Using Test Presentation Measurements to Understand Student Learning in Introductory Science Classes

An Honors Thesis submitted in partial fulfillment
of the requirements of Honors Studies in Physics.

By
Shawn C. Ballard

Spring 2008
Physics
J. William Fulbright College of Arts of Sciences
University of Arkansas

Acknowledgements

This is for mom and dad, who never let me fall, even when I wanted to.

Table of Contents

Abstract

This work examines the effect of students' written test presentation, their use of language, mathematics, and drawing, on success in an introductory physics class, University Physics II, in order to determine which characteristics of the presentation correlate most strongly with the students' overall performance in the class. From student exams, the number of presentation elements including words; sentences; symbols; operators; equal signs; numbers; lines, curves and objects drawn on a graph; numbers and symbols used to describe graphed work; and words used to describe graphed work are counted. Additional variables that capture important features of written presentation are calculated from these observations.

Using a statistical analysis program, SAS, correlations are calculated between all combinations of the tabulated variables to determine which variable or combination of variables has the greatest impact on overall performance. A multiple regression analysis determines the extent to which student performance is predicted from students' presentation of written work.

Chapter 1: Introduction

According to surveys by the National Science Foundation, Americans are very interested in science and technology. However, they do not generally feel knowledgeable about these topics. Surveys from 2001 indicate that less than 15% of the population thinks they are well-informed about current issues in science in technology [10]. This feeling of inadequacy on matters scientific is viewed as a serious problem given the prevalence of constantly developing technology in most people's everyday lives. Additional motivations for a focus on science and technology are more global. Most people recognize that global warming and environmental pollution are serious issues, though they may not be primary concerns at the national level.

Other surveys show that roughly 70% of Americans do not understand the scientific process. Furthermore, performance on questionnaires designed to measure science literacy has declined in recent years [10]. While scientific knowledge is not necessary for becoming a productive member of society, the critical thinking and problem solving skills gained from science classes have applications in many fields. Boosting performance in science and subsequently increasing practical mental ability is seen by Americans as very important. Possible improvements in this area are generally linked to the quality of science education in the United States, which many feel is lacking.

Scientists in the physics community are answering this call for education reform. Particular attention is paid to high school and introductory college courses where a firm foundation in science may be cultivated. Discovering what most inhibits conceptual gains in physics and how best to maximize these gains constitute much of the research occurring in this field.

Research in physics education frequently seeks to determine what the greatest predictors of student performance are so that learning may be enhanced through either different models of instruction or a modification of course content and requirements. In their 1982 study (Table 1), Hudson and Liberman address the impacts of different modes of learning on performance in a physics course [5]. They identify formal operational reasoning and mathematics as the two primary measures of physics achievement and seek to determine what their relative importance is in the final grade distribution. Here formal operational reasoning refers to a student's logical ability and mathematics takes into account his or her math background and general ability.

Variable	**r**	**R^2**
Mathematics Pretest	0.345	0.119
Formal Operational Reasoning	0.517	0.268

Table 1. Multiple regression coefficients for total physics grade as dependent variable, and mathematics pretest and formal operations scores as independent variables [5]. The statistic R^2 measures the amount of initial variation in the data that is removed by a model using the variable in question. R^2 varies from 0 to 1. The statistic r is the correlation coefficient.

Hudson and Liberman found that "while neither mathematics ability nor reasoning ability alone will suffice to enable a student to do well, these two variables act in independent ways to influence student performance dramatically" [5]. This interaction of results rather than a purely mathematics-intensive approach implies that conceptual reasoning may play a major role in performance on primarily quantitative exams traditionally given in physics courses.

Meltzer examines the role of mathematics in conceptual gains from physics courses [8]. His research indicates that correlations between mathematics skill and physics performance do not hold consistently. To address this gap, he investigates other potential factors such as pre-existing physics and mathematics background. Meltzer found that students' normalized gains do not depend on their pretest scores (see Table 2).

Correlations Between Normalized Learning Gain and Physics Pretest Score				
Sample	**N**	**r**	**R^2**	**P**
SLU 1997	45	0.15	0.02	0.35
SLU 1998	37	0.10	0.01	0.55
ISU 1998	59	0.00	0.00	0.98
ISU 1999	78	0.10	0.01	0.39

Table 2. Displayed are the correlation coefficients, r, and their significances, P, for correlations between performance on a physics pretest and the normalized learning gain [8]. The number of students in each sample is given by N. The samples are individual semesters at two different institutions.

Not only are these correlations low, but they are also not significant. In other words, a student's ability to achieve conceptual gains is independent of the level of physics concept knowledge with which he begins the class. Such results would imply an equal playing field for students of various physics backgrounds. A student with no prior physics instruction could learn as much in an introductory course as a student with extensive previous experience.

The same independence of background preparation is not observed for the mathematics preparation for a physics class. In Meltzer's study, students with more pre-instruction in mathematics had much greater learning gains on the physics concepts, regardless of their initial knowledge of physics, than students with a lower entering mathematics skill level. Students' math abilities were evaluated using a pretest. Analyzing these test results versus overall course grade revealed that prior math instruction has a much stronger correlation than prior physics instruction (Table 3).

Correlation Between Normalized Learning Gain and Mathematics Pretest Score				
Sample	**N**	**r**	**R^2**	**P**
SLU 1997	45	0.38	0.14	<.01
SLU 1998	37	0.10	0.01	0.55
ISU 1998	59	0.46	0.21	0.0002
ISU 1999	78	0.30	0.09	<0.01

Table 3. Displayed are the correlation coefficients, r, the variance, R^2, and their significance for correlations between performance on a math pretest and the normalized learning gain [8].

Based on these results, Hudson and Liberman's implications about the benefits of conceptual instruction on quantitative problems must be re-evaluated. Increased mathematics knowledge will necessarily correspond to higher test outcomes on tests whose results correlate strongly with quantitative, math presentation rather than qualitative, conceptual presentation. Assuming that both conclusions are true and there is some combined correlation at work in the relationship between conceptual and quantitative presentation, the next step is to determine how best to maximize the performance outcomes due to this interaction.

Halloun and Hestenes conclude that conventional instruction falls dismally short of meeting expectations for course gain. "Diagnostic test results show that a student's initial knowledge has a large effect on his performance in physics, but conventional instruction produces comparatively small improvements in his basic knowledge" [4]. They investigate a number of factors influencing student outcomes (Table 4) [4].

Regression Values for Analysis of Course Achievement Indicators		
Variable	**University Physics, R^2**	**College Physics, R^2**
Physics pretest	0.30	0.32
Math pretest	0.26	0.22
Physics and math pretests	0.40	0.42
Physics courses	0.07	0.12
Math courses	0.10	0.04
Physics and math courses	0.15	0.16
All pretests and courses	0.49	0.51

Table 4. Variance as measured by R^2 for regressions of class performance in both University and College Physics versus pre-test and pre-instruction markers [4].

Halloun and Hestenes found that if physics and math pretests and the results of all previous classes taken are considered, then an R^2 of 0.5 results. Adding to their results, Hake studied different methods of course instruction, which indicated that the use of interactive engagement (IE) strategies increases learning gains well beyond those obtained with traditional methods [3]. Hake recommends the use of hands-on laboratory activities and peer/instructor discussion with immediate feedback as preferable to lecture alone. Further, Hake suggests that the impacts of substantial improvement in physics education would be culturally pervasive and interdisciplinary, not limited to areas of hard sciences and mathematics.

Hake developed the normalized, or Hake, gain (denoted <g> and defined further on page 13) to measure how effective a course is in promoting conceptual understanding. Based on normalized gain results from various institutions and skill levels, Hake concludes that "problem-solving capability is actually *enhanced* (not sacrificed as some would believe) when concepts are emphasized" [3]. For this reason, test presentation even on a primarily quantitative exam can relate to conceptual gain, as indicated by Hudson and Liberman.

Based on these previous results and proposals for improvement, potential factors influencing science classes fall into essentially two categories: mathematics-based and more conceptual logic-based causes. Pretesting in mathematics and physics can help determine how much success in physics courses is due to prior experience with the techniques and concepts. General abilities in critical thinking and logic can also boost performance.

There is a further factor that influences how a student performs in a science class, the actions a student takes to address the class. These factors include the time a student spends on a class, the actions a student performs (taking lecture notes for example), and the care a student invests in his or her work. Many classes now invest substantial resources in managing these factors. The class under study uses laboratory and lecture quizzes to manage attendance and has policies in place to encourage careful, well-explained work.

The effect of student action and time choices was investigated by Stewart [11]. The variance explained by behavior is reported in Table 5. The variables that went into the regression include total time, reading time, practice test use, and 23 other measures of behavior. Two versions of a class were studied where Version 1 was less challenging than Version 2. Student behavior was more predictive of outcome in the more challenging version of the course [11].

Regressions with Measures of Behavior to Predict Student Performance		
	R^2-Test Average	**R^2-Normalized Gain**
Version 1 (Easier)	0.26	0.19
Version 2 (More Challenging)	0.41	0.44

Table 5. This table displays the regression values for how a student's behavior relates to his test average and normalized gain [11].

Another aspect of student behavior not investigated by the Stewart study is the role of student writing behavior in student outcomes. This study investigates the effects of student presentation in order to address two issues:

- What factors in a student's exam presentation (mathematical, graphical, or linguistic) most affect student learning?
- To what extent can a student's performance on class examinations and conceptual instruments be predicted from his or her written presentation?

Answering these questions has the potential to provide insight into the role of language, mathematical, and graphical presentation in students' learning and to inform modifications in the classroom that will improve learning.

Chapter 2: Set Up and Data Collection

Over the course of each semester, students in the introductory electricity and magnetism course at the University of Arkansas, University Physics II, take four examinations, each with a large open-response component. The written work on these open-response questions is examined in this study. A student's solution to a physics problem is a complex combination of English, mathematics, and drawings. In UPII, instructors encourage students to give written solutions symbolically before performing numeric calculations. These numeric calculations are also expected to include proper units where appropriate. Additionally, certain exam questions ask specifically for graphical or English description. Though they may enhance the general readability of the solutions, these encouraged behaviors are not graded and so do not directly affect the students' performance in the course.

Observation of the student's work is organized into these three major categories, mathematics, English, and graphics. Each group can be divided into subcategories. The mathematics group is divided into symbols, operators, equal signs, numbers (distinguished as correct or incorrect); the drawing or graphed portion into graphed objects, graphed symbols, graphed words; and the language group into words and sentences. The definition of the constituents of each group follows.

Math Variables

Symbols: Variables (e.g. x, y, t), constants (π, ε_0, etc.), and letters in languages other than English (e.g. π, φ, θ).

Operators: Mathematical operations including +, -, /, trigonometric functions such as sine and cosine, integrals, derivatives, absolute values, cross products, and dot products. Scalar or implied multiplication is not counted. For example, in $xy = 2$, the multiplication of x and y does not contribute to the operator count. Division within fractions that are already simplified does not count, either. These numbers are already in an accepted form and require no further operation.

Equal signs: Relations in mathematical expressions, i.e. inequalities and equal signs.

Numbers: Constants (2, 3), fractions (¾), simple numbers with units (1.2 m, 5000 V), and numbers in scientific notation (4×10^{-12} F). These have been subdivided into correct numbers and incorrect numbers only in the extra presentation analysis. For all other purposes, the number count is simply a measure of how many numbers appear on the open response section of the test.

- **Correct numbers:** Any numbers that have been assigned correct units where required. This count also includes natural numbers, constants, and any other dimensionless numbers. All the examples listed above fall into this category.
- **Incorrect numbers:** Numbers for which units are required and have not been provided by the student. This count does not reflect the correctness of the problem. For example, in reporting an answer to a problem asking for magnetic field, if the student reports 3.6 instead of 3.6 T, then that is an incorrect number.

Graphics Variables

Graphed objects: Lines and curves, such as circles, parabolas, or exponential functions. This category also contains any special composite objects drawn on the test, such as solenoids or batteries.

Graphed symbols: All presentation included on a graph that would normally fall under mathematics, such as symbols, numbers, or other symbolic or numeric graph labels.

Graphed words: The number of words written on a graph.

Language Variables

Words: English letters listed together in a recognizable pattern such that this string is distinguishable from a chain of variables. Spelling errors do not affect this count as long as the first criterion is met. For example, "Gauss' Law" counts as two words, while "abcde" is a string of five variables.

Sentences: Complex word strings. They must include a verb and/or some form of punctuation, although strict rules of grammar, like strict rules of spelling, may not be observed. Simple sentences such as "Apply Gauss' Law" are common in this category.

Using these definitions, all writing on University Physics II tests were counted for the Fall 2006 and Fall 2007 semesters. This yielded N = 222 students each taking four exams with 24 total open-response questions. This researcher and a few helper students performed the measurements. Comparison of test observations for pairs of observers yielded very good agreement between observers.

In addition to the tests, students' progress is also measured through the use of the Conceptual Survey of Electricity and Magnetism (CSEM) completed as the first and last laboratory quizzes. According to its designers, **"**The CSEM was developed as a pretest and posttest to assess students' initial knowledge in electricity and magnetism as well as the effect of various forms of instruction on changing that knowledge base" (Maloney

S12). The test includes questions on a range of electricity and magnetism topics designed to probe students' previously existing conceptions as well as the formal knowledge gathered from the course.

From the pre- and post-tests, the normalized gain, g, or Hake gain is used to characterize conceptual improvement. <g> is defined as the ratio of the actual average gain (%<post> - %<pre>) to the maximum possible average gain (100 - %<pre>) (Hake 64). This measures a student's improvement as the actual improvement over total possible improvement creating a scaled value for overall gain.

Correlating the gathered test presentation data with both the test average and the normalized gain yields information about the importance of how a student presents physics to their exam performance and conceptual understanding. These two separate correlations will not necessarily indicate the same factors as responsible for success in both modes of assessment. Determining which parts of student presentation relate to a particular achievement and why will be important in choosing areas of improvement in teaching methods and course goals.

Chapter 3: Variable Definitions

From these observed data, additional variables were formed. It would be no surprise if a student who wrote more on an exam performed better. To remove an expected correlation with quantity of presentation, a set of variables that are scaled by the total amount of presentation were formed (scaled variables). New variables were also introduced to explore interactions between important variables and to capture educationally interesting features (special variables).

The observations were formed into a hierarchy and totals of each branch formed for each student.

- **Total Elements –**
  - **Total Mathematics –**
    - **Symbols**
    - **Operators**
    - **Equal Signs**
    - **Numbers**
      - **Correct Numbers**
      - **Incorrect Numbers**
  - **Total Graphs – (all work drawn or written on drawings)**
    - **Graphed Objects**
    - **Graphed Symbols**
    - **Graphed Words**
  - **Total Language –**
    - **Words**
    - **Sentences**

Each variable in the second and third tiers is divided by total elements on the top tier to create a scaled ratio. Each third tier total is similarly analyzed as a ratio to its second tier super-category. In this manner, all ratios were formed to scale out effects of the total quantity of writing. The complete set of scaled variables is defined as follows (note that this detailed variable description continues through page 20),

Simple Ratios

English

Language Ratio (Words+Sentences/Total Elements) divides the total language elements, all words and sentences, by the total elements. This variable measures the relative amount of English in the student's writing.

Words Ratio (Words/Total Elements) divides the total number of words by the total number of elements.

Sentences Ratio (Sentences/Total Elements) divides the total number of sentences by the total elements.

Mathematics

Math Ratio (Symbols + Operators + Numbers + Equal Signs/Total Elements) divides the total number of math elements; operators, symbols, equal signs, and numbers; by the total elements. This variable measures the relative amount of mathematics presentation in the total work.

Symbols Ratio (Symbols/Total Elements) divides the total symbols by the total elements.

Operators Ratio (Operators/Total Elements) divides the total operators by the total elements.

Equal Signs Ratio (Equal Signs/Total Elements) divides the total number of equal signs by the total elements.

Numbers Ratio (Numbers/Total Elements) divides the total numbers by the total elements.

Graphics

Graph Ratio (Graphed Words + Graphed Symbols + Graphed Objects/Total Elements) divides the total graph elements (graphed words, symbols, and objects) by the total elements to measure the relative frequency of graphed presentation in the total work. A certain amount of graphed presentation is required in the examined course. However, the extent to which graphs must be explained is generally unspecified.

Graphed Objects Ratio (Graphed Objects/Total Elements) divides the total graphed objects by the total elements.

Graphed Symbols Ratio (Graphed Symbols/Total Elements) divides the total graphed symbols by total elements.

Graphed Words Ratio (Graphed Words/Total Elements) divides the total graphed words by the total elements.

The next set divides each base variable by its category, leaving a ratio that indicates the relative importance of each variable within its math, graph, or language category.

Language

Words Language Ratio (**WLR, Words/Total Language**) divides the total words by the total language. This will be larger if there are longer sentences. Extended sentences are one measure of linguistic complexity.

Sentences Language Ratio (**SLR, Sentences/Total Language**) divides the number of sentences by the total number of language elements.

Mathematics

Numbers Math Ratio (**NMR, Numbers/Total Math**) divides numbers by the total math elements.

Symbols Math Ratio (**SMR, Symbols/Total Math**) divides symbols by the total math elements.

Equal Signs Math Ratio (**EMR, Equal Signs/Total Math**) divides the total number of equal signs by the total number of mathematics elements.

Operators Math Ratio (**OMR, Operators/Total Math**) divides operators by the total math.

Graphical

Graphed Objects Graph Ratio (**OGR, Graphed Objects/Total Graph**) divides the total graphed objects by the total graph elements.

Graphed Symbols Graph Ratio (**SGR, Graphed Symbols/Total Graph**) divides the graphed symbols by the total graph elements.

Graphed Words Graph Ratio (**WGR, Graphed Words/Total Graph**) divides the total number of graphed words by the total graphed elements.

Beyond these scaled variables, special variables were created to explore interesting relationships. These special variables are defined as

Special Variables

Numeric Expression divides numbers by symbols, which shows how much symbolic work a student does before simply plugging in values. More symbolic work is specifically encouraged in the observed class.

Graphical Description divides the total number of graphed words and graphed symbols by the total graph elements. This quantifies how well described a graph is rather than being drawn with no explanation about what it means.

Language Complexity takes a ratio of words to sentences to quantify the complexity of language expression. This is related to traditional measures of reading level.

Math Complexity divides the sum of all the symbols, operators, and numbers by the equal signs. This variable is an analogue to Language Complexity in that it essentially measures the complexity of complete math sentences.

Abstract Mathematics Use divides the sum of all symbols, equal signs, and operators by the total numbers. This extends the Numeric Expression by determining the relative importance of all non-numeric mathematical expression.

Abstract Language Use divides the sum of the total language, symbols, equal signs, and operators by the total elements. This is a ratio of how much work in both mathematic and English sentences without any numeric evaluation is done compared to the total number of presentation elements on an exam.

Graph Language Balance creates a ratio of graphed words to the sum of the graphed objects and graphed symbols. This variable describes what portion of the graph's description is in English rather than mathematics.

Sentence Impact divides sentences and equal signs by the total elements to determine how important the presentation in complete sentences (both in mathematics and in English) is to overall presentation.

The selection of these special variables occurred systematically based on the total and scaled test presentation data as well as through the employment of certain intuitions and practical knowledge of which components might have specific bearing on student learning.

As previously mentioned in the definitions of the base variables, these scaled and special variables were further supplemented by a division of numbers into correct and incorrect categories. Variables involving the correct/incorrect distinction are termed "extra" and treated apart from the other variables. Extra scaled and special variables include the following:

Correct Numbers Ratio divides correct numbers by the total elements.

Incorrect Numbers Ratio divides incorrect numbers by the total elements.

Incorrect Numbers Math Ratio (INMR) divides the total incorrect numbers by the total mathematics elements.

Correct Numbers Math Ratio (CNMR) divides the correct numbers by the total mathematics elements.

Number Quality creates a ratio of correct numbers to the total numbers. This value for the relative importance of correct numeric expression answers the question, "When numbers are used, how vital is it that units be included?"

Correlations involving the two types of numbers were analyzed separately from the rest of the variables because the numbers variables are the only ones that make any assumptions about the correctness of the work on an exam.

Chapter 4: Statistical Analysis

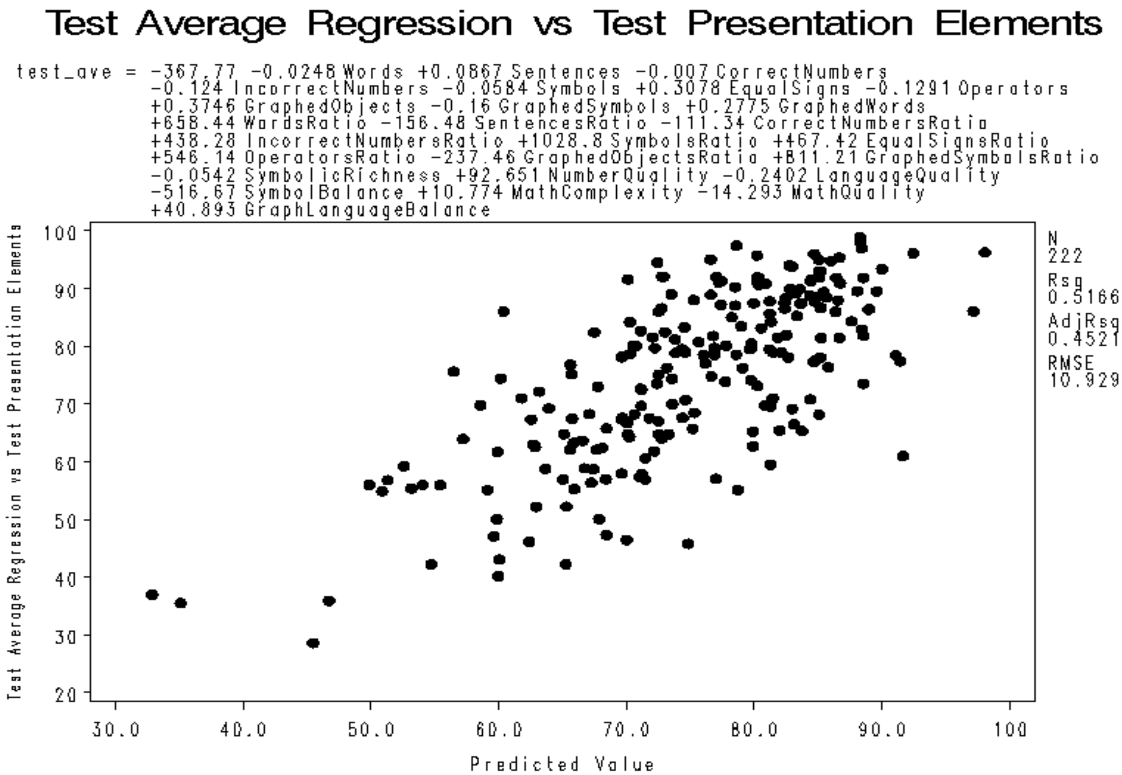
Two statistical methods will be used to evaluate the relations between the observed presentation variables and student performance measured by test average and normalized gain. According to Samprit Chatterjee, "The task of regression analysis is to learn as much as possible about the environment represented by the data" [1]. To this end, the computer program SAS will be used to perform a linear regression analysis and correlation analysis. Regression analysis is a statistical technique that relates one dependent variable to other variables [6]. In this study, regression models will be calculated for test average and normalized gain.

The regression model in this study is linear. That is, the dependent variable will be related to the independent variables by an equation containing only x to the first power and constants. For example, the expression $y = ax_1 + bx_2 + c$ is a linear model with constants a, b, and c, independent variables x_1 and x_2, and dependent variable y.

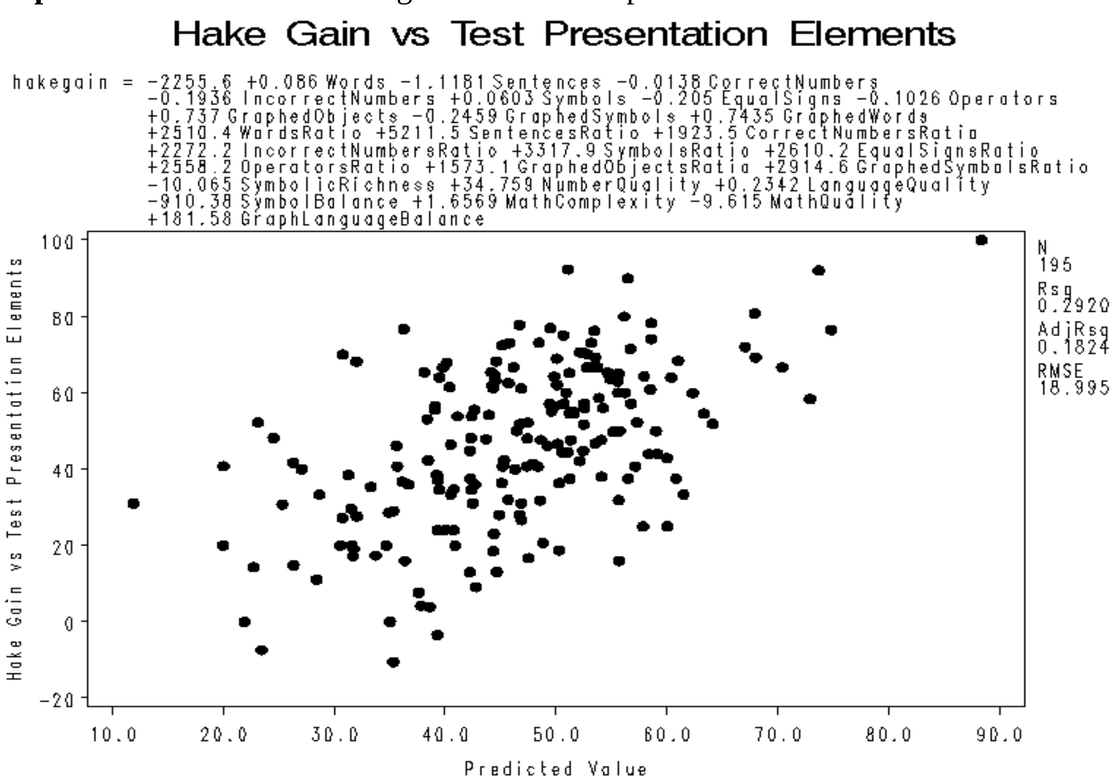
R-squared (R^2) is used to describe the predictive power of regressions. R^2 is the ratio of the variation explained by the assumption of the model to the variation in the initial data. This value ranges from zero to one. Zero signifies no linear relation. A plot with zero R^2 would be a spherical collection of data points. One indicates a perfect linear relation with all data points on a line. See the figures below (Graph 1 and Graph 2) for examples of regressions with R^2 values of 0.30 and 0.50.

These graphs are regressions of test average and normalized gain compared with all the tested variables from both semesters. They provide an idea of what different regressions look like, but more specifically are indicative of the data which composes this study.

Graph 1. Model of test average vs. all tested elements with $R^2 = 0.5166$.



Graph 2. Model of normalized gain vs. all tested presentation with $R^2 = 0.2920$.



In general, regression models become better as more variables are added. The adjusted R^2 (R^2_{adj}) eliminates this effect by correcting for the loss of degrees of freedom due to the addition of more independent variables. It also helps to prevent over-fitting the model, or adding terms that are unnecessary [9]. In other words, R^2_{adj} shifts according to actual changes in the variance of the fit, not simply because the model contains more variables. The maximum R^2_{adj} will be used to select the regression model that best describes the data.

Correlations are also calculated for the two dependent variables, test average and normalized conceptual gain, versus independent variable using the SAS program. The measure of correlative strength, or goodness of linear fit, is called the correlation coefficient and signified by r. This value ranges from negative one to one and measures the strength of the linear relationship between y and x.

The statistical analysis also reports the probability that the calculated correlation is due to random factors. This value, reported as P, indicates the probability that the correlations resulted from chance. A correlation will be considered statistically significant if $P < 0.05$, that is if there is a less than 5% chance the correlation arose randomly. In other words, the probability values are the odds that a correlation is a result of chance and that the model does not actually describe the relationship of the variables [2]. Probability values range from zero to one. High probabilities indicate that the correlation is possibly due to chance; low probability values mean that the likelihood of the relationship being a random occurrence is small.

Chapter 5: Descriptive Statistics

Table 7 presents averages, minimum, and maximum values for the collected test presentation data. Comparing these values in Fall 2006 and Fall 2007 indicates major differences in the two data sets. These variations between semesters are probably due to the tests themselves. Each semester presents 12 free response test questions. A greater emphasis on conceptual exam questions, for example, might significantly alter the amount of mathematics and language presentation. Multi-semester test average and normalized gain data show that the student population is fairly stable. Regardless of the cause, the large variation between semesters suggests that the semesters should be analyzed separately.

Mean Values for Collected Test Presentation Data								
	Fall 2006				**Fall 2007**			
Variable	**N**	**Mean**	**Min**	**Max**	**N**	**Mean**	**Min**	**Max**
Words	87	163.92	21	753	135	208.47	52	1213
Sentences	87	11.47	1	54	135	14.52	3	58
Numbers	87	324.18	168	598	135	234.68	67	360
Symbols	87	453.20	179	853	135	303.95	101	509
Equal Signs	87	196.11	101	338	135	138.07	50	250
Operators	87	248.20	97	470	135	200.56	68	342
Graphed Objects	87	107.78	61	182	135	90.01	46	146
Graphed Symbols	87	72.67	18	174	135	118.61	60	720
Graphed Words	87	7.40	0	43	135	10.05	0	90
Test Average	87	70.23	28.62	98.88	135	77.89	42.25	97.94
Hake Gain	87	43.42	-7.41	92.31	121	48.30	-10.53	100.00

Table 7. This table reports the mean values as well as the minimum and maximum number of variables counted from each semester. N is the number of observations.

Table 7 has several interesting features. While Fall 2006 has much more mathematics presentation, its test average and Hake gain are both lower than Fall 2007. Fall 2007 on the other hand has more language use and more graphed words and symbols for roughly the same number of graphed objects.

Table 7 shows the students' written work in the class and undoubtedly contains the effect that a student who shows little work cannot complete the problems. To investigate the intrinsic effect of the pattern of student presentation on performance, the counts were scaled by the total number of elements, as shown in Table 8.

Mean Values Scaled Test Presentation Data				
	Fall 2006		**Fall 2007**	
Variable	**N**	**Mean**	**N**	**Mean**
Ratios of Base Variables to Total Elements				
Total Language/Total Elements	87	0.10348	135	0.15908
Words/Total Elements	87	0.09674	135	0.14877
Sentences/Total Elements	87	0.00674	135	0.01031
Total Math/Total Elements	87	0.77664	135	0.67252
Symbols/Total Elements	87	0.28597	135	0.23172
Numbers/Total Elements	87	0.20849	135	0.18139
Equal Signs/Total Elements	87	0.12549	135	0.10644
Operators/Total Elements	87	0.15670	135	0.15298
Total Graphed Elements/Total Elements	87	0.11988	135	0.16840
Graphed Words/Total Elements	87	0.00438	135	0.00699
Graphed Symbols/Total Elements	87	0.04473	135	0.09098
Graphed Objects/Total Elements	87	0.07077	135	0.07043
Ratios of Base Variables to Total Math, Language, or Graph				
Words/Total Language Elements	87	0.93696	135	0.93633
Sentences/Total Language Elements	87	0.06304	135	0.06367
Numbers/Total Math Elements	87	0.26837	135	0.26961
Symbols/Total Math Elements	87	0.36855	135	0.34481
Equal Signs/Total Math Elements	87	0.16155	135	0.15833
Operators/Total Math Elements	87	0.20152	135	0.22725
Graphed Objects/Total Graph Elements	87	0.59893	135	0.42266
Graphed Symbols/Total Graph Elements	87	0.36550	135	0.53661
Graphed Words/Total Graph Elements	87	0.03557	135	0.04073

Table 8. Like Table 7, this table presents mean values. These are for the counted variables divided by the total elements and their total category elements.

Much of the variation between the semesters vanishes with the scaled variables. The scaled and special variable averages indicate that the problems encountered in merging the test totals are not fully present once the totals are scaled out. Although there is some variation between the two semesters, these differences are not as large. The language variables again have a particularly large variation compared to the other variables. This is due to the selection of exam questions between the semesters.

As one would expect, the majority of the presentation went into mathematics (Total Math/Total Elements or Math Ratio). In fact, more presentation went into graphing than writing in English, which illustrates the resistance, or perhaps inability, of beginning engineers and scientists to explain their work in English.

The special variables created to examine more complicated relationships in the data feature a similar general agreement between semesters. Even those special variables involving language remain comparable from semester to semester.

Mean Values for Special Test Presentation Data				
	Fall 2006		**Fall 2007**	
Variable	**N**	**Mean**	**N**	**Mean**
Numeric Expression	87	0.73704	135	0.79886
Graphical Description	87	0.40107	135	0.57734
Language Complexity	87	17.20797	135	15.82353
Math Complexity	87	5.25501	135	5.40454
Abstract Mathematics Use	87	2.76918	135	2.74780
Abstract Language Use	87	0.67163	135	0.65021
Graph Language Balance	87	0.03860	135	0.04487
Sentence Impact	87	0.13223	135	0.11675
Test Average	87	70.22943	135	77.88852
Hake Gain	74	43.41703	121	48.29537

Table 9. Presents averages for the combined variables. Special variable definitions: Numeric expression = Numbers/Symbols, Graphical Description = (Graphed Words + Graphed Symbols)/Total Graphed Elements, Language Complexity = Words/Sentences, Math Complexity = (Symbols + Numbers + Operators)/Equal Signs, Abstract Mathematics Use = (Symbols + Equal Signs + Operators)/Total Numbers, Abstract Language Use = (Total Language + Symbols + Equal Signs + Operators)/Total Elements.

Examination of Table 9 shows that students are using symbols more than numbers, behavior that is encouraged in the examined course. The large sentence length (Language Complexity, or Words/Sentences) is inflated by the large number of words that are not in sentences.

Chapter 6: Correlation Analysis

Calculating the correlations between the two dependent variables, test average and normalized gain, and the collection of independent variables provides a measure of the strength of their relationship. The relationship of the data is measured by r, the correlation coefficient, ranging from negative one to one. This value tells how well the linear model y (dependent variable) = a (some constant) x (independent variable) + b (intercept) describes the collection of data.

Based on the mean values presented in the data collection section, although working with a larger pool of students is desirable, simply merging the two data sets for the unscaled variables does not accurately retain the important correlations in each individual set. Therefore the data from Table 7 was analyzed separately for the two semesters. From Tables 8 and 9, it was determined that such issues do not arise with scaled variables. So the two semesters were combined and compared with the individual semesters. This pooled data set contains both the full set of 24 exam questions and the complete sampling of students from two semesters (222 students). Correlations from the individual semesters examined separately and then combined may indicate continuous important factors in the prediction of test average and normalized gain.

Total Test Presentation Data Correlations: Test Average					
	Fall 2006		**Fall 2007**		**Average**
Variables	**r**	**P**	**r**	**P**	**r**
Total Math	0.600	<.0001	0.595	<.0001	0.59739
Total Elements	0.603	<.0001	0.568	<.0001	0.58566
Operators	0.510	<.0001	0.632	<.0001	0.57104
Symbols	0.616	<.0001	0.486	<.0001	0.55078
Equal Signs	0.601	<.0001	0.476	<.0001	0.53885
Numbers	0.521	<.0001	0.554	<.0001	0.53763
Total Language	0.456	<.0001	0.374	<.0001	0.41512
Words	0.459	<.0001	0.369	<.0001	0.41376
Sentences	0.387	0.0002	0.409	<.0001	0.39833
Total Graphed Elements	0.167	0.1227	0.258	0.0025	0.21236
Graphed Words	0.220	0.0406	0.183	0.0341	0.20127
Graphed Symbols	0.193	0.0740	0.193	0.0248	0.19284
Graphed Objects	0.074	0.4961	0.298	0.0005	0.18587

Table 10. This table reports the correlations of the total presentation elements counted in each category with the test average. These are sorted from most to least correlated. Highlighted rows are correlations that are significant for both semesters.

The total number of presentation elements correlates strongly with test average (Table 10). Increased presentation output from a student generally indicates that the student can do more, so such a correlation was anticipated. The relative strength of this correlation can be measured against correlations of traditional markers of student success, such as homework average and the total number of missing assignments.

For Fall 2006, the test average compared with the homework average yields a highly significant correlation with a correlation coefficient of 0.548. Fall 2007 does even better with 0.681. Homework average includes information about class performance, so it naturally has a strong relationship to the student's overall success in the class. The total number of missing assignments includes only data about student behavior without any components that affect the grade. For Fall 2006, test average has a relatively weak -0.276 correlation with the total missing assignments. Fall 2007 correlates similarly at -0.348. Both are significant but relatively weak correlations. Compared to these values, the total elements correlation at 0.58 is about the same as homework average and much stronger than the total missing number of assignments. Because test presentation data does not take into account any graded components, it is more comparable to total missing. The correlation of 0.6 for total elements is particularly impressive when compared with the missing assignments. Although this total correlation does not contain specific markers of student success, its strength indicates the amount of student presentation is of equal importance to more traditional markers of student performance.

The total mathematics presentation elements have an even stronger correlation with test average, 0.59, than the total presentation elements. In physics courses in particular, the importance of mathematics as the primary method of expression merits specific attention to its role in overall performance. The total numbers, symbols, operators, and equal signs all exhibit similarly strong correlations, roughly 0.55 each, which indicates that individual mathematics components play a large role in student performance. Each part of math presentation is important in its own right, not just as an additive part of the total work in math.

The total language variable and its constituent words and sentences all have significant, relatively strong correlations. Though the presentation in English does not predict test average as well as the mathematics variables, its predictive power indicates an important relationship between average test performance and written English work.

With easily the lowest predictive power and least significance of the total test presentation, the graphed portion of the test presentation data seems to have very limited bearing on test performance. The total graphed elements have a much weaker correlation than either the total math or total language elements. The graphed objects variable, a measure of what one would generally think of as the main content of the graph itself, has not only a very weak correlation, but also a high probability that this is a product of chance. This is somewhat surprising since graphical presentation is important in scientific communication.

Total Test Presentation Data Correlations: Normalized Gain					
	Fall 2006		**Fall 2007**		**Average**
Variables	**r**	**P**	**r**	**P**	**R**
Sentences	0.245	0.0355	0.393	<.0001	0.31882
Total Elements	0.312	0.0068	0.316	0.0004	0.31379
Operators	0.311	0.0070	0.295	0.0010	0.30311
Total Language	0.230	0.0490	0.372	<.0001	0.30101
Words	0.227	0.0515	0.368	<.0001	0.29763
Total Math	0.320	0.0054	0.251	0.0055	0.28560
Symbols	0.328	0.0044	0.206	0.0233	0.26698
Numbers	0.268	0.0210	0.210	0.0206	0.23912
Equal Signs	0.286	0.0135	0.191	0.0356	0.23857
Graphed Objects	0.019	0.8723	0.136	0.1365	0.07758
Graphed Words	0.101	0.3909	0.006	0.9450	0.05378
Total Graphed Elements	0.055	0.6445	0.0002	0.9986	0.02734
Graphed Symbols	0.062	0.6025	-0.045	0.6250	0.00833

Table 11. Correlations of the total presentation elements counted in each category with the normalized gain. The significant correlations are highlighted.

The correlations between test presentation data and the normalized gain express a far weaker connection than the test average correlations (Table 11). This is expected somewhat because the test presentation data is measured from the same exam as the results it is predicting, whereas the normalized gain is drawn from a different instrument. Even so, comparing these results with the traditionally predictive homework average and total number of missing assignments shows that an average correlation coefficient of 0.31 for total elements is comparable to traditional measures. Homework average versus normalized gain yields correlation values of 0.373 and 0.322 for Fall 2006 and Fall 2007, respectively. The total number of missing assignments versus the normalized gain yields weaker, non-significant results.

Referring to Table 11, among the subcategories English language elements predict normalized gain particularly well. Total language, words, and sentences all have significant, relatively strong correlations around 0.30. Of the individual variables, sentences ranks highest, indicating that structured, complete work in language corresponds to enhanced conceptual mastery.

Somewhat less than total language, total math and its constituent components also exhibit considerable comparable predictive power to the homework average. Graphed elements, however, again have a very weak correlation and similarly low significance. While they do not necessarily detract from conceptual gain, graphed work provides no positive impacts.

Scaled Test Presentation Data Correlations: Test Average						
Ratios of Base Variables to Total Elements						
Variables	**2006, r**	**P**	**2007, r**	**P**	**All, r**	**P**
Language Ratio	0.329	0.0019	0.286	0.0008	0.36653	<.0001
Words Ratio	0.333	0.0016	0.278	0.0011	0.36464	<.0001
Sentences Ratio	0.241	0.0246	0.326	0.0001	0.33967	<.0001
Graphed Words Ratio	0.092	0.3986	0.114	0.1861	0.13954	0.0378
Operators Ratio	-0.019	0.8646	0.184	0.0327	0.06453	0.3399
Graphed Symbols Ratio	-0.145	0.1802	-0.205	0.0173	0.04888	0.4687
Graph Ratio	-0.410	<.0001	-0.306	0.0003	-0.13983	0.0373
Symbols Ratio	0.181	0.0938	-0.080	0.3557	-0.14325	0.0329
Math Ratio	-0.109	0.3128	-0.101	0.2440	-0.23725	0.0004
Equal Signs Ratio	-0.106	0.3284	-0.225	0.0088	-0.26798	<.0001
Numbers Ratio	-0.340	0.0013	-0.173	0.0446	-0.33190	<.0001
Graphed Objects Ratio	-0.605	<.0001	-0.452	<.0001	-0.51550	<.0001

Table 12. Lists the correlations of the counted test presentation variables divided by the total elements. Highlighted rows maintain their significance from semester to semester.

Scaled variables remove the effects of the total counts from the correlation, in effect normalizing all the elements. After this re-weighting, language variables maintain their positive, strong, significant correlations (Table 12). Language ratio, words ratio, and sentences ratio all have correlation coefficients of roughly 0.35. This correlation is stronger than that of all the other ratio variables, but is much weaker than the correlation of homework average (0.546).

Besides the strong positive correlations in the language variables, important negative correlations are also highly descriptive. The ratio of graphed objects to total elements has easily the most negative correlation, -0.51, and is highly statistically significant. Combined with all the graphed elements' weak performance in the total correlations, this negative correlation indicates that graphs without supporting description are not a positive component for learning. This could be a result of the class policy that encourages graphical work, but is currently unsuccessful in producing the well-described graphical work that is important to learning.

Within the math category, the numbers ratio has a significant, negative correlation of 0.33. Although these correlations are not as negative as the graphed objects ratio discussed above, detrimental effects from increased math presentation seem unusual for a physics course. In the case of the numbers, this correlation shows that mathematics done largely with numbers rather than symbols is detrimental to student learning.

Scaled Test Presentation Data Correlations: Test Average						
Ratios of Base Variables to Total Math, Language, or Graph						
	Fall 2006		**Fall 2007**		**All**	
Variables	**r**	**P**	**r**	**P**	**R**	**P**
Operators/Total Math	0.054	0.6220	0.372	<.0001	0.31326	<.0001
Graphed Symbols/Total Graph	0.225	0.0362	0.052	0.5482	0.29106	<.0001
Graphed Words/Total Graph	0.214	0.0468	0.165	0.0552	0.19033	0.0044
Sentences/Total Language	0.028	0.7968	0.212	0.0138	0.11169	0.0969
Symbols/Total Math	0.352	0.0008	-0.001	0.9894	0.03901	0.5631
Words/Total Language	-0.028	0.7968	-0.212	0.0138	-0.11169	0.0969
Equal Signs/Total Math	-0.065	0.5519	-0.215	0.0121	-0.16336	0.0148
Numbers/Total Math	-0.347	0.0010	-0.154	0.0742	-0.23029	0.0005
Graphed Objects/Total Graph	-0.299	0.0049	-0.187	0.0303	-0.35104	<.0001

Table 13. Displays the correlations of the ratios of each element to the total elements in its particular category. All math elements are divided by total math elements, all language elements by total language, and all graph elements by total graph.

The next tier of scaled variables contains the base variables balanced against their subcategories. That is, each math variable is weighted against the total math elements, each graph variable is weighted against the total graph elements, and each language variable is weighted against the total language elements. These ratios indicate the relative importance of the base variables within each subset rather than their strength in the presentation as a whole (Table 13).

Although several of the variables have significant, relatively strong correlations in the total set, none except the ratio of graphed objects to total graphed elements maintains its negative correlation and significance through both individual semesters. Using more graphed objects, without similarly high graphed words and symbols, again comes through as a considerable detriment to overall test performance. This was observed in Table 7, so it would seem that the ratios of the variables to their subcategories do not contribute additional insight into student performance.

Special Test Presentation Data Correlations: Test Average						
	Fall 2006		**Fall 2007**		**All**	
Variables	**r**	**P**	**r**	**P**	**r**	**P**
Graphical Description	0.299	0.0049	0.187	0.0303	0.35104	<.0001
Abstract Language Use	0.535	<.0001	0.352	<.0001	0.34341	<.0001
Abstract Mathematics Use	0.318	0.0027	0.139	0.1085	0.21074	0.0016
Graph Language Balance	0.209	0.0516	0.160	0.0630	0.18531	0.0056
Math Complexity	0.047	0.6662	0.213	0.0132	0.15629	0.0198
Sentence Impact	-0.034	0.7548	-0.136	0.1155	-.019668	0.0033
Numeric Expression	-0.383	0.0002	-0.081	0.3477	-0.13739	0.0408
Language Complexity	-0.109	0.3153	-0.211	0.0140	-0.17322	0.0097

Table 14. Special variables do not contain total elements so they are in a sense scaled, but they are less natural than the scaled variables. They must indicate interesting features very strongly to be considered relevant to the discussion.

The final set of variables investigated are the "special" variables; variables constructed to look for specific educational effects. The special test presentation variables as a whole have substantially less significance to the prediction of test average than the scaled variables (Table 14). In general, a P value must be less than 0.05 to be considered significant. Only Abstract Language Use maintains both a high significance and strong correlation through both semesters and the combined set. Abstract Language Use describes the relative amount of abstract description. It measures the ratio of all non-graphical, non-numeric variables to the total elements (i.e. (Total Language + Symbols + Operators + Equal Signs)/Total Elements). Strength in this area reaffirms the previous conclusions about the detrimental effects of a test heavy in numbers and unexplained graphs. It also points to the use of complex mathematical and linguistic description as the most important measures of overall success.

Graphical description also has a highly significant, strong correlation of 0.35. The strong correlation of words, symbols, and other descriptors on a graph complements the negative correlation seen for graphed objects. An increase in the use of graphs is only beneficial to test average if the graphs have ample explanation.

The same analysis was performed for the normalized gain (Table 15). Although the normalized gain continues to have generally weaker and fewer significant correlations with test presentation data than test average, the correlations that do exist are maintained in the scaled variables. The English language variables' values from the first three rows of Table 15 are comparable to homework average (0.373 for 2006 and 0.314 for 2007). At a highly significant positive 0.30 each, no other scaled variables come close to matching their predictive power. Also unusual about these variables is their low significance in Fall 2006. For Fall 2007 and the combined set, which includes Fall 2006, the language variables are strongest and the most significant.

Scaled Test Presentation Data Correlations: Normalized Gain						
Variables Divided by Total Elements	**Fall 2006**		**Fall 2007**		**All**	
	r	**P**	**r**	**P**	**r**	**P**
Language Ratio	0.189	0.1059	0.349	<.0001	0.31122	<.0001
Sentences Ratio	0.204	0.0811	0.354	<.0001	0.31936	<.0001
Words Ratio	0.186	0.1118	0.343	0.0001	0.30676	<.0001
Math Ratio	-0.068	0.5630	-0.168	0.0649	-0.17482	0.0145
Symbols Ratio	0.054	0.6467	-0.158	0.0830	-0.14161	0.0483
Operators Ratio	0.105	0.3717	0.047	0.6123	0.05796	0.4209
Numbers Ratio	-0.199	0.0889	-0.161	0.0786	-0.20595	0.0039
Equal Signs Ratio	-0.135	0.2514	-0.193	0.0340	-0.20401	0.0042
Graph Ratio	-0.232	0.0471	-0.293	0.0011	-0.16269	0.0231
Graphed Objects Ratio	-0.337	0.0033	-0.252	0.0054	-0.28365	<.0001
Graphed Symbols Ratio	-0.010	0.3973	-0.243	0.0073	-0.07103	0.3238
Graphed Words Ratio	0.042	0.7206	-0.076	0.4049	-0.02047	0.7764
Variables Divided by Total from Subcategory	**Fall 2006**		**Fall 2007**		**All**	
	R	**P**	**r**	**P**	**r**	**P**
Sentences/Total Language	0.028	0.5014	0.167	0.0668	0.12458	0.0827
Words/Total Language	-0.028	0.5014	-0.167	0.0668	-0.12458	0.0827
Symbols/Total Math	0.352	0.2315	-0.024	0.7961	-0.01554	0.8293
Operators/Total Math	0.054	0.1015	0.220	0.0153	0.23733	0.0008
Equal Signs/Total Math	-0.065	0.2343	-0.101	0.2716	-0.12148	0.0907
Numbers/Total Math	-0.347	0.0889	-0.082	0.3725	-0.12063	0.0930
Graphed Objects/Total Graph	-0.299	0.1735	0.0003	0.9974	-0.13557	0.0588
Graphed Words/Total Graph	0.214	0.0468	-0.002	0.9797	0.05007	0.4870
Graph Symbols/Total Graph	0.225	0.0362	0.002	0.9850	0.12343	0.0856

Table 15. Counted variables scaled against the total elements and the math, graph, and language total elements. Ratios by category, rather than overall total, are named with initials. Highlighted variables maintain their significance for all three sets.

Though its correlation is roughly half what it was in test average, the graphed objects ratio still has a relatively strong, highly significant, negative correlation. In predicting a student's normalized gain as well as test average performance, the trend continues to be that graphs in themselves are not a key to learning. Once again, drawing a graph without enhancing it with description is not an effective intellectual strategy. None of the other graph or mathematics variables are significant enough to add to this assessment, but the overall results weakly mirror those from test average. This would indicate a consistency in the conclusion that more use of English language increases performance while added graphical inferior analysis decreases it.

Special Test Presentation Data Correlations: Normalized Gain						
	Fall 2006		**Fall 2007**		**All**	
Variables	**r**	**P**	**r**	**P**	**R**	**P**
Abstract Language Use	0.309	0.0075	0.333	0.0002	0.28374	<.0001
Graphical Description	0.160	0.1735	-0.0003	0.9974	0.13557	0.0588
Math Complexity	0.135	0.2518	0.095	0.2987	0.11778	0.1010
Abstract Mathematics Use	0.195	0.0954	0.077	0.3984	0.11733	0.1024
Graph Language Balance	0.132	0.2638	0.011	0.9047	0.05943	0.4092
Numeric Expression	-0.177	0.1307	-0.028	0.7581	-0.04787	0.5064
Sentence Impact	-0.084	0.4793	-0.085	0.3558	-0.12779	0.0750
Language Complexity	-0.139	0.2391	-0.155	0.0895	-0.15117	0.0349

Table 16. List of correlations between the special variables and the normalized gain. Abstract Language Use is highlighted because it is the only variable that is consistently significant.

The correlations of special test presentation variables to normalized gain display almost no important relationships (Table 16). The strongest and only significant variable is Abstract Language Use, a ratio of all language and math elements excluding numbers to the total elements. The Abstract Language Use variable was also important in predicting test average. Here it has similar strength and meaning. If a student works more in complex mathematical and linguistic forms, his conceptual gain will increase. The complement of that is again that using more graphs and numbers decreases performance.

Chapter 7: Regression Analysis

To determine how much of a student's performance in the class is predictable from the presentation of their tests, multiple linear regression is used. The amount of variation in the initial data explained by the linear model is measured by R^2. R^2_{adj} is calculated to scale out the effects of adding extra variables and determine the best, most efficient model for the data. In order to make useful comparisons with the traditional markers in the regression analysis, regressions of test average and normalized gain with homework average and the total number of missing assignments were created. These regressions produced the following results:

R^2 Values for Regressions of Traditional Education Markers: Test Average						
	Fall 2006		**Fall 2007**		**All**	
Variable	R^2	$R^2_{adj.}$	R^2	$R^2_{adj.}$	R^2	$R^2_{adj.}$
Homework Average	0.3004	0.2922	0.4633	0.4593	0.2977	0.2945
Total Missing Assignments	0.0761	0.0652	0.1209	0.1143	0.1120	0.1079

Table 17. This reports the R^2 values for regressions between the test average and homework average and total missing.

Omitted from this table are the regression coefficients for the normalized gain's relationship with homework average and the total number of missing assignments. These values were either very weakly related or they were not reported because their P values are non-significant. All the other reported R^2 have very significant probability values, though the total missing assignments values are quite low. These values can be compared with the regression values in the following tables to assess the relative value of the calculated models.

Table 18 presents the R^2 and R^2_{adj} values for linear models that predict the test average and normalized gain from the variables in Tables 12, 13, and 15, called scaled variables.

Regression Values and Best Models for Scaled Test Presentation Data					
Test Average	**N**	**R^2**	**$R^2_{adj.}$**	**P**	**Best Model**
Fall 2006	87	0.5207	0.4783	<.0001	Sentence Ratio (-), Numbers Ratio (-), Equal Signs Ratio, Graphed Objects Ratio (-), NMR, WLR (-), EMR (-)
Fall 2007	135	0.3289	0.3135	<.0001	Sentences Ratio, Graphed Objects Ratio (-), OMR
All	222	0.4269	0.3998	<.0001	Symbols Ratio, Equal Signs Ratio, Math Ratio (-), Language Ratio, NMR, OMR, OGR (-)
Normalized Gain	**N**	**R^2**	**$R^2_{adj.}$**	**P**	**Best Model**
Fall 2006	74	0.2554	0.1764	0.0052	Numbers Ratio (-), Equal Signs Ratio, Operators Ratio (-), Graphed Symbols Ratio (-), SMR (-), EMR (-), SGR
Fall 2007	121	0.2982	0.2548	<.0001	Sentences Ratio, Graphed Objects Ratio, Graph Ratio (-), Symbols Ratio (-), Graphed Symbols Ratio, OMR, OGR (-)
All	195	0.2477	0.2154	<.0001	Symbols Ratio, Graphed Objects Ratio, Language Ratio, SMR (-), OMR, OGR (-)

Table 18. This presents the best regression models relating all the scaled variables (variables divided by the total elements overall and in each respective category) to both the test average and the normalized gain. Variables listed as ratios are divided by the total elements, while those named with initials are scaled by total math, language, or graph elements: NMR = Numbers/Total Math, WLR = Words/Total Language, EMR = Equal Signs/Total Math, OMR = Operators/Total Math, OGR = Graphed Objects/Total Graphed Elements, SMR = Symbols/Total Math, SGR = Graphed Symbols/Total Graphed Elements, and SMR = Symbols/Total Math.

Table 18 shows the best regression models for the variables that have been scaled as a ratio to the total elements and to their respective math, graph, or language category. The best models for test average are all strongly significant with R^2 values in the same range as those for homework average. Placing these values in the context of the existing research reported in chapter one shows that these regressions compare favorably overall with what is currently available in physics education. The test average regression is slightly below Halloun and Hestenes' model that includes all pretests and prior courses

(0.50), but is comparable to the regression for pretests only (0.41) and is over twice as strong as that for courses taken (0.16) (values from Table 4) [4]. Additionally, the test average models are much better than those found by Hudson and Liberman (from Table 1, only 0.27 for formal reasoning and 0.12 for mathematics pretest) [5]. These models are also generally much better than those determined by the Stewart study for the role of a student's behavior in class performance (Table 5) [11]. Since writing behavior would naively be ranked below all these measures by most physics instructors, this is a very impressive result.

Similarly assessing the results of the normalized gain models in context, the regressions' predictive strength is much stronger than that of comparable regressions with homework average and total missing. The regressions predict well below Meltzer's reported R^2 values of about 0.40 [8], but are considerably better than the behavior data found by Stewart and McGee for the easier version of the course [11].

These results are remarkable in that the regression models for written test presentation, without any inclusion of performance information, are more predictive than previous courses taken and performance on pretests. Additionally, a student's written work tells more than his or her behavior toward the class, including time spent on homework and studying.

Unfortunately, the variables included in each model differ significantly in number and type. The best models in the normalized gain face similar issues of variation in the number and type of determining factors. This issue continues through all regression models and makes determining any constant predictive equation impossible. It is probable that this is due to the low number of test questions each semester. The regression that pools both semesters should overcome this problem because it contains 24 questions.

Regression Values and Best Models for Special Test Presentation Data					
Test Average	**N**	**R^2**	**$R^2_{adj.}$**	**P**	**Best Model**
Fall 2006	87	0.4503	0.4235	<.0001	Graphical Description, Abstract Mathematics Use (-), Abstract Language Use, Sentence Impact
Fall 2007	135	0.2938	0.2607	<.0001	Numeric Expression, Graphical Description, Math Complexity, Language Complexity (-), Abstract Mathematics Use, Abstract Language Use
All	222	0.3614	0.3466	<.0001	Numeric Expression, Graphical Description, Abstract Language Use
Normalized Gain	**N**	**R^2**	**$R^2_{adj.}$**	**P**	**Best Model**
Fall 2006	74	0.1592	0.1104	0.0164	Graphical Description, Math Complexity, Abstract Language Use, Sentence Impact
Fall 2007	121	0.2227	0.1959	<.0001	Numeric Expression, Abstract Language Use, Sentence Impact (-)
All	195	0.1968	0.1756	<.0001	Graphical Description, Numeric Expression, Abstract Language Use

Table 19. This presents the best regression models relating all the special variables (complex variables chosen to explore a specific educational effect) to both the test average and the normalized gain. Variables are defined as follows: Graphical Description = (Graphed Words + Graphed Symbols)/Total Graphed Elements, Math Complexity = (Symbols + Operators + Numbers)/Equal Signs, Abstract Language Use = (Total Language + Symbols + Equal Signs + Operators)/Total Elements, Sentence Impact = (Sentences + Equal Signs)/Total Elements, Numeric Expression = Numbers/Symbols, Language Complexity = Words/Sentences, Abstract Mathematics Use = (Symbols + Equal Signs + Operators)/Total Numbers.

Analyzing the special variables in the same manner as the scaled variables yields similar results. All the correlations with test average are strong (0.36 overall) and significant. They compare very well with previous research, with R^2 well above Hudson/Liberman (from Table 1, 0.12 for math pretest and 0.27 for reasoning) [5] and Stewart/McGee (from Table 5, roughly 0.20 average for behavior predictions) [11]. As before, the test average value is on par with Halloun and Hestenes pretest regressions and well above their prior courses taken relationship [4].

Models for normalized gain fail to maintain this continuity. At only 0.26 R^2 the normalized gain predictability is well below Meltzer (Table 3) [8] and Halloun and Hestenes pretest values (Table 4) [4]. The normalized gain for the special variables is comparable to the behavior data measure by Stewart and McGee (Table 5) [11].

Regression Values and Best Models for Scaled and Special Test Presentation Data					
Test Average	**N**	$\mathbf{R^2}$	$\mathbf{R^2_{adj.}}$	**P**	**Best Model**
Fall 2006	87	0.5499	0.4973	<.0001	Sentences Ratio (-), Numbers Ratio (-), NMR, EMR (-), Graphed Objects Ratio (-), Language Complexity (-), Sentence Impact
Fall 2007	135	0.3521	0.3164	<.0001	Sentences Ratio, Graphed Objects Ratio (-), OMR
All	222	0.4360	0.4093	<.0001	Words Ratio, Symbols Ratio, Equal Signs Ratio, OMR, Math Ratio (-), NMR, OGR (-), Language Complexity (-)
Normalized Gain	**N**	$\mathbf{R^2}$	$\mathbf{R^2_{adj.}}$	**P**	**Best Model**
Fall 2006	74	0.2651	0.1872	0.0066	Numbers Ratio (-), Equal Signs Ratio, Operators Ratio (-), Graphed Objects Ratio, EMR (-), Numeric Expression, Abstract Mathematics Use (-)
Fall 2007	121	0.3234	0.2751	<.0001	Graphed Objects Ratio, Graphed Words Ratio (-), OMR, OGR (-), WGR (-), Abstract Mathematics Use (-), Abstract Language Use, Graph Language Balance
All	195	0.2780	0.2311	<.0001	Words Ratio (-),Graphed Objects Ratio, Graphed Symbols Ratio, SMR (-), EMR (-), SGR, Operators Ratio (-), WGR (-), Abstract Language Use, Graph Language Balance

Table 20. This table combines tables 18 and 19 in an effort to determine if a total model can significantly improve the regression values or provide a complete model to predict students' success. The variables are defined as in the previous two tables.

Combining the results of the scaled and special variable regressions yields similar R^2 values to those seen in each separate data set. Their placement in the referenced physics education results is exactly the same as that of the scaled variables. The normalized gain has slightly weaker regression values, but features a more constant number of variables.

In an effort to normalize the best models such that they have the same number of factors while maintaining comparable regression coefficients, regressions were calculated using the scaled variables and only the special variables that maintained their significance (i.e. sentential [non-numeric and non-graphical] description and graphical description). This did not produce a standard model either. Even selecting only variables already proven to be important, the fluctuations in various factors from semester to semester cannot be eliminated. Furthermore, this model confirmed that the combined set of scaled and special variables does not significantly improve what is seen for the scaled variables alone. Gains in predictive strength are minor; the changes in variables and model are practically nonexistent from scaled to scaled and special.

Taking the scaled and special regressions individually and as a composite set in various combinations fails to yield a specific, overall equation for success in either test average or normalized gain. The subdivision of numbers into correct and incorrect did not add any beneficial models to the regression analysis, and so this table is omitted.

While it is disappointing that a single linear model did not emerge that allows the prediction of test average or normalized gain from an observation of written presentation, it is not surprising. The correlations observed in Chapter 6 did not show a single, small set of variables as significantly more predictive than others. So the same variation in model is expected between semesters. The predictive power of the scaled presentation variables is extremely impressive when compared to measures that include evaluated student work such as pretests, homework, and logical or mathematical reasoning tests as discussed in Chapter 1.

Chapter 8: Implications for Instruction

Based on these conclusions, there are three main implications for improvement in instruction of physics courses. First, teachers should encourage more work in written English. In particular, requiring a logically complex explanation in words and sentences of how the student arrived at his conclusion seems to ensure the highest standard of comprehension. Second, work in graphs and drawings must be required to have appropriate explanations in words and symbols. While graphics are a common mode of expression in science, using drawings without carefully integrating them in the full solution process by using them for reasoning does not contribute to learning. Finally, policy encouraging symbolic (over numeric) calculations is supported. Though the math variables were not as strongly correlated as the other types, this may simply have occurred because such policy is already in place.

Chapter 9: Further Research

Given the predictive power of the presentation data, written presentation is an important feature in student learning. The variation of the models needs to be studied for more semesters to continue to expand the number of test questions and students. Large samplings from various institutions with many different types of students would be ideal for testing whether the selected distinctions in presentation are truly predictive of success in introductory physics courses.

This study is limited in that it examines only calculus-based physics. Determining the predictive strength of written presentation in algebra-based, high school physics or advanced physics would indicate the scope of the importance of language. Furthermore, an experiment that compared a control group of students with one that was encouraged to improve their presentation in specific ways would be very enlightening.

Chapter 10: Discussion

The consistent significant correlations between the English language variables, words and sentences, and both test average and normalized gain strongly indicate that more explanation in English corresponds to increased performance in class and increased conceptual gain. Increased English presentation would, according to Hake, Hudson, and Liberman, correspond to increased conceptual gain and thus increased performance on quantitative examinations.

In nearly all the correlations, increased use of graphed objects resulted in decreased overall performance in both test average and normalized gain. The total graphed elements also had a generally negative correlation with performance, probably because it contains graphed objects. The somewhat positively correlated graphed words and graphed symbols variables were clearly major factors in cases where the total graphed elements corresponded positively to performance. The implication here is that increased use of symbols and words on a graph is required for a drawing to be beneficial. While the inclusion of a drawing as a detriment to learning is counter-intuitive, this is probably a result of students being specifically asked for a drawing rather than integrating drawing as part of their logical solution process. Without some evidence of the thought behind them, graphs provide no support to student performance.

While the mathematics variables are generally in agreement with intuitions about the role of mathematics in physics courses as well as the assessments from the previous research (they have positive correlations overall), they do not provide as striking an insight as the graph and language elements. This may be due simply to the fact that the course examined already has policy in place to encourage a particular type of mathematical presentation. Symbolic work before numbers are substituted is required for full credit on exams. Thus, the mathematics variables' success in predicting test average is not as interesting as its lack of correlation for the normalized gain. The difference in these two correlations is likely related to the fact that the symbolic mathematical reasoning is a very different mode of thinking than the logical conceptual reasoning required on concept inventories.

All the regressions suffer from a lack of continuity, which prevents the development of any specific formula for what elements to require in students' written presentation. The failure of any particular variable to stand out as more predictive than the rest indicates that a combination of factors, like that described in Hudson and Liberman, is at work here. The best general predictions for total gains come from work in English language, which correspond to both increased class performance and increased conceptual understanding.

Chapter 11: Conclusion

Scaled written presentation explains 42% of the variation in test average and 24% variation in normalized conceptual gain. In almost all cases, this is higher than the variation explained by logical reasoning, mathematics and physics pretests, previous mathematics and physics courses, and student behavior. The use of language shows strong positive correlations with student performance on exams and is a particularly strong feature in student performance on conceptual inventories. Continued and enhanced class support of writing in English, good presentation of mathematics, and carefully explained graphed work is indicated.

Bibliography

[1] Chatterjee, Samprit, *Regression Analysis by Example*, John Wiley & Sons, Inc. 1977.

[2] Freund, John and Ronald Walpole, *Mathematical Statistics*, 3rd Ed., Prentice-Hall, Inc. 1980.

[3] Hake, R. R., "Interactive-Engagement vs. Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses," *American Journal of Physics* 66, 64-74, 1998.

[4] Halloun, Ibrahim Abou and David Hestenes, "The Initial Knowledge State of College Physics Students," *American Journal of Physics* 53, No. 11, 1985.

[5] Hudson, H.T. and Dov Liberman, "The Combined Effect of Mathematics Skills and Formal Operational Reasoning on Student Performance in the General Physics Course," *American Journal of Physics*, Vol. 50, No. 12, 1982.

[6] Littel, Ramon C., Walter W. Stroup, and Rudolf J. Freund, *SAS for Linear Models*, 4th Ed., SAS Institute, Inc., 2002.

[7] Maloney, David P., Thomas L. O'Kuma, Curtis J. Hieggelke, and Alan Van Heuvelen, "Surveying Students' Conceptual Knowledge of Electricity and Magnetism," *Physics Education Res., American Journal of Physics Suppl.*, 69, No. 7, 2001.

[8] Meltzer, David E., "The Relationship Between Mathematics Preparation and Conceptual Gains in Physics," *American Journal of Physics* 70, No. 7, 2002.

[9] Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining, *Introduction to Linear Regression Analysis*, 3rd Ed., Wiley Interscience, 2001.

[10] National Science Board, “Science and Technology: Public Attitudes and Public Understanding,” *Science and Engineering Indicators 2000*, Vol. 1, 2000.

[11] Stewart, John, Jennifer McGee and Gay Stewart, “Using Time-on-Task Measurements to Understand an Introductory Science Class,” APS Presentation, Summer 2006.