An Objective Rating Scale for the Difficulty of Introductory Mechanics Problems

By

Paul Fischer

A senior thesis submitted to the faculty of
Loyola Marymount University
in partial fulfillment of the requirements for the degree of

Bachelors of Science

Department of Physics

Loyola Marymount University

June 2016

Loyola Marymount University

THESIS COMMITTEE APPROVAL

of a thesis submitted by

Paul Fischer

This thesis has been read by each member of the following thesis committee and by majority vote has been found to be satisfactory.

Date	Dr. Jeff Phillips, Thesis Advisor Department of Physics
Date	Dr. David Berube Department of Physics
Date	Dr. John Bulman Department of Physics

Abstract

This research aimed to test the efficacy of an objective rating scale for the difficulty of introductory mechanics problems. Problems that require more bits of thought to obtain a solution have an increased likelihood that the solver will get an incorrect answer to do an excess of cognitive load. To quantify the difficulty of these problems, the minimum number of bits of thought required to obtain a correct answer were catalogued using an Excel spreadsheet based on the ACE-M framework. Using data from the online physics homework site WebAssign, the percent of students who answered seventeen selected problems incorrectly on their first try was graphed separately against each component of the ACE-M Framework. The result was an R^2 value of 0.00359 for the A component, 0.18446 for the C component and 0.15081 for the E component. Future research could include an in-person questionnaire that will provide the opportunity to probe students to help refine the coding of bits of thought that increase the cognitive load.

Table of Contents

List of Figures

Acknowledgments

I would like to express my sincere appreciation for Dr. Jeff Phillips and Dante Sblendorio allowing me to continue and add to their research project. I also would like to thank Dr. Phillips for all of the time he set aside for the meetings and edits that made this thesis possible. I want to share the gratitude I have for my parents because of their continued support which has allowed me to study Physics at Loyola Marymount University. I would also like to thank Dr. David Berube for his help as an advisor and Dr. Bulman for his general support as well as membership in my thesis committee. Finally, I would like to thank the entire LMU Physics Department for helping me to grow in both knowledge and passion for a subject that I love.

Chapter 1: Introduction

To address the needs of a teacher challenging his or her students at each level of expertise, certain textbooks contain difficulty ratings; however, these are subjective and do not match all students' views. These rating scales also cannot be used to effectively in other research projects. Students, teachers and researchers can be assisted by the creation of an objective rating scale for the difficulty of textbook problems. Further, if such a scale contained multiple dimensions, rather than the singular star rating found in textbooks, it could help to diagnose in what part of the problem-solving process students have difficulties and be used in targeted practice as students improve the specific strategies that they need to become more effective problem solvers.

Creating a system to measure difficulty can be difficult as what constitutes a problem depends on the solver and her knowledge and skill set. Caldwell and Goldin's experiment resulted in students rating concrete and factual problems as substantially easier than abstract or hypothetical problems.[1] There is current research in mathematics education comparing the subjective difficulties of different types of problems, but there has not yet been an explicit attempt to quantify the difficulty of all types of problems adhering to one coding scheme. Students have also been shown to struggle with mathematics more as they age. Lester found that the number of students asked to solve mathematics problems declined from having an average of sixteen percent satisfactory answers from the end of elementary school to only nine percent by the end of high school.[2] Gender has been shown to be less of a factor in determining student performance. Bray et al. gave intellectual problems to groups of varying numbers of participants from different genders and found that there was virtually no difference between the performance of the different genders. However, there was an interesting dynamic that sparked a discussion for future research: as the groups grew in number so did the number of nonparticipants, decreasing the size of the functional group.[3] These are a few examples on the study of students performance in math and science based on demographic information, which is not the topic of this research.

In addition, Gire et al. observed discrepancies in the subjective difficulty ratings students and teachers assigned to a list of juxtaposed problems.[4] The teachers in the study could not accurately predict which problems the students would rate as being among the most difficult since the students mostly used their familiarity with certain concepts as a measure of difficulty. Nathan and Petrosion tested the "expert blind spot" hypothesis, which expresses the struggle an experienced teacher faces when trying to empathize with the limited amount of knowledge possessed by their students. Their research revealed that the teachers considered increasingly complex equations to be a measure of difficulty which, for a reason similar to the aforementioned study, was not the case according to the students.[5] This had to do in part with the nature of expert vs. novice problem solving, which is not the focus of this research. There will not be an attempt in this paper to rate the difficulty of problems to novices or experts selectively; the goal instead will be a scale that represents the factors that make one exercise more difficult than another to a solver regardless of his or her level of knowledge and experience with problem solving techniques.

There are two terms central to this thesis whose context will now be explicated: “monitoring” and "cognitive load". According to the research of Phillips et al., monitoring is the process of evaluating and correcting one’s own work.[6] This process occurs more often if a solver is not experienced at the particular type of problem being solved or if the problem contains many components. The more processing or decision-making a solver needs to make, the more monitoring that is required. The greater the number of chunks of information, called "bits of thought", required to solve a problem the more challenging the problem.

The retention and command of several chunks of information takes a toll on one's brain due to what is called the cognitive load. This can be thought of as the amount of "bits" (similar to "schemas" in Psychology research) one can accuratmely hold and merge in one’s mind simultaneously. According to Sweller, learning in mathematics comes from schema acquisition.[7] This means that chunks of information about concepts taught in class and their applications must be stored in a student's brain in order for him or her to perform well on a mathematics test. If that student faces a problem with a large number of different types of schemas that require effort to integrate, his or her cognitive load can be exceeded.

Granholm et al. tested the pupillary responses of students who were asked to remember increasingly large sequences of numbers and effectively repeat them from memory.[8] The study showed that the limit of where the average student will exceed his or her cognitive load and cease to effectively monitor more bits of thought is between five and nine numbers. Powell et al. tested how the position of information affected difficulty in mathematical word-problems.[9] This did not affect the student scores, so the bits of thought within each problem sum up regardless of the order in which they are presented.

Given these results, it is hypothesized that there is a direct relationship between the number of bits of thought that a student must monitor in a problem and the likelihood that he or she will arrive at an incorrect answer. The expected result is a linear increase in the percentage of students who answered a question incorrectly on their first try with respect to the number of bits of thought they must monitor as they approach and exceed their respective cognitive loads. To objectively break down the number of bits of thought contained in an introductory mechanics problem, a problem solving framework must be used which represents the minimum number of decisions solvers of any skill level have to make to obtain a correct solution.

One of the earliest and most famous mathematical problem solving schemes comes from Hungarian mathematician George Polya: Understand the Problem, Devise a Plan, Carry Out the Plan and Look Back.[10] Although this scheme is viable for exercises, it is not practical for problems that involve extensive monitoring, like in the following example:

"You are a lawyer with a background in physics. Your client is accused of firing a gun inconspicuously at the plaintiff, where the bullet missed and hit a wooden chair resting on a wooden floor such that the coefficient of kinetic friction between the two surfaces is 0.2. You collect that the mass of the chair is 20 kg, the mass of the bullet lodged in the chair was 5 g and the skid mark it left was 5 cm long. The plaintiff claims that your client owns a gun such that it must have been him who fired the bullet, but you know that your client's gun has a muzzle velocity of only 200 m/s. Is the defendant guilty of firing the bullet from the gun in question?"

This problem does not explicitly state the variable that will be used to solve the problem. It is instead left to the discretion of the solver, who might need to experiment with carrying out multiple possible plans before making that decision. To include the added difficulty caused by the cognitive load of monitoring that information back and forth between the different steps of the solution process, the ACE-M framework will be used as the problem solving system for this research. The components of the ACE-M framework are: Analyze the Problem, Create a Plan, Execute the Plan and Monitoring.

Solving a problem using this framework can be compared to getting hired as an engineer and being given an ambiguous project that requires you to make decisions about how to construct the most efficient pathways between roads over a lake. Analyze the Problem (A) involves the clarity of the beginning and endpoints. This means that the steps in A, one of which is choosing the desired variable(s), are analogous to deciding which roads (initial and final conditions) are to be connected through the construction of the bridge. For simple plug and chug problems, the desired goal(s) may be explicitly stated, which is like having only one road on both sides of the lake that needs to be connected, eliminating the need to make such a decision at the discretion of the solver. The sample problem posed above, however, contains multiple possible variables from which to build a solution, analogous to having to make a choice of the most efficient connections between multiple options of roads. The Monitoring (M) in this case is analogous to experimenting with the planning of multiple bridges before figuring out which boundary conditions best fit what was originally an ambiguous blueprint. Create a

Plan (C) can be thought of as completing a detailed blueprint for the construction of the bridge. Execute the Plan (E) is analogous to the actual construction of the bridge, where challenges arising in the algebra can be likened to the situations that arise while following through with the blueprints, including hiring employees, ordering parts, etc. that may require back and forth planning and experimentation before the construction is complete. Thusly, the ACE-M Framework provides a three-vector system where the challenges within each section are explored separately while taking into account the monitoring that goes on between them. It is expected that such a multi-dimensional coding scheme will more accurately represent the perceived difficulty than a singular value as has been attempted in previous research. [11, 12]

This thesis is organized as follows: Chapter 2 explores the ACE-M spreadsheet that was used to mark the difficulty of introductory mechanics problems and provide a sample coding of a problem, Chapter 3 discusses the acquisition of the student data and provides an analysis of the results and Chapter 4 discusses the conclusion of the research and the possibilities for supplemental future research projects.

Chapter 2: The ACE-M Spreadsheet

The research done by Sblendario and Phillips culminated in a Microsoft Excel spreadsheet that attempted to catalogue all of the possible bits of thought that can occur in the solution to an introductory mechanics problem.[13]. The motions, interactions, etc. inherent to each one of the three vectors of difficulty are checked off of the spreadsheet as demonstrated in Figure 2.1. In this chapter, the details of each section of the spreadsheet will be discussed, followed by a walkthrough of the coding of a sample problem which will demonstrate how the spreadsheet is used.

The A section of the ACE-M Framework involves reading the question, making representations and drawings, accounting for an explicitly desired quantity and the clarity of the boundary conditions. As shown in Figure 2.1, the first cell used in the coding of the data accounts for an explicitly desired quantity, like "find the initial velocity" as opposed to the implicit variable asked for in the sample problem provided in the Introduction, "is the defendant guilty?" The following cells to be coded in the A section are First Order Motions and Interactions. These are stated by the language used in a given problem and can either be explicit or implicit. For example, a problem stating that there is a particle undergoing “non-constant rectilinear motion” in addition to “kinetic friction” has explicitly stated a motion and interaction. These would be marked under the corresponding cells in the A section. There are also implicit First Order Motions and Interactions. These are stated by the problem but in a wording different from that of the

	1	2	3	14	13	12	11	10	9	8	7	6	5	4
A			problem #	585046049	5853963	58558358	58544183	583611037	582600180	582073944	585611037	582934342	582561000	584099183
B	1. Analysis		Desired quantity explicitly stated											
C		1. Explicit or Implicit (First Order) Motions	A. Constant velocity											
D			B. Constant Acceleration											
E			C. Uniform Circular Motion											
F			D. Non-constant (or unspecified) rectilinear motion											
G			E. Non-constant (or unspecified) rotional motion							1				
H			F. Constant angular velocity											
I			G. Constant angular acceleration											
J			H. Relative velocity											
K			I. Rolling without slipping (incl. pulleys)											
L			TOTAL #	0	0	0	0	0	0	1	0	0	0	0
M		2. Explicit or Implicit (First Order) Interactions	A. Weight near earth(or any other massive body)											
N			B Universal gravitation											
O			C. Friction- kinetic									1		
P			D. Friciton- static							1				
Q			E. Tension/ ideal rope					1			1		1	
R			F. Spring/ Hooke's Law	1		1	1		1					
S			G. Drag- linear											
T			H. Applied						1					
U			I. Collision/ contact		1							1		
V			TOTAL #	1	1	1	1	1	2	1	1	2	1	0
W	2. Creating a Plan		Desired quantity implicitly determined											
X			Extraneous or Researched Information											
Y		1. Implicit (Second Order) Motions	A. Constant velocity									1		
Z			B. Constant Acceleration	1								1		
AA			C. Uniform Circular Motion											
AB			D. Non-constant (or unspecified) rectilinear motion		1	1	1		1					1
AC			E. Non-constant (or unspecified) rotional motion					1			1			
AD			F. Constant angular velocity										1	
AE			G. Constant angular acceleration											
AF			H. Relative velocity											
AG			I. Rolling without slipping (incl. pulleys)											
AH			TOTAL #	1	1	1	1	1	1	0	1	2	1	1
AI		2. Implicit (Second Order) Interactions	A. Weight near earth(or any other massive body)	1	1	1	1	1		1			1	
AJ			B Universal gravitation											
AK			C. Friction- kinetic											
AL			D. Friciton- static											
AM			E. Tension/ ideal rope											
AN			F. Spring/ Hooke's Law											
AO			G. Drag- linear											
AP			H. Applied											
AQ			I. Collision/ contact											1
AR			TOTAL #	1	1	1	1	1	0	1	0	0	1	1
AS		3. Models	A. Linear kinematics									1		
AT			B. Rotational Kinematics											
AU			C. Newton's Second Law- linear			1	1					1		
AV			D. Newton's Second Law Rotational					1		1	1		1	
AW			E. Newton's Third Law											
AX			F. Conservation of Rectilinear Momentum									1		1
AY			G. Conservation of Energy	1	1	1			1					1
AZ			H. Conservation of Angular momentum											
BA			TOTAL #	1	1	2	1	1	1	1	1	3	1	2
BB	3. Execution		1. Total Equations Used	3	1	6	3	3	6	3	3	3	2	4
BC			2. Unit Conversion	0	0	0	0	0	0	0	0	0	0	0
BD			3. Geometry	0	0	0	0	0	0	0	0	0	1	0
BE			4. Trigonometry	0	0	0	0	0	0	0	2	0	2	0
BF			TOTAL #	3	1	6	3	3	6	3	5	3	5	4
BG			Analyzing the Problem Total	1	1	1	1	1	2	2	1	2	1	0
BH			Creating a Plan Total	3	3	4	3	3	2	2	2	5	3	4
BI			Execution Total	3	1	6	3	3	6	3	5	3	5	4

Figure 2.1 The full ACE-M Spreadsheet (displayed sideways)

spreadsheet. For example, another problem claiming to have a particle move "periodically along a straight line" while its surface is "losing energy due to contact with the ground" is implying the same motion and interaction as the previous example, but using different vocabulary. The spreadsheet contains all of the explicit motions and interactions, but the corresponding implicit motions and interactions that are mentioned in the problem are coded in these cells as well. Once all of these quantities have been accounted for, the A score is revealed.

The coding of the C section begins with the elements that are not involved in the direct language of the question but must be introduced by the solver to arrive at the correct solution. According to Figure 2.1, the first value to be marked is an implicitly desired quantity. This is because the student must experiment with multiple plans to make a decision about which variable(s) to pursue, which was demonstrated in the sample problem in Chapter 1. Next, the spreadsheet takes into account the times students get an incorrect answer because they neglected to search for a variable in a table. For example, a student might use the wrong equation for a problem involving a massive body attracted to Mars because he or she was not explicitly given the mass of the planet. The student may try to use an incorrect equation that avoids the missing mass instead of just looking it up in a table. Unused variables are also considered for a similar reason. For example, problems involving projectile motion often give the mass of an object even when it is not required to solve the problem, which could confuse the solver into using an incorrect equation just because it contains the mass. The next quantities to be coded in the C section are Second Order Motions and Interactions. These are not given explicitly or implicitly in the language of the problem, they are up to the discretion of the solver's thorough conceptual understanding of the physical situation that is being described. A Second Order Interaction occurs in a problem that includes two surfaces in contact but neglects to mention whether or not the friction between them must be considered. To get the right answer, the solver must make the choice to include friction because of an understanding of the physics between the two objects in question. Thusly, any motion or interaction that is not explicitly or implicitly stated in the problem is Second Order and is to be marked in the C section. Once all of the motions and interactions needed to obtain a solution are accounted for, then the student must make a decision about with which mathematical models all of the motions and interactions correspond (i.e. Newton's Second Law Linear, Linear Kinematics, etc.).

Executing the Plan deals with the mathematical details of the solution. The coding of this section begins with the total number of equations required to get the right answer, since each equation lends itself to algebraic errors. Then, the number of unit conversions is considered due to the necessary prior knowledge of how to correctly perform such operations. Finally, any formulas from geometry and trigonometry are accounted for since their applications require background understanding that could considerably increase the difficulty of the algebra.

The coding of problem 4.119 from the WebAssign data will now be demonstrated:

> While moving in, a new homeowner is <u>pushing</u> a box across the floor at a <u>constant velocity</u>. The <u>coefficient of kinetic friction</u> between the box and the floor is 0.49. The pushing force is directed downward at an angle θ below the horizontal. When θ is greater than a certain value, it is not

possible to move the box, no matter how large the pushing force is. Find that value of θ.

Information pertaining to the A section has been underlined for convenience. First, it is noted that the desired quantity is explicitly stated: "Find that value of θ." The problem expresses explicitly that the box is moving at a constant velocity, so the corresponding "constant velocity" cell in the First Order Motions section of the spreadsheet in Figure 2.2 will receive a mark. Then, there is an explicit reference to the coefficient of kinetic friction. This is marked under the "Friction" cell within the First Order Interactions section. It is then stated that there is a "pushing" force, which is an implicit reference to the "Applied" force cell in the First Order Interactions section. The final A score is then 4.

		1. Analysis																			
		1. Explicit or Implicit (First Order) Motions										2. Explicit or Implicit (First Order) Interactions									
problem #	Desired quantity explicitly stated	A. Constant velocity	B. Constant Acceleration	C. Uniform Circular Motion	D. Non-constant (or unspecified) rectilinear motion	E. Non-constant (or unspecified) rotional motion	F. Constant angular velocity	G. Constant angular acceleration	H. Relative velocity	I. Rolling without slipping (incl. pulleys)	TOTAL #	A. Weight near earth(or any other massive body)	B Universal gravitation	C. Friction- kinetic	D. Friciton- static	E. Tension/ ideal rope	F. Spring/ Hooke's Law	G. Drag- linear	H. Applied	I. Collision/ contact	TOTAL #
4.119	1	1									1			1					1		2

Figure 2.2: The coding of the A section from problem 4.119 of the WebAssign data

	2. Creating a Plan																														
			1. Implicit (Second Order) Motions										2. Implicit (Second Order) Interactions										3. Models								
problem #	Desired quantity implicitly determined	Extraneous or Researched Information	A. Constant velocity	B. Constant Acceleration	C. Uniform Circular Motion	D. Non-constant (or unspecified) rectilinear motion	E. Non-constant (or unspecified) rotional motion	F. Constant angular velocity	G. Constant angular acceleration	H. Relative velocity	I. Rolling without slipping (incl. pulleys)	TOTAL #	A. Weight near earth(or any other massive body)	B Universal gravitation	C. Friction- kinetic	D. Friciton- static	E. Tension/ ideal rope	F. Spring/ Hooke's Law	G. Drag- linear	H. Applied	I. Collision/ contact	TOTAL #	A. Linear kinematics	B. Rotational Kinematics	C. Newton's Second Law- linear	D. Newton's Second Law Rotational	E. Newton's Third Law	F. Conservation of Rectilinear Momentum	G. Conservation of Energy	H. Conservation of Angular momentum	TOTAL #
4.119												0	1									1			1		1				2

Figure 2.3: The coding of the C section from problem 4.119 of the WebAssign data

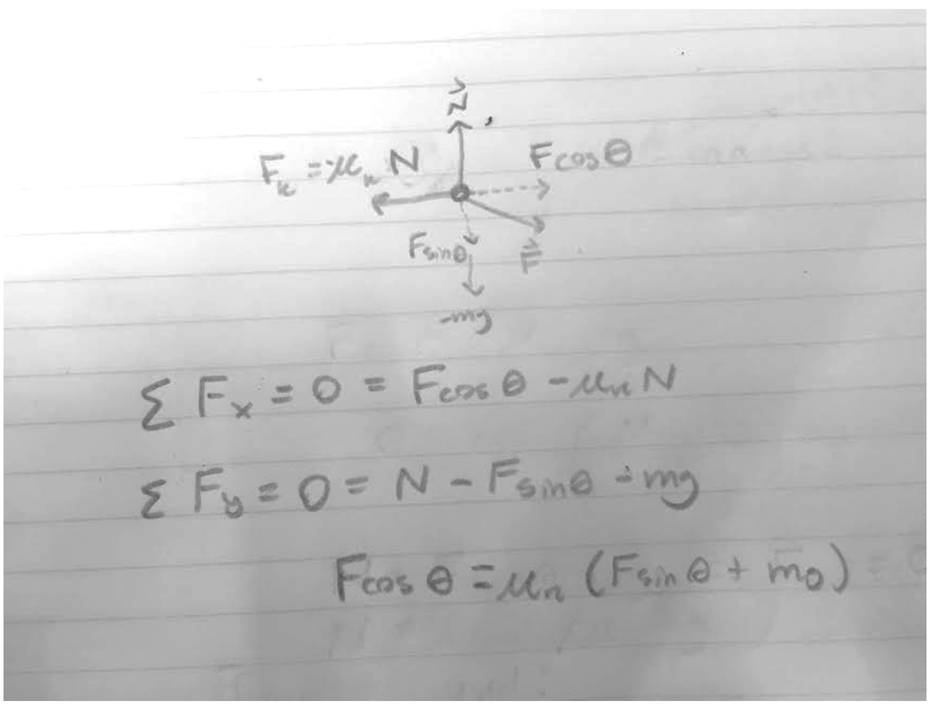

Figure 2.4: The free body diagram and equations used to solve problem 4.119 of the WebAssign data

To represent the sum of the interactions, the weight of the box had to be considered despite not having been mentioned in the language of the problem. (Figure 2.4) This is marked as a Second Order Interaction under the cell for "Weight near earth" in Figure 2.3. Two models were also required to make the diagram: Newton's Third Law pertaining to the equal and opposite Normal Force in the y-direction and Newton's Second Law Linear used to model the weight of the box as $F_G = -m*g$. These are correspondingly marked in their respective cells in the Models section in Figure 2.3. This reveals a total C score of 3.

3. Execution				
1. Total Equations Used	**2.** Unit Conversion	**3.** Geometry	**4.** Trigonometry	**TOTAL #**
2	0	0	2	4

Figure 2.5: The coding of the E section from problem 4.119 of the WebAssign data

For the E section of the spreadsheet, first refer to Figure 2.4 for the equations for the superposition of forces in the x and y – directions. This reveals a mark of "2" underneath the "Total Equations Used" cell in Figure 2.5. The final equation in Figure 2.4 shows an equation with two separate trigonometric functions. This significantly increases the difficulty of the algebra that will need to be carried out. Thusly, the "trigonometry" cell will receive a mark of "2". This reveals a total score of 4 for the E section, and a total difficulty rating of 4 –3 – 4 for the entire problem according to the ACE-M Spreadsheet.

Chapter 3: Data and Analysis

WebAssign is a software that allows students to view homework assignments, enter their answers and receive feedback online. It contains a database of problems from many different Physics textbooks, data on the percentage of students who answered each question incorrectly on their first try and the total number of students who attempted each solution. For the purpose of collecting reliable data, there was an attempt to keep the average number of students per problem at the very least in the hundreds. For the graphs in Figures 3.1 – 3.3, seventeen problems were coded from three different textbooks:

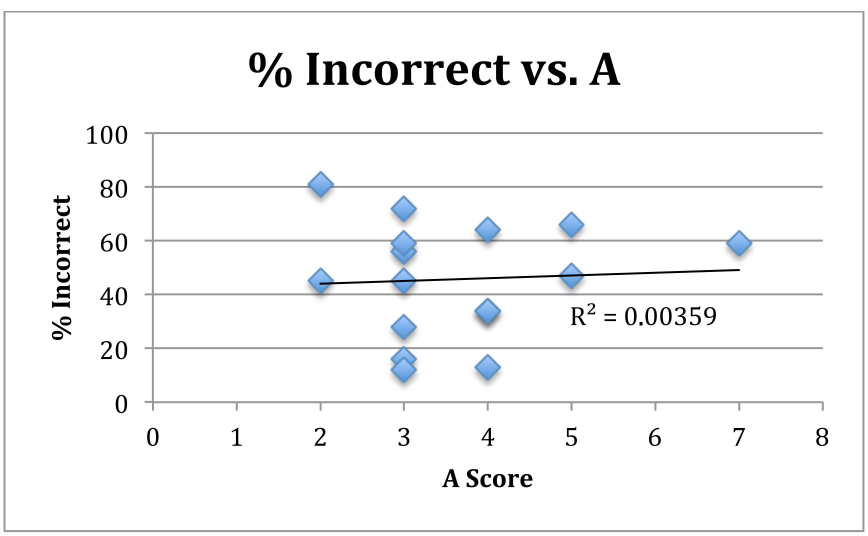

Figure 3.1: The graph of the percentage of students who answered a question incorrectly on their first try vs. the total A scores

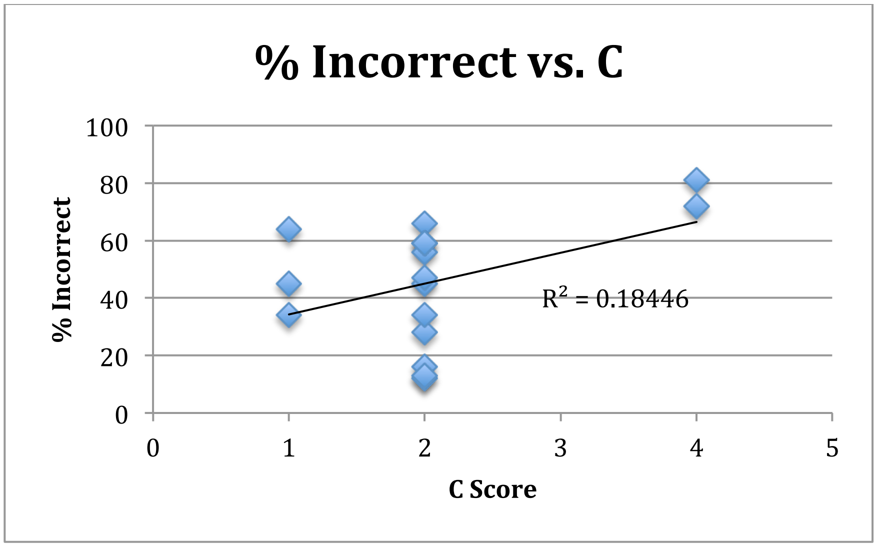

Figure 3.2: The graph of the percentage of students who answered a question incorrectly on their first try vs. the total C scores

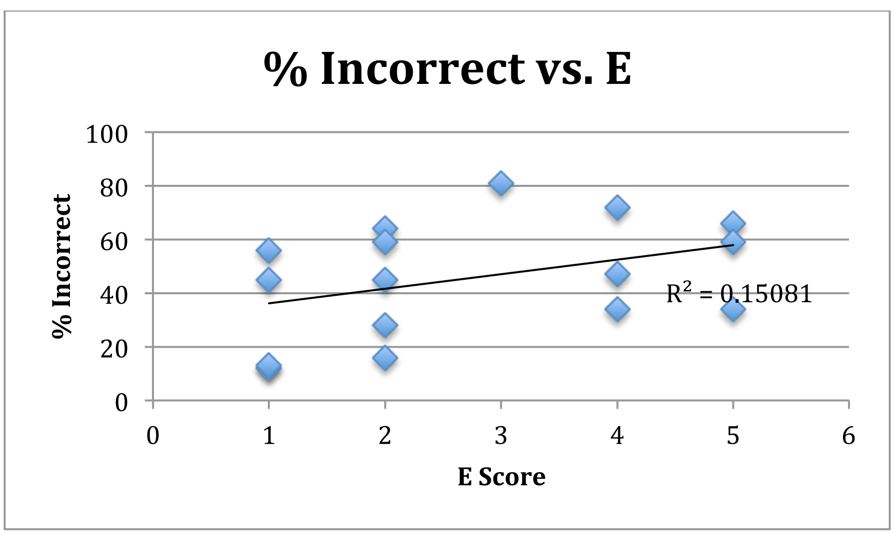

Figure 3.3: The graph of the percentage of students who answered a question incorrectly on their first try vs. the total E scores

Physics for Scientists and Engineers (8th Edition) by Raymond A. Serway and John W. Jewett, *Fundamentals of Physics* (10th Edition) by David Halliday, Robert Resnick and Jearl Walker and *University Physics with Modern Physics* (13th Edition) by Hugh D. Young, Roger A. Freedman and A. Lewis Ford. There was an attempt made to collect problems that spanned a wide range of the amount of necessary bits of thought while engaging multiple concepts. The percentage of students who received an incorrect score on their first tries was graphed against the score of each section of the ACE-M framework to test the correlation of each component with the data.

Problems integrating multiple concepts were chosen to minimize any bias in the data that could arise from a level of comfort students have with problems they have seen before in textbook examples. The definition of the A section is still ambiguous since it has not yet been proven which elements add bits of thought that require monitoring. It has been considered that marking the explicitly stated quantities might not represent the difficulty of the problem since these need not be monitored by the student and are instead accounted for on paper such that there is not strain on the student's cognitive load. Since answers are entered into WebAssign online, the data might also have been affected by answers that were off by insignificant decimal amounts that arose from round-off errors, even if the student followed through with the correct process. The steps in C that could go wrong from students managing to forget to plug in variable or use unnecessary variables may also cause the data to show discrepancies since these problems are more likely to affect novices as opposed to experts.

The R^2 value for the A section is 0.00359, which is much less than the ideal value of 1. This might be due in part to the aforementioned difficulties with defining the bits of thought in the A section as well as round off errors when entering answers into WebAssign. The C and E sections had R^2 values of 0.18446 and 0.15081 respectively, which are still much less than 1 but also much larger than the A correlation. These low values could be due in part to the round-off errors inherent in typing in answers online. The A score might reflect a better correlation if more research was done to define the elements that add bits of thought to that section.

Chapter 4: Conclusions and Future Research

This research sought to chart the minimum number of bits of thought required to solve a physics problem in an effort to test the efficacy of the ACE-M spreadsheet as an objective rating scale for the difficulty of introductory mechanics problems. The correlation for the A section was much smaller than for the C and E sections, so there could be research done in the future defining the elements that make the A section more difficult. Sources for error include the round-off errors in entering answers on WebAssign and the ambiguity in defining the bits of thought that impact the cognitive load of a student in the A section.

These two problems could be alleviated by a future study involving an in-person questionnaire collecting students' answers in person. This would provide an opportunity to probe them for the difficulties they face as they attempt or after they attempt to solve problems. This could be used to reformulate the spreadsheet and then collect more data to be tested and hopefully yield a stronger correlation. If there is a stronger correlation, an equation could be developed to weight the difficulty of different models, as opposed to assigning a value of 1 to each bit of thought. This could provide insight into which subjects within introductory mechanics teachers should be sensitive to providing thorough understandings and student assessments. If such an objective difficulty rating scale is developed, it could be used as a basis for other subjects such as electricity and magnetism, quantum mechanics, statistical mechanics, etc. If this scale becomes the standard for all physics textbooks, it could be used in mathematics and chemistry as well to supplement teachers and researchers alike on a whole variety of different projects.

References

1. Caldwell, J. H., & Goldin, G. A. (1987). Variables affecting word problem difficulty in secondary school mathematics. *Journal for Research in Mathematics Education*, 187-196.
2. Lester, F. K. (1994). Musings about mathematical problem-solving research: 1970-1994. *Journal for research in mathematics education*, *25*(6), 660-675.
3. Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty, and sex on group performance and member reactions. *Journal of Personality and Social Psychology*, *36*(11), 1224.
4. Gire, E., Rebello, N. S., Singh, C., Sabella, M., & Rebello, S. (2010, October). Investigating the perceived difficulty of introductory physics problems. *AIP Conference Proceedings* (Vol. 1289, No. 1, p. 149).
5. Nathan, M. J., & Petrosino, A. (2003). Expert blind spot among preservice teachers. *American Educational Research Journal*, *40*(4), 905-928.
6. Phillips, J.A., McCallum, J.E.B., Clemmer, K.W., Zachariah, T.M. " A Problem-Solving Framework to Assist Students and Teachers in STEM Courses" submitted to the *Journal of College Science Teaching*
7. Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295-312.
8. Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457-461.

9. Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T., & Fletcher, J. M. (2008). Do word-problem features differentially affect problem difficulty as a function of students' mathematics difficulty with and without reading difficulty?.*Journal of Learning Disabilities*.

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

11. Knaus, K., Murphy, K., Blecking, A., & Holme, T. (2011). A Valid and reliable instrument for cognitive complexity rating assignment of chemistry exam items. *Journal of Chemical Education*, 88(5), 554-560.

12. Gire, E., Rebello, N. S., Singh, C., Sabella, M., & Rebello, S. (2010). Investigating the perceived difficulty of introductory physics problems. *AIP Conference Proceedings* 1289 (1) 149.

13. Sblendorio, D. and Phillips, J. A. (2013) "Locating Introductory Mechanics Problems Along the Well-structured- Ill-structured Continuum" presentation at the Summer Meeting of the American Association of Physics Teachers in Portland, OR.