constant 6.0 amps. Find the numerical answers below, including the variable for time t when needed.

(a) What are the magnitude and direction of the magnetic field in the center of this solenoid as a function of time for all times $t > 0$? (Ignore the B-field created by the induced current in the small wire loop.)
(b) What is the magnetic flux through the small wire loop as a function of time, again, for all times $t > 0$? (Ignore the B-field created by the induced current in the small wire loop.)
(c) What is the inductance of the solenoid?
(d) What is the induced voltage around the small wire loop as a function of time for all times $t > 0$?
(e) What is the current through the small wire loop as a function of time for all times $t > 0$?



Question 20 of 20: A parallel-plate capacitor has circular plates of radius $R = 0.30$ m. Its plates are separated by a distance $d = 0.10$ mm. The capacitor is being charged with a constant current $I = 7.0$ A.

a) What is the magnitude of the magnetic field between the plates at a distance $r = 0.20$ m from the central axis of the capacitor?
b) If you are looking down the axis of the capacitor with the positive plate closer to you and the negative plate further from you, in which direction will the B-field loop around (clockwise or counterclockwise)?



Spring 2012

The final test of spring 2012 was given without a formula sheet. The 5 open response problems are shown in the original format below.

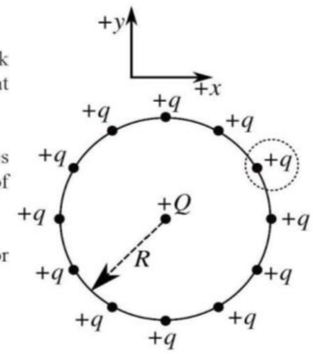
Problem 1: **[25 points]** Twelve equal charges $+q$ are situated in a circle of radius R, and they are equally spaced (see the figure).

(A) What is the net force on a charge $+Q$ at the center of the circle?

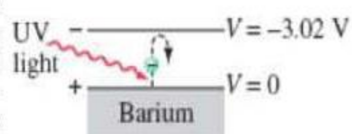
(B) You remove **only** the charge $+q$ located at "2-o'clock" (the circled charge in the clock figure to the right). Now, what is the force (magnitude and direction) on the charge $+Q$ at the center of the circle? (Find the x- and y-components of the force.)

(C) Set $V = 0$ at $r = \infty$. Replace the $+q$ charge so there are once again twelve point charges in a circle surrounding one charge $+Q$. Calculate the electrostatic potential at the center of the circle, and the potential energy of the $+Q$ charge.

(D) If the circle were smaller, would the energy of the central charge be the same, larger, or smaller? Explain.



Problem 2: **[25 points]** In a photocell, ultraviolet (UV) light provides enough energy to some electrons in barium metal to eject them from the surface at high speed. To measure the maximum energy of the electrons, another plate $d = 1$ cm above the barium surface is kept at a negative enough potential that the emitted electrons are slowed down and stopped, and return to the barium surface. If the plate voltage is -3.02 V (compared to the barium) when the fastest electrons are stopped, what was the maximum speed of these electrons when they were emitted? (The mass and charge magnitude of an electron are 9.11×10^{-31} kg and 1.6×10^{-19} C, respectively.)

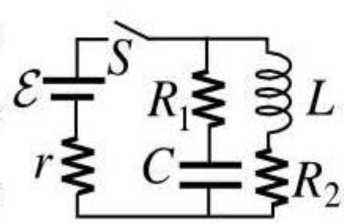


Problem 3: **[25 points]** In the circuit illustrated below, a battery with an EMF of $\mathcal{E} = 10.0$ V and internal resistance $r = 5.00\ \Omega$ is attached in parallel to an RC circuit ($R_1 = 20.0\ \Omega$, $C = 2.00\ \mu F = 2.00 \times 10^{-6}$ F) and an RL circuit ($R_2 = 40.0\ \Omega$, $L = 4.00$ mH $= 4.00 \times 10^{-3}$ H). Before the switch S is closed, there is no current in the circuit nor is there any charge on the capacitor.

(A) Just after switch S is closed ($t = 0$), what is the power supplied by the battery to the circuit?

(B) After the switch as been closed for a long time, ($t \to \infty$), what is the power supplied by the battery to the circuit?

(C) Again, after the switch has been closed for a long time, what is the charge on the positively charged plate of the capacitor?

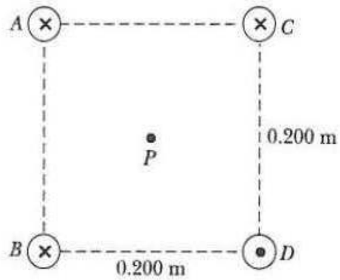


Problem 4: [25 points] Four long, parallel conductors carry equal currents of $I = 5.00$ A. The current direction is into the page at points A, B and C and out of the page at point D.

(A) Calculate the magnetic field, magnitude and direction, at point P, located at the center of the square with an edge length of 0.200 m.

(B) Calculate the magnetic force, magnitude and direction, on an electron moving into the page with speed $v = 1.00 \times 10^6$ m/s at point P.

Recall that $e = 1.60 \times 10^{-19}$ C. Also, $\mu_0 = 4\pi \times 10^{-7}$ Tm/A

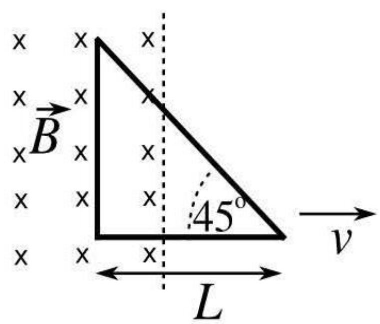


Problem 5: [25 points] A right (45-45-90) triangular loop of leg length L and with a resistance R lies in a plane perpendicular to a constant magnetic field of magnitude B pointing into the page. As shown in the figure below, the field exists only in a certain region (left of the dashed line), with a sharp boundary. An external force pulls the loop out of this region at a constant speed v in the direction shown.

(A) What is the induced current (in terms of the listed variables) through the loop as a function of time t if, at time $t = 0$, the bottom right corner of the triangle is just leaving the region with the magnetic field? Note: using the chain rule, $\frac{d}{dt}(x^2) = [\frac{d}{dx}(x^2)]\frac{dx}{dt}$.

(B) What is the maximum current through the loop (in terms of the listed variables)?

(C) Does this current flow in the clockwise or counterclockwise direction?



Appendix C: Reliability Test

Figures 16 and 17 show the relationship between the student scores on these problems and the difficulty rating of the two graduate students. The gray area on each graph shows the 95% confidence interval for the straight line fit. The correlation coefficient for non-expert rater 1 was -0.58 with a probably of no correlation of 0.007. For non-expert rater 2, the correlation coefficient was -0.42 with of no correlation probability of 0.068. This data shows that the simple directions for applying the difficulty measure were sufficient to give an indication of relative problem difficulty but not reliable enough for decision making. An additional study could exam whether there is a simple set of written instructions that would make this useful to a single, isolated instructor.

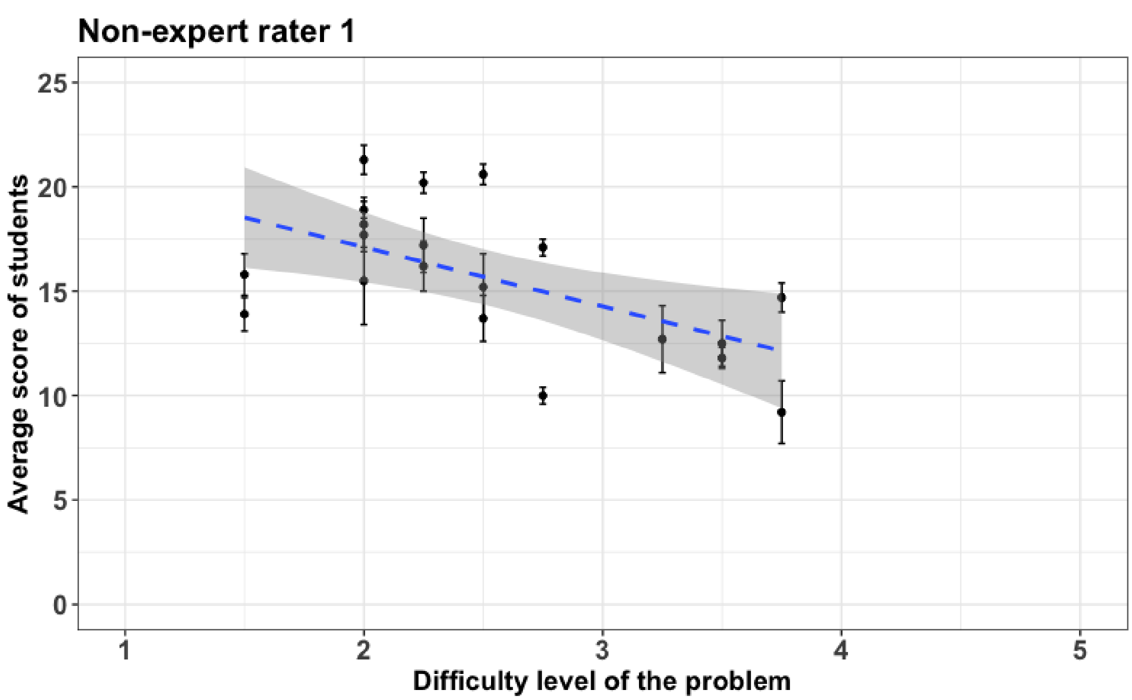


Figure 16: The relationship between students' average score and the assigned difficulty score by non-expert rater 1.

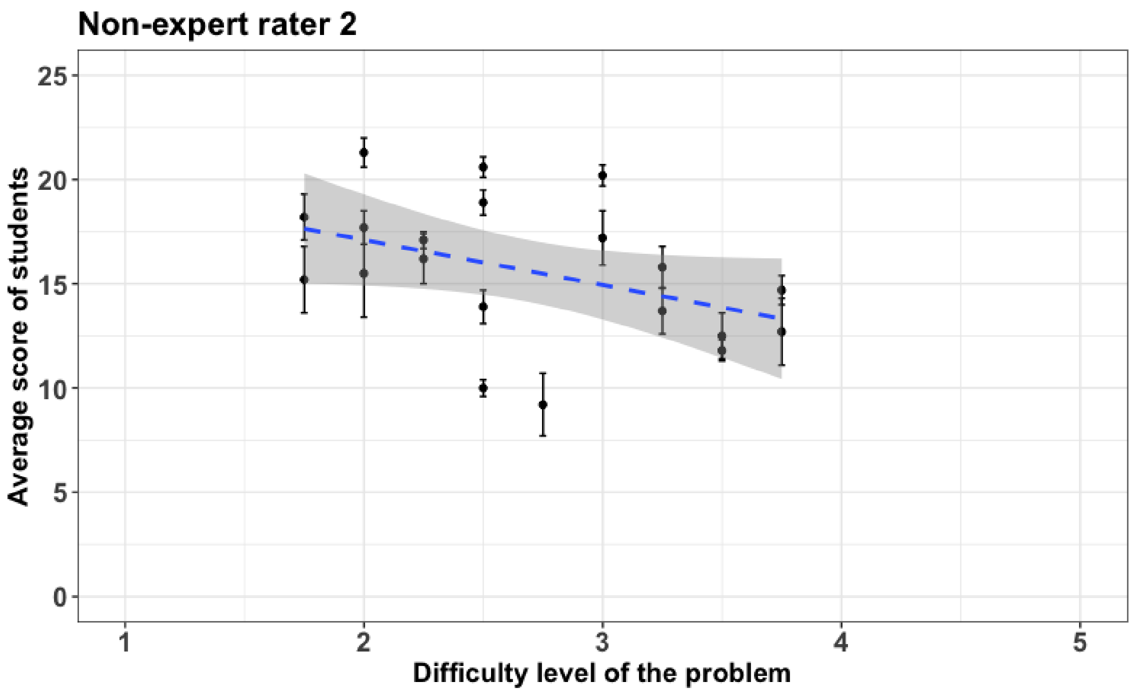


Figure 17: The relationship between students' average score and the assigned difficulty score by non-expert rater 2.