Using Language Frequency Measurements to Understand Student Learning in Introductory Physics

Honors Thesis

James Roller

Abstract

A new kind of class assessment measuring quantitative properties of the communication that exists within introductory science classes is presented. Frequency measurements of specific language, math, and graph elements taken from student work in an introductory electricity and magnetism course are analyzed and used as a quantitative measure of the communication that exists within a science class. The correlation of these measurements with student conceptual gain and test average is calculated. Regression analyses are used to determine both the important communication variables affecting test average and conceptual gain and whether or not student presentation plays a statistically significant role in student learning.

Chapter 1: INTRODUCTION

Considering our increasing exposure to technology and our growing acceptance and incorporation of it within our lives, it should come as no surprise that surveys conducted by groups such as the National Science Foundation indicate that Americans are growing more and more interested and enthusiastic about technological and scientific advancements [Jennifer 1]. From iPods and camera-phones to rotary engines and motorcycle build-offs, technology is not only becoming more advanced, but also becoming much more of an integral part of our daily existence; however, despite this increase in interest and exposure to science through evolving technology, research also shows that Americans know alarmingly little about the science that governs the world around them [Jennifer 1]. Examples of this include a survey conducted by the National Science Foundation in which less than half of 2000 people questioned correctly answered that the earth revolves around the sun once every year [Jennifer 2].

Because of this alarming deficiency in scientific knowledge, many researchers in the scientific community have focused on improving student understanding of specific scientific fields. In particular, the physics community has seen a strong surge of interest from many of its professors towards increasing the understanding of their students. Initial research into these issues include work done by Ibrahim Halloun and David Hestenes which suggests that a student's initial "common-sense" understanding of physics is not only typically "incompatible with Newtonian theory", but is also not significantly changed by "conventional physics instruction"[heather 2]. As a result, the two researchers claimed that physics students tend to misinterpret the material presented in introductory physics classes. As a result of research such as this, reform efforts within physics classes have been implemented, and results from these reform efforts suggest that introductory courses engineered towards an increase in student conceptual gain do indeed yield increases in student fundamental understanding [Jennifer 4].

Here at the University of Arkansas, the introductory electricity and magnetism course – University Physics II – has been the focus of efforts such as these. The class covers electrostatics, magnetism, circuits, and optics, and is a required class for many science and engineering students. The material from this class is also tested on the

MCAT, which results in students from many other fields taking the class as well. "Revision of the structure of UPII began in 1994 and continued under NSF funding from 1995 to 2000 to provide a more hands-on lab-based learning experience. The revised course was very successful, generating high student scores on standardized evaluations and high instructor scores on course evaluations. Currently, it is being optimized under another NSF grant for distribution to other institutions [Jennifer]." Detailed data on both student performance as well as behavior tendencies is taken each semester to aid in the optimization of the course.

Most students begin UPII with little exposure to a majority of the subject matter the course presents. This is particularly true for students from Arkansas, where only a small fraction of high school students (28%) are enrolled in upper level science classes [Ellen's Thing], and since most students have never thought about such things as what an electric field is, one should not make the assumption that students' preconceived notions of electric fields hinder their true understanding of the concepts involved. With this in mind, one can reasonably assume that students gain at least a partial understanding of the concepts presented throughout the course through the work done for the course. In short, this means that student work should have a clear correlation with student understanding. The goal of this project is to evaluate student language, mathematics, and graphical presentation in such a way as to correlate student work with measured conceptual gain and test average, therefore determining what exactly physics professors can do to maximize student understanding of the concepts presented in introductory physics courses.

Chapter 2: DATA COLLECTION

Course Design

University Physics II (UPII) is the calculus-based introductory electricity and magnetism course at the University of Arkansas. The lecture meets twice a week, on Mondays and Wednesdays. A homework assignment is due at the beginning of each class, and a lecture quiz is given at the beginning and end of each class in order to both ensure attendance and gauge the class' understanding of basic concepts from the assigned readings and in-class lectures.

Labs are scheduled for two hours and meet twice a week. Students are given a lab quiz covering topics pertaining to concepts presented in the day's activity at the beginning of each lab session. Lab activities are done in groups and scheduled so as to coincide with the lecture and homework topics. Students are required to have the lab instructor review their activity work before leaving for the day.

Practice tests with solutions are published to aid in preparation for examinations. Use of these practice exams is encouraged, but not required. During examinations students are given as much time as needed to complete the exams. Exams include both multiple choice and free response questions, and partial credit is given for work shown on all non-multiple choice questions.

Concept Inventories

Students in UPII are given the Conceptual Survey of Electricity and Magnetism (CSEM) as a lab quiz at the beginning and end of each semester. This survey is designed to evaluate student knowledge and conceptual understanding of the fundamental concepts of physics dealing with electricity and magnetism [jennifer 9]. Student performance on the CSEM is recorded as a "pretest" and "posttest" score. A comparison is made between the two, and conceptual gain from the beginning to the end of the class is then measured.

A statistic called the Normalized Conceptual Gain, or "Hake Gain", the ratio of the concept percentage gained from pretest to posttest to the maximum possible concept percentage gained:

$$HakeGain \equiv \frac{(posttest)\% - (pretest)\%}{100\% - (pretest)\%} \text{ [Jennifer 4],}$$

is then used to measure each student's improvement over the course of the semester. The Hake Gain was first introduced by R. R. Hake as a method of evaluating student performance in traditional versus new interactive course structures and is considered very useful in gauging a student understanding over and entire course [Jennifer 4]. Because it is only the ratio of improvement to total possible improvement, it does not penalize courses with students that have a good understanding of the concepts to be presented.

Student Work

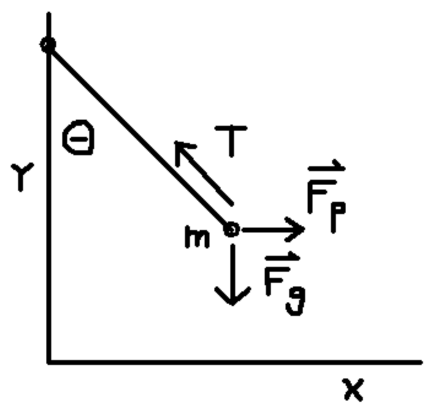
Student work was measured throughout the course of the semester. Two out of six lab sections of the class, one honors and one non-honors, were used to collect data forming a test group of approximately 30 students per analyzed assignment. Data was taken on three homework sets, six laboratory activities, and all hourly examinations.

For each of the pieces of student work analyzed, specific identifiable items within the student work were counted. These include:

- The number of words in a given student response;
- The number of sentences in a given student response;
- The number of symbols in a given student response;
- The number of numbers with associated units in a given student response;

- The number of operators in a given student response;
- The number of graphed lines in a given student response;
- The number of graphed curves in a given student response;
- The number of graphed symbols in a given student response;
- The number of words in a given student graphical response;
- The number of numbers with associated units in a given student graphical response;
- The number of equals signs in a given student response.

For example, if a student were to draw the following graph on any analyzed assignment, the following data would be recorded:



- 6 graphed lines,
- 7 graphed symbols.

Another example is the following recorded data from a response to a qualitative question about Kirchhoff's Junction equation:

"Kirchhoff's Junction equation states that the current flowing into any junction in a circuit is equal to the current flowing out the junction."

- 23 words,
- One sentence.

A final example is of a quantitative problem asking the student to calculate the electric flux through a uniform electric field. Given the following response, the following data would be recorded:

For a uniform field, $\phi_e = (\vec{E} \cdot \hat{n})A$. The given field has the form $\vec{E} = E_0\hat{x}$ where $E_0 = 2 \times 10^3\mathrm{N/C}$. For a surface in the $y - z$ plane, the surface normal is $\pm\hat{x}$, so the magnitude of the flux is

$$|\phi_e| = |E_0\hat{x} \cdot \hat{x}A| = |EA| = (2 \times 10^3\mathrm{N/C})(0.01\mathrm{m}^2)$$

$$|\phi_e| = 20\frac{\mathrm{N}}{\mathrm{C}}\mathrm{m}^2$$

- 26 words,
- 4 sentences,
- 17 symbols,
- 4 numbers,
- 6 operators,
- 7 equal signs.

All data was recorded on a per student basis, and serial numbers rather than student names were used in order to ensure confidentiality. All data was taken by hand and then entered into Microsoft Excel spreadsheets.

Total and Scaled Variables

All values for the variables previously listed were totaled for each student and each type of assignment (homework, lab activity, and hourly exam).

Total Variables

- **Total Language** was calculated by summing the total values for all words and sentences recorded for a specific assignment type:

$$(TotalWords+TotalSentences) = TotalLanguage$$

- **Total Math** was calculated by summing the total values for all symbols, numbers, operators, and equals signs recorded for a specific assignment type:

$$(TotalSymbols+TotalNumbers+TotalOperators+TotalEqualsSigns) = TotalMath$$

- **Total Graph Elements** was calculated by summing the total values for all graphed lines, graphed curves, graphed symbols, words appearing in graphs, and numbers appearing in graphs for a specific assignment type:

$$(TotalGraphedLines+TotalGraphedCurves+TotalGraphedSymbols +TotalGraphedWords+TotalGraphedNumbers) = TotalGraphElements$$

- **Total Elements** present in an assignment type was calculated by summing the total values for all of the variables:

$$(TotalLanguage+TotalMath+TotalGraphElements) = TotalElements$$

Using these variable totals, a number of new variables were then calculated in order to further understand the relationships between variables within the data sets by discounting the effect total language has on student performance. This was done since a student that turns in more work necessarily has more class assignments they can address.

Scaled Balance Variables

Balance variables represent the ratio of total elements of each category, language, math, and graphs, to the total number of elements for each of the three assignment types. These variables measure the relative distribution of language elements among the major groups: language, math, and graphs. Balance variables used include:

- **Language Balance** was defined as the ratio of Total Language to Total Elements:

$$(TotalLanguage/TotalElements) = LanguageBalance$$

- **Math Balance** was defined as the ratio of Total Math to Total Elements:

$$TotalMath/TotalElements = MathBalance$$

- **Graph Balance** was defined as the ratio of Total Graph Elements to Total Elements:

$$(TotalGraphElements/TotalElements = GraphBalance)$$

Scaled Quality Variables

Quality variables were used to define what an instructor would view as good student presentation. This includes the ratios of the number of sentences to total language, which would show a less fragmentary use of language. Another example would be the ratio of total symbols to total numbers, which would show a better algebraic presentation. Quality variables used include:

- **Language Quality** was defined as the ratio of Total Sentences to Total Language:

$$TotalSentences/TotalLanguage = LanguageQuality$$

- **Math Quality** was defined as the ratio of the sum of Total Symbols, Total Operators, and Total Equals Signs to Total Math. It measures the extent to which the student is reasoning symbolically and not numerically:

$$(TotalSymbols+TotalOperators+TotalEqualsSigns)/TotalMath = MathQuality$$

- **Graph Quality** was defined as the ratio of the sum of Total Graphed Symbols and Total Graphed Words to Total Graph Elements. It measures whether the student drawing is enriched with symbols and explanatory words:

$$(TotalGraphedSymbols+TotalGraphedWords)/TotalGraphElements = GraphQuality$$

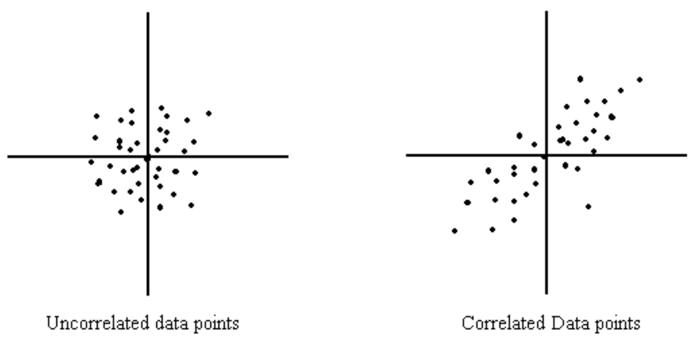
Concept Inventory

Aside from these composite variables, each student’s pretest score on the CSEM was also used as a variable.

Chapter 3: CORRELATION ANALYSES

Statistical Methods

To measure the correlation of the variables in our data set to the test average and Hake Gain, we used a sample correlation coefficient, r. As a statistic, r represents how much one variable deviates from the other. The statistic can have values within the range $-1 < r < 1$. If there is little to no correlation between the data points, $r = 0$ and the distribution of data points will be round. In contrast, if $r > 0$ or $r < 0$ then the distribution will approach a line. Perfect positive correlation is achieved when all points lie on a line and $r = 1$. Perfect negative correlation, $r = -1$, is achieved when one variable gets smaller as the other gets larger [Jennifer real, Jennifer 13].



The SAS statistic program was used to calculate the correlation coefficient and its significance [Jennifer 13].

The square of the correlation coefficient, R^2, is easier to visualize. A more important tool for understanding the measurements presented in this research, R^2 is called the coefficient of determination. This statistic measures how much variation in one variable can be explained by its relation to another variable. For each correlation, the probability that the correlation coefficient determined by the model is a result of random chance rather than actual correlations is reported [Jennifer 13]. Larger probability values suggest that the correlation between two variables is just random, while lower probability values suggest that the correlation is significant. Typically, a probability of 5% or less is needed for a correlation to be held significant [Jennifer real].

Correlation Results

Correlations for the language variables were calculated for both test average and conceptual gain. Although some results were negligible and probably the result of random occurrence, other values were quite significant. Correlation values with probabilities are listed in Table I and Table II below for all assignment types using the scaled variables.

Table II – Test Average Presentation Correlation

	Lab Presentation Correlation	Lab Presentation Probability	Homework Presentation Correlation	Homework Presentation Probability	Test Presentation Correlation	Test Presentation Probability
Language Quality	**0.38674**	**2.62%**	-0.24800	15.73%	-0.04220	80.98%
Math Quality	-0.30004	8.98%	0.24806	15.72%	**0.53443**	**0.09%**
Graph Quality	0.32338	6.64%	0.30683	7.76%	**0.33685**	**4.78%**
Language Balance	**0.42701**	**1.32%**	0.18742	28.85%	**0.39534**	**1.87%**
Math Balance	***-0.43847****	***1.07%****	-0.08383	63.74%	-0.17237	32.21%
Graph Balance	-0.09473	60.00%	-0.06860	69.99%	-0.28754	9.40%

Results in bold show significant correlation.

Table II – Hake Gain Presentation Correlation

	Lab Presentation Correlation	Lab Presentation Probability	Homework Presentation Correlation	Homework Presentation Probability	Test Presentation Correlation	Test Presentation Probability
Language Quality	0.07260	71.35%	-0.32315	8.73%	0.24241	20.25%
Math Quality	-0.21128	28.05%	0.19214	31.80%	0.22295	24.50%
Graph Quality	0.28607	14.00%	**0.40297**	**3.02%**	-0.08224	67.15%
Language Balance	0.17822	36.42%	0.19989	29.85%	**0.46468**	**1.11%**
Math Balance	-0.32778	8.86%	-0.21473	26.33%	-0.08844	64.83%
Graph Balance	0.13338	49.86%	0.08939	64.47%	***-0.54618****	***0.22%****

Results in bold show significant correlation.

Discussion of Results

Presentation and Test Performance

Test performance correlations clearly indicate that better student presentation does yield better understanding. The significant variables for this correlation are Math Quality, Language Balance, and Graph Quality.

That Math Quality is a significant factor in student understanding is no surprise since it relates all non-numerical elements of math presented in student work to the total amount of math elements presented. A higher Math Quality represents more symbolic, abstract work. This means that students that work out mathematical problems farther with symbols tend to understand more than those that plug numerical values into equations early in the solution process.

That Language Balance plays a significant role in student understanding means that students that communicate more with words and sentences typically will understand more than those that just do the mathematical aspects of the assigned work. This result makes sense as well since students that explain their answers with words as well as math will be more likely to develop a qualitative sense of what the math they did represents, and by writing an explanation of their answers, these students are forced to put concepts learned in the course into their own words.

The significance of Graph Quality implies that students that express ideas in graphs with more than just the actual drawings tend to understand more. This means that students that use words and symbols to label and explain their drawings will have a better understanding of course materials. This fits together well with the significance of Language Balance, since Graph Quality is a measurement of the relationship between student language and student drawings; so the use of language in problems that may not require it results in increased understanding.

Presentation and Conceptual Gain

Conceptual gain results further show that student performance is correlated with language presentation. The primary effectors of conceptual gain in this correlation are Graph Quality and Language Balance.

The fact that Language Balance and Graph Quality play significant roles in both the Test Average and Conceptual Gain Correlations is worth noting. Students that are more thorough in their explanations will understand concepts better than those that do not. The use of language in course work is a clear benefit to student understanding both for examination purposes and overall conceptual gain.

Chapter 4: LINEAR REGRESSION ANALYSES

Statistical Methods

Linear regression analysis is used to determine the linear equation of a line that represents the best relationship between two or more variables [Jennifer 15]. Since multiple linear regression analyses were used to determine the relationship between the many variables analyzed in this research, some statistical adjustments had to be made.

For a simple two-variable data set, the line fit is of the form $y = \alpha + \beta x$. Each value of x_i is used to predict $\hat{y}_i$ through $\hat{y}_i = \alpha + \beta x_i$, where α and β are the estimates of the slope found in the regression [Jennifer 15].

Data for cases like this are in the form of ordered pairs, (x_i,y_i). Each value of x_i can be used to calculate a corresponding value of $\hat{y}_i$. Total variation within a data set is measured by the total sum of squares,

$$SST = \sum (y_i - \bar{y})^2$$

with $\bar{y}$ equal to the average of y [Jennifer 15], so the remaining error left after the model of the data is applied is determined by the variation from the actual value, y, and its predicted value, $\hat{y}_i$

$$SS_{res} = \sum_i (\hat{y}_i - y_i)^2 \text{ [Jennifer 15].}$$

The percentage of variation measured by the assumption of a linear relation regression is called the Coefficient of Determination, R^2.

$$R^2 = 1 - \frac{SS_{res}}{SST} \text{ [Jennifer 15]}$$

Values of R^2 close to one indicate that most of the variation is removed and yield a straight line, while values close to 0 indicate low significance between the two [Jennifer 15].

Unfortunately, as variables are added to the regression, R^2 will increase regardless of correlation significance. In order to produce the best statistical model of a data set with a large number of variables, an adjusted R^2 value that accounts for the loss in degrees of freedom from an increase in parameters must be calculated, $R^2_{adj.}$ [Jennifer 15]. The formula for $R^2_{adj.}$ is

$$R^2_{adj} = 1 - \frac{SS_{res}/(n-p)}{SST/(n-1)} \text{ [Jennifer 15],}$$

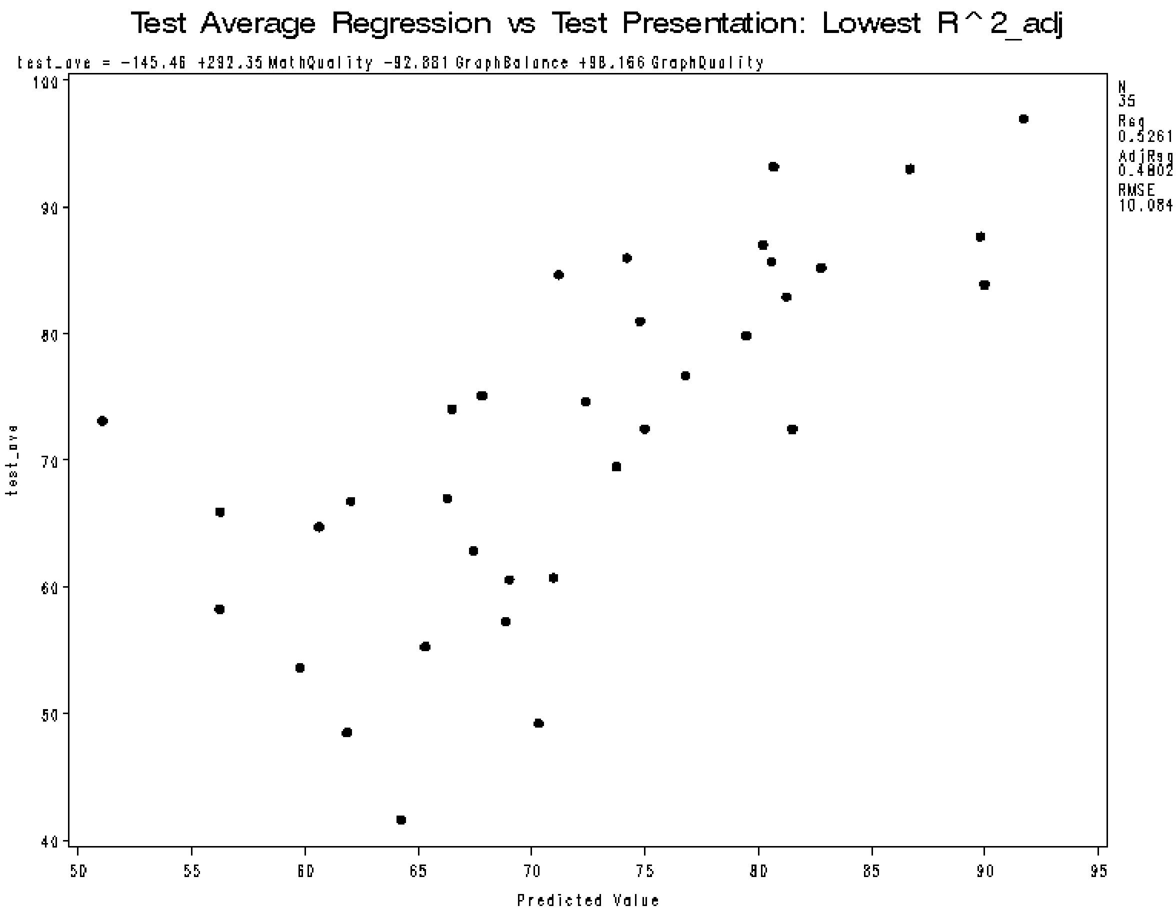
where p is the number of fitting parameters and n possible parameters, and n is the number of data points [Jennifer real].

Linear regression analyses were carried out for all types of student work using the PROC REG procedure in SAS. These were used to model both Hake Gain [Jennifer 14] and test average. The $R^2_{adj.}$ statistic was used to select the most significant set of regression variables.

Linear Regression Results

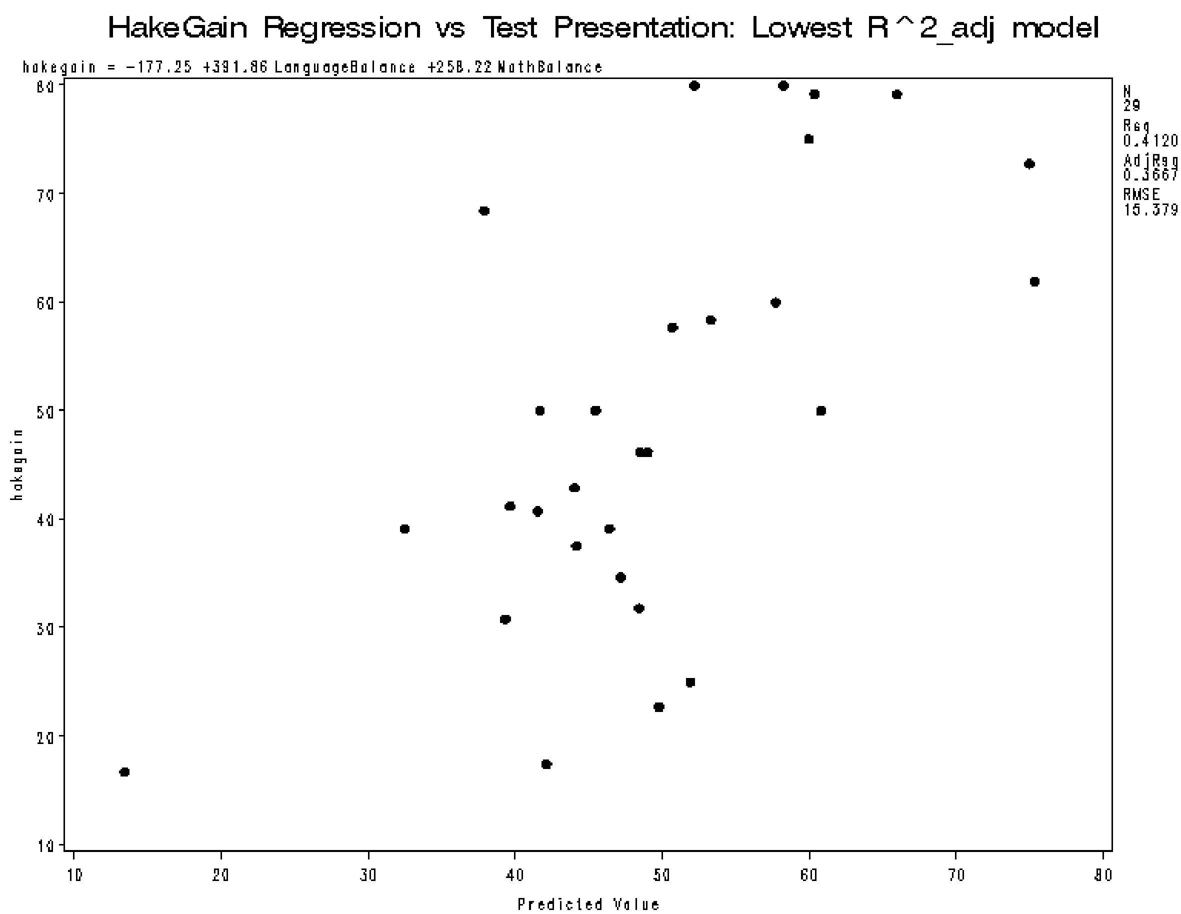
Regressions were performed using language variables to predict both test average and Hake Gain. The most significant linear regression plots for both of these are presented in Figure I and Figure II along with an explanation of the resulting coefficients of determination for all assignment types in Table III.

Figure I



The graph plots the observed test average vs. the test average predicted by the regression for the lowest R_{adj}^2 model (3 variables).

Figure II



The graph plots the observed Hake Gain vs. the Hake Gain predicted by the regression for the lowest R_{adj}^2 model (2 variables).

Table III – Adjusted Coefficient of Determination

	Lab R^2 (R_{adj}^2)	**Homework R^2 (R_{adj}^2)**	**Test R^2 (R_{adj}^2)**
Test Average	0.3549 (0.2882)	0.2791 (0.1504)	0.5261 (0.4802)
Hake Gain	0.2045 (0.1409)	0.3048 (0.1889)	0.4120 (0.3667)

Explanation of Results

Presentation and Test Performance

Both lab and test presentation show a substantial amount of the initial variation of a linear relation with presentation variables. Results show that approximately 35% of the variance can be explained by lab presentation, and over 50% can be explained by test presentation.

Presentation and Conceptual Gain

While results for conceptual gain were not as strong as those for test average, test presentation still shows a high R^2 at 37%.

Chapter 5: COMPARISON TO PREVIOUS METHODS

Using presentation variables as predictors of student performance shows much more predictive power than other variables related to student preparation and behavior. Previous research conducted on UPII measuring the relationship between student performance and behavior with test average and conceptual gain yielded 23% correlation for test average and 17% on concept inventories [Jennifer real].

Variables for this study measured student actions such as taking lecture notes, studying for exams, and reading course assignments, and other variables gauging student preparation for the course were calculated using the Conceptual Survey of Electricity and Magnetism [Jennifer 9]. Results of regression analysis using behavior variables with and without pretest are shown in Table IV.

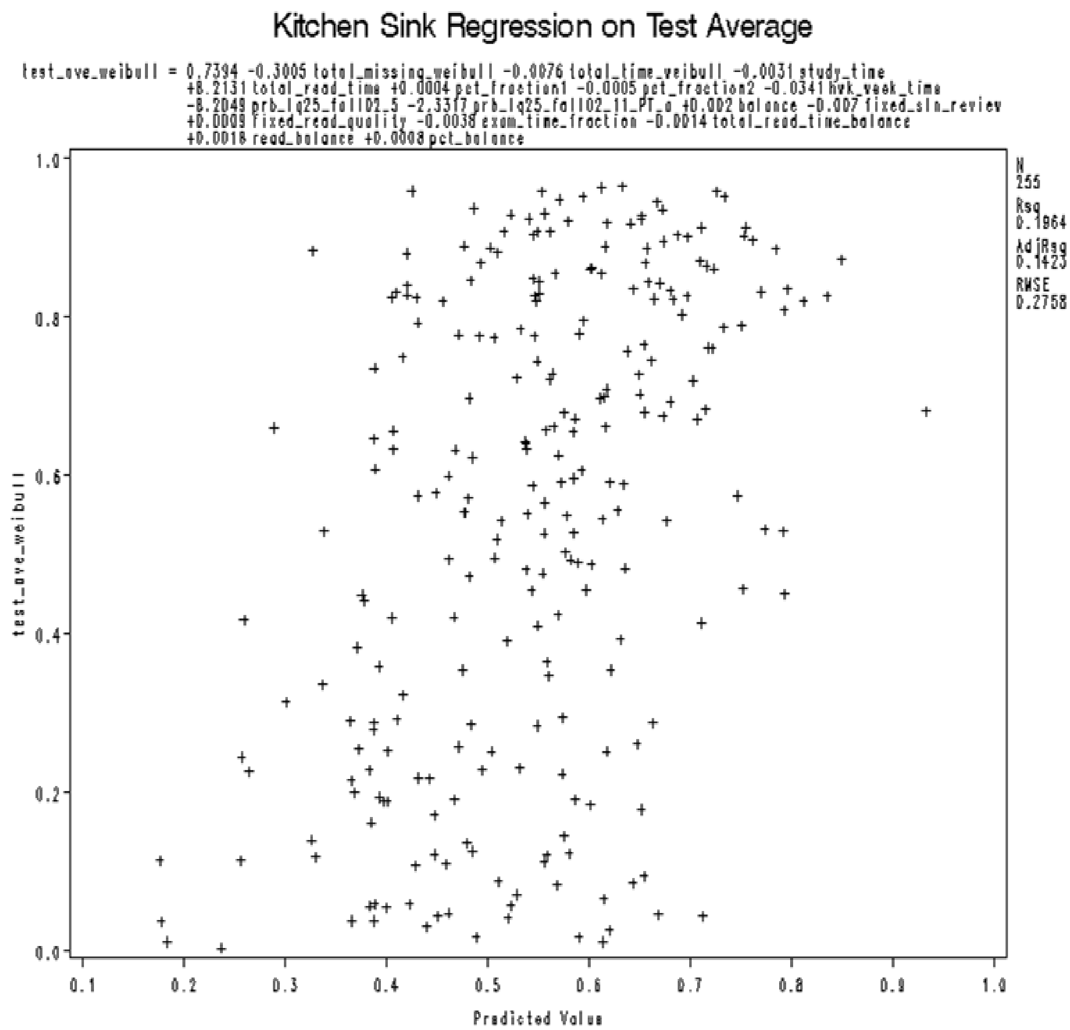
Table VI – Behavior and Preparation vs. Student Performance

	R^2 (R_{adj}^2)
Behavior with Test Average	0.1967 (0.1651)
Behavior and Preparation with Test Average	0.2654 (0.2345)
Behavior with Hake Gain	0.1845 (0.1507)
Behavior and Preparation with Hake Gain	0.2065 (0.1701)

[Jennifer real]

The prediction of the test average from 17 behavior variables is plotted against observed test average in Figure III. A comparison of Figure I and Figure III show that the presentation variables used in this study remove far more noise from the data than either the pretest or student behavior.

Figure III



[Jennifer real]

For test average, behavior and preparation explained 23% of variance, while the new assessment yields results of 48%. Furthermore, behavior and preparation only accounted for 17% of conceptual gain, while presentation accounted for 37%.

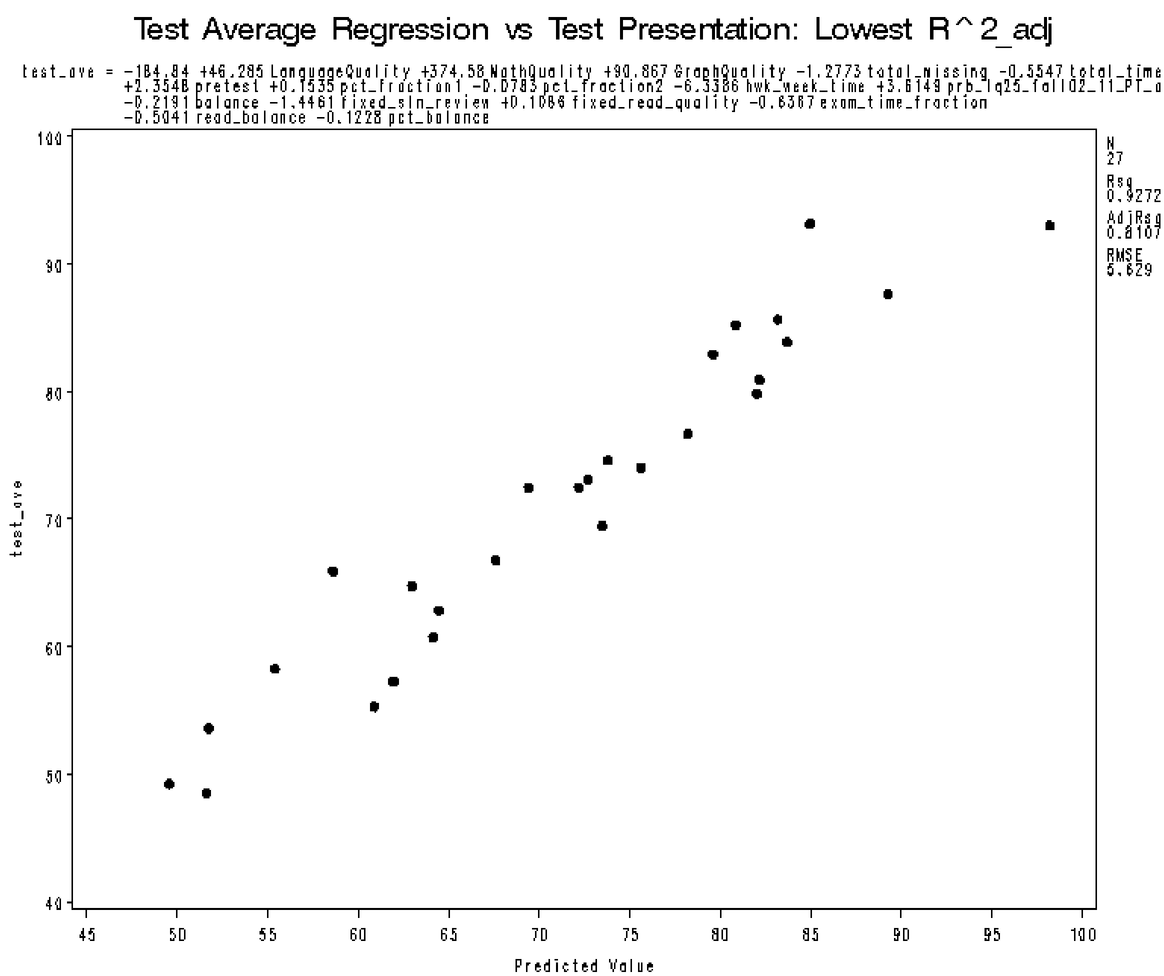
Chapter 6: COMBINED REGRESSIONS

Because of the utility the language variables proved to have in predicting student performance, regressions containing these variables along with those for student behavior were carried out. Because of the large number of variables used to fit the data, both the most significant R^2 value and the R^2 for the most significant five variables are reported for each assignment type.

Table V – Adjusted Coefficient of Determination

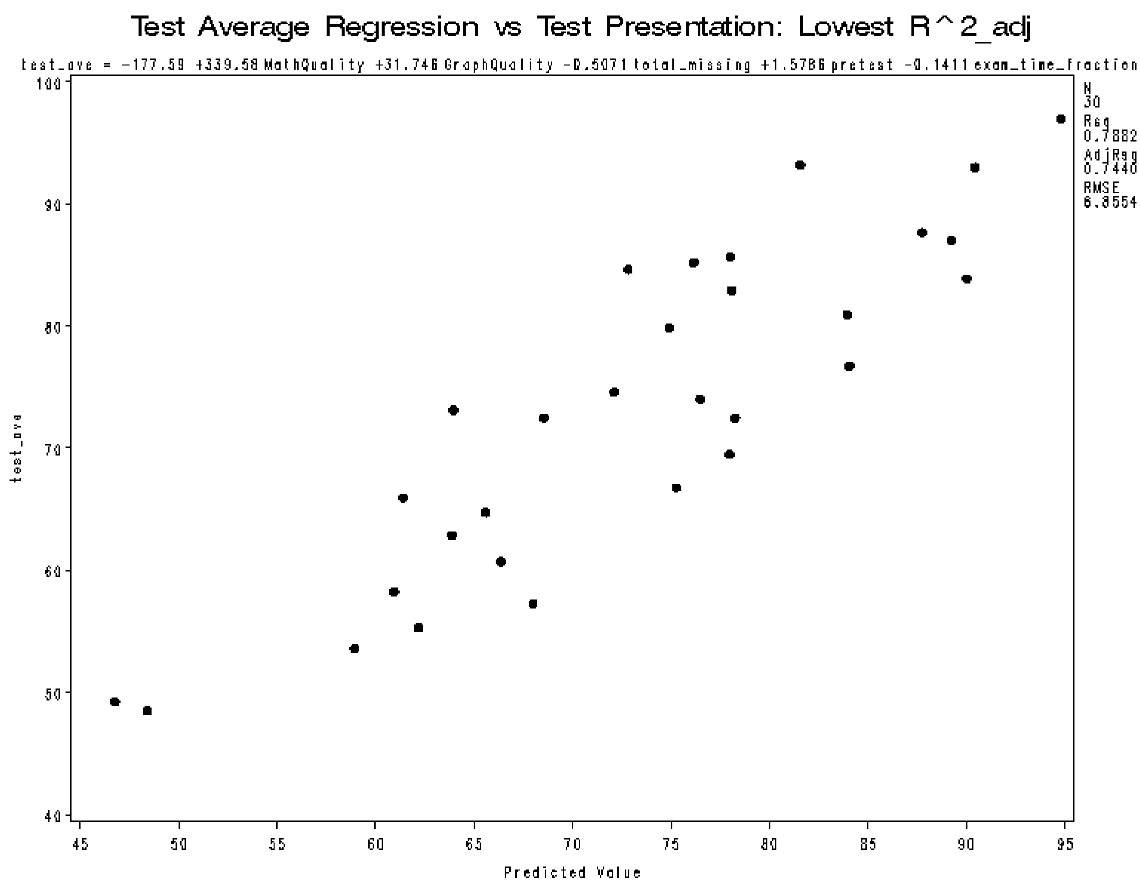
	Lab with Behavior R^2 (R_{adj}^2) N = 33	Lab with Behavior R^2 (R_{adj}^2) N = 5	Homework with Behavior R^2 (R_{adj}^2) N = 35	Homework with Behavior R^2 (R_{adj}^2) N = 5	Test with Behavior R^2 (R_{adj}^2) N = 35	Test with Behavior R^2 (R_{adj}^2) N = 5
Test Average	0.9198 (0.7973)	0.5669 (0.4586)	0.4198 (0.2263)	0.3495 (0.2081)	0.9272 (0.8107)	0.7784 (0.7257)
Hake Gain	0.8954 (0.7718)	0.4739 (0.3354)	0.7033 (0.5364)	0.6275 (0.5344)	0.9616 (0.8932)	0.6418 (0.5523)

Figure IV



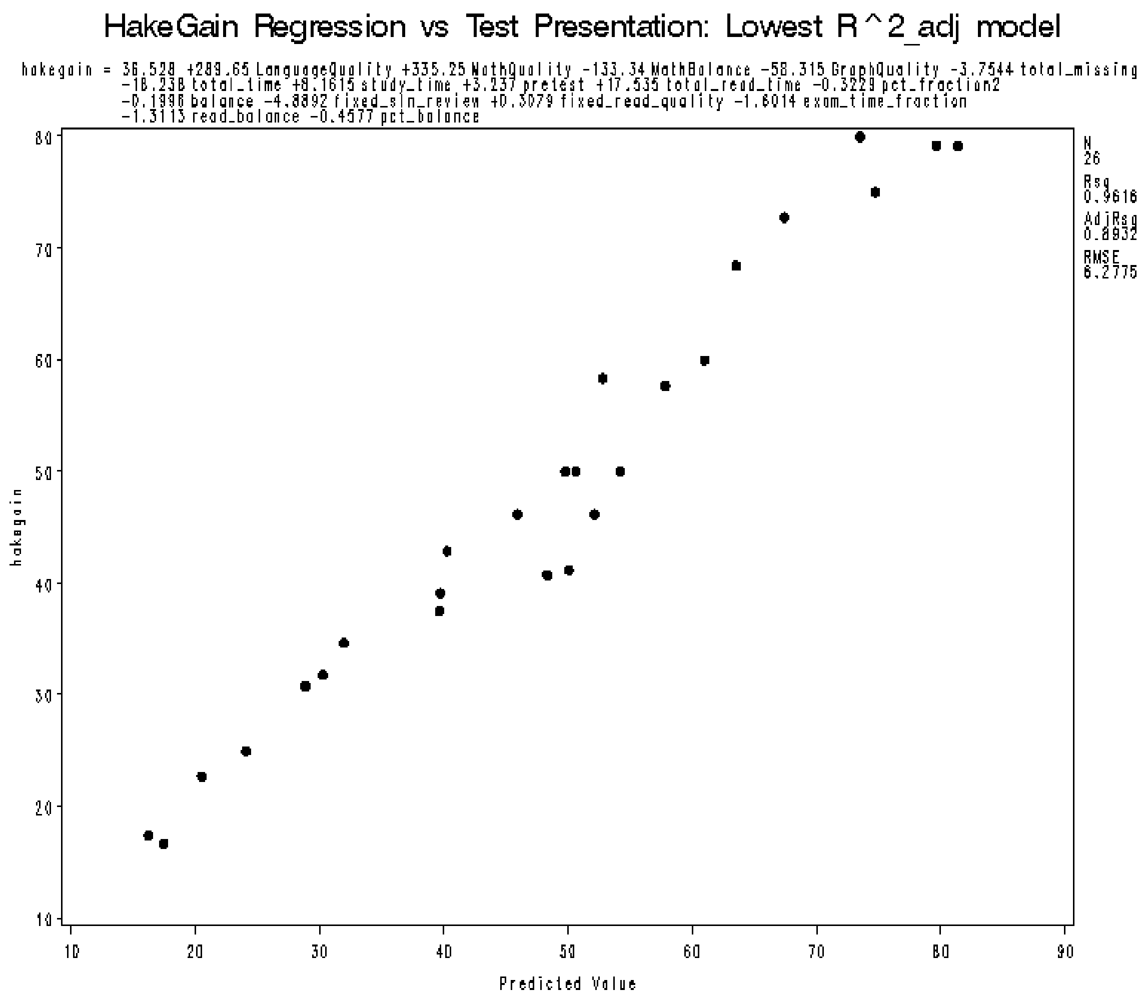
N = 35

Figure V


Test Average Regression vs Test Presentation: Lowest R^2_adj
test_ave = -177.59 +339.58 MathQuality +31.746 GraphQuality -0.5071 total_missing +1.5786 pretest -0.1411 exam_time_fraction
N
30
Rsq
0.7882
AdjRsq
0.7440
RMSE
6.8554
test_ave
Predicted Value
100
90
80
70
60
50
40
45
50
55
60
65
70
75
80
85
90
95

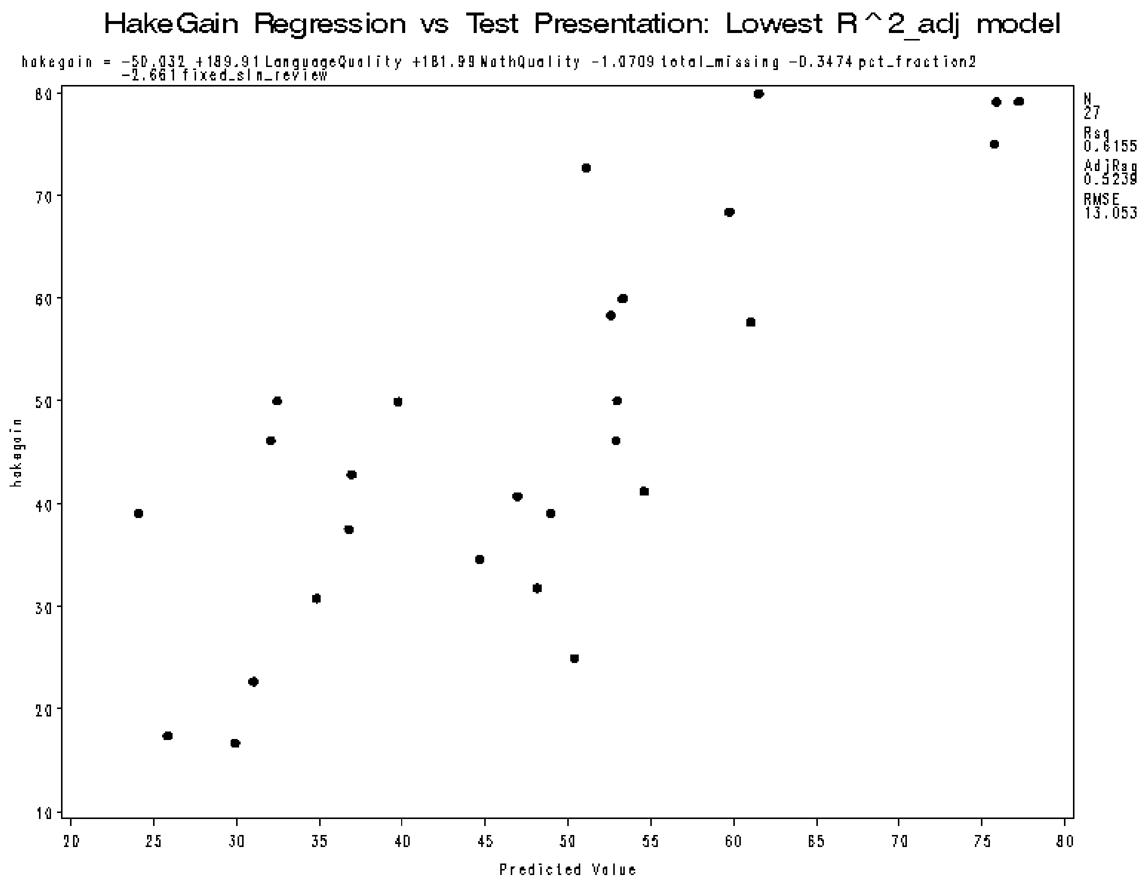
N = 5

Figure VI



N = 35

Figure VII



N = 5

Explanation of Results

The combination of behavior and language variables removed an impressive amount of the variation in the data.

Presentation and Behavior vs. Test Average

Test average showed a high correlation with both the lab activity and test presentation when combined with behavior statistics. Nearly 92% of student performance can be explained with behavior and lab presentation, with a chance of random recreation of the data at 0.14%, and limiting the number of variables to five still yielded results of 57% variance explained.

Test presentation was even more conclusive when combined with behavior variables. Results indicate that 93% of student performance on exams can be predicted when using a behavior and presentation assessment, with the chance of random recreation at 0.1%, and even with the five variable limitation, 78% of variation could still be explained.

Presentation and Behavior vs. Conceptual Gain

Significance in the regression for conceptual gain existed in all three assignment types. Data indicates that 77% of conceptual gain can be explained by lab activities with a chance of random recreation at 0.12%, and over one third of the variance can be explained with using the five variable regression.

Homework presentation had a 70% significance with student conceptual gain when combined with behavior variables at a probability of random generation at 0.6%. At the five variable mark, this was only reduced to 63% of variance explained.

Test average again yielded the highest results, with 96% of variance explained when combined with behavior. This result had a probability of random generation of 0.02%, and even at only five variables, test presentation and behavior had a predictive index of 64% for conceptual gain.

Chapter 7: DISCUSSION

Results

Presentation showed impressive predictive power for performance. Correlation analysis indicates that students that use more variable notation in calculations as well as explain answers and graphs in words and sentences will learn more of the concepts presented in class and perform better on examinations.

Future Analysis

Further measurements over a much larger group of students are needed. Due to time constraints, only semester totals of the communication elements measured were used in analysis. Measurements taken for specific problem types, such as the value of student presentation in only quantitative problems or the value of student presentation on graphing vs. qualitative problems, could be analyzed for a much greater level of detail in the understanding of communication within science classes.

Possible Applications of Results

Using the variable correlations calculated, many applications of this study could be implemented in a classroom setting. One possible application of the results would be to have all responses be verbally explained in order for credit to be given. A second possibility is to have all mathematical work shown in variable notation in order to receive credit, and another would be to require all elements in graphical responses to be properly labeled with a significant level of detail in order for students to receive credit. UPII currently uses some policy to encourage all these possible outcomes.

Chapter 8: CONCLUSION

Results provide clear indications that student presentation is correlated with higher performance. Results of this research indicate that especially when combined with other performance assessments, presentation can be used to determine as much as 92% of test performance and 96% of conceptual gain, and this is clear improvement to previously used methods.