Predictors of final exam score in CS1: Language ability is important

Approved: ____Mike Rowe__ Date: _February 27, 2013__

Committee Chair - Prof. Mike Rowe, Ph. D. (UWP)

Approved: __Dr. Bettina Harriehausen__ Date: __February 22, 2013__

Committee Member - Prof. Dr. Bettina Harriehausen(h_da)

Approved: __Robert Hasker__ Date: _March 1, 2013__

Committee Member - Prof. Rob Hasker, Ph. D. (UWP)

Predictors of final exam score in CS1: Language ability is important

A Thesis

Presented to

The Graduate Faculty

University of Wisconsin-Platteville

In partial Fulfillment of the Requirement for the Degree
Master of Science in Computer Science

by

Robert W. McKinney

2013

PREDICTORS OF FINAL EXAM SCORE IN CS1:
LANGUAGE ABILITY IS IMPORTANT

Robert W. McKinney
Under the Supervision of Dr. Michael Rowe, Dr. Robert Hasker
and Dr. Bettina Harriehausen

Abstract

Over the past 30 years there have been several studies that investigated various variables in an attempt to predict success in computer science related courses. One of the more commonly studied courses has been CS1, Introduction to Programming. The motivation for many of these studies is the high average attrition rate of first and second year computer science students. Attrition rate has been reported by some schools to be as low as 19% and others are reporting attrition to be as high as 66% (Cohoon and Chen). Over the last three years the University of Wisconsin-Platteville has experienced a 70% success rate in the CS1 equivalent course, CS 1430, as measured by the number of students who earn a 70% or better.

This study examined several college entrance exam scores. These scores are available through the institutional research department for most students attending the University of Wisconsin-Platteville. The purpose of this study was to determine which college entrance exam scores are good predictors of the CS 1430 final exam score at the University of Wisconsin-Platteville. The University of Wisconsin-Platteville is primarily an undergraduate school consisting of approximately 8,000 students. The introductory computer science course, CS 1430, is taught by the Computer Science and Software Engineering department. This department is part of the university's engineering college.

This study used the final exam score from the CS 1430 course as the dependent variable. The final exam in the CS 1430 course consists of two sections. This study used the objective section of the final exam, which consisted of 50 multiple choice and true/false questions. The second section, which is the more subjectively scored programming section of the final exam, was not used. A number of independent variables were used including ACT Composite and subscale scores and Wisconsin placement scores for English and Mathematics.

Stepwise multiple regression analysis was performed to determine the best set of independent variables for predicting scores on the department's standardized final exam. The Wisconsin Placement Test Math score was the best single predictor of the final exam score, with a correlation of 0.501. The next variable added by the step-wise regression was the Wisconsin Placement Test English score, which increased correlation to 0.570. The third variable added by the step-wise regression was the ACT English score, which had a negative coefficient that decreased the total uncorrelated area and which increased the correlation to 0.594. No other variables were found to increase the statistical significance. Of the 504 students that completed the CS 1430 course between the fall semester of 2008 and the fall semester of 2010, 102 students had complete data allowing this study to be conducted. This study successfully constructed a prediction model using the data collected. The prediction model maybe used to accurately place students into the correct introductory programming course.

Table of Contents

List of Tables

List of Diagrams

1. Introduction

One goal of an academic advisor and/or academic department at the university level is to place students into the appropriate entry level class. Currently, that goal is achieved through a number of standardized exams and personal interviews with the student. Specifically, in computer science departments this goal is often achieved through a combination of evaluating such factors as high school programming background, previous academic performance in math and occasionally a personal interview of programming knowledge. Often, such evaluations fall short and are not based on any empirical evidence. This study addresses the need to produce a comprehensive tool for placing students into the appropriate introductory programming course. There has been a need for such a tool since computer science courses have been offered. Both students and computer science departments can benefit from a tool that appropriately matches the student's interest in programming and aptitude. (T. R. Hostetler)

Several studies, beginning in the late 1970's attempted to predict a student's success in a programming course based off past academic achievement using the student's grade point average as the single best predictor. (T. R. Hostetler) The world has progressed significantly since then, especially in the computing science fields. Even though grade point average was the best predictor of success then, this may not be true today.

Over the past twenty five years computers have gone from not being a part of daily life to being intertwined with every aspect of one's daily routine. These studies date back to the first days of personal computers when having a computer in the home was the exception. In 1984, only 8.2% of households owned a computer. By 2003, 61.8% of households owned at least one computer. (United States Census Bureau) Children start interacting with computers even before they start school. One goal of this thesis is to reevaluate the previous findings in light of the dramatic increase in the popularity and daily use of the personal computer.

At the university level, there are often two introductory computer science courses: a CS 0.5 course and a CS1 course. CS1 is typically a standard introduction to programming course which presumes little to no previous experience and prepares students for the CS2 course on data structures. The CS 0.5 course is typically a slower-paced course that is available to students who may not be ready for a more rigorous CS1 course.

The purpose of this thesis was to research, analyze and develop tools and measures that will allow advisors to effectively place students into the appropriate introductory computer science course, whether it is CS1 or a slower paced CS 0.5 computer science course. The need to place students into the appropriate introductory computer science course is more important and more difficult than ever before. The tools developed in this study will make the course placement process much easier and will give the student the greatest chance for success while at the same time lowering attrition rates.

2. Previous Work

The topic of predicting performance in CS1 has been the subject of many studies. Three types of predictors have been studied; these include measures available prior to the course, during the course and after the course. All of the measures have the goal of selecting the students who will perform well in CS1. The following sections summarize these findings.

2.1 Predictors Available Prior to CS1

There were a number of studies that used combinations of the students' personality traits, cognitive skills and previous academic performance to build models that would predict CS1 performance. Hostetler successfully classified 61 of 79 students (77.2%) into low and high aptitude groups using a multiple regression equation developed from 5 predictors. The study investigated to what extent certain cognitive skills, personality variables and past academic achievement were predictive of CS1 success. Of the 600 students enrolled in the CS1 course, 120 students were randomly selected. Missing data reduced the final sample size to 79 students. The dependent variable was the final exam numerical score in the CS1 course. The five independent variables found to have a significant correlation to CS1 success included Diagramming and Reasoning from the Computer Programmer Aptitude Battery Test, college GPA, math background and the personality trait variable PF05. A high score on personality trait PF05 represented a "Sober/Happy-go-lucky" personality type. This was a very good prediction model for the time with a multiple correlation value of 0.653. Hostetler concluded that changing educational trends and the introduction of the personal computer into the home prompted for a new model to be developed.

Rauchas, Rosman and Konidaris performed a study that used language performance at the high school level as a predictor of CS1 success. The authors collected data through two different methods. In method one they surveyed newly enrolled students in three areas: their comfort with language, their reading habits, and their perception of the importance of language in studying computer science. The second method was a more in-depth quantitative analysis of the high school examination results. The study focused more on the latter: the student's performance on the high school final examination to predict their success in CS1. The study did try to distinguish between performance in computer science and more general computing topic of programming. Their study also dealt with the wide range of English language ability as it was based in South Africa where English is not the first language of all students. Although, all students enrolled in the CS1 course had passed the English matriculation exam. The study included 107 students across two CS1 courses using four predictors for correlation (mathematics, English as a first language, English as a second language, and all first language). The study found that English as the first language had a strong statistically significant positive correlation for both CS courses. The study suggests that language courses at the high school level are better predictors of success for CS1 courses than mathematics at the high school level. One of the goals of this thesis is to determine if there is a similar correlation for students whom English is a first language.

2.2 Predictors Available During a CS1 course

Bruce, Buckingham and Hynd categorized the students into five learning types:

1. Following: getting through the unit.
2. Coding: learning the code.
3. Understanding and integrating: through understanding and integrating programming concepts.
4. Problem solving: do what it takes to solve a problem.
5. Participating or enculturation: discovering what it means to be a programmer.

A question posed by the authors was, "How can curriculum support the ways of going about learning?" (Bruce, Buckingham and Hynd 143) In that question, the authors observed that students either focused on programming in 'parts' or in 'wholes'. The students that focused on 'parts' desire the information in more manageable amounts in order to receive continuous feedback while completing individual tasks. The students that focused on the 'wholes' were more concerned with syntax and coding. The ability to break programming requirements down into parts or read the programming requirements as a whole would suggest the students would need strong English and reading comprehension skills to be successful. Students that lacked such skills preferred to learn in alternate ways. In types three and four, the student focused on 'understanding and integrating' and 'problem solving'. The authors saw the programming activity as a learning experience that would suggest the student had a more developed reading comprehension skill set, giving them the ability to clearly read and understand programming requirements.

Bruce, Buckingham and Hynd also suggested an incremental, mathematical approach that compliments the ability to understand the programming requirements. Once the student clearly understood the programming requirements, English and reading comprehension, the most successful students had the ability to break those requirements into more manageable pieces to solve one at a time. Using these solved pieces, the student then may have a broader understanding of the 'whole' problem. This is a very similar mechanism used by some for solving complex mathematical problems. The difference is that they are not just solving mathematical equations, but doing abstract problem solving.

Most CS1 courses do not utilize just one or even two of these categories of learning. Most touch on all five category types. Bruce, Buckingham and Hynd suggested that teachers of programming who want their students to progress should expose them to the more sophisticated ways of learning. This would suggest that a student who possesses the broadest range of core academic skills (math and language) would have the greatest chance of success.

Black focused less on changing the curriculum or altering teaching methods and more on accessing the student's data available previous to entering CS1. Hagan and Markham suggested that students with previous programming experience perform significantly better than students without any programming experience.

The Black study and the Roddan study both shared the same goal, which was to identify students at risk as early as possible so that help could be provided. Black's goal was to advance the research by Roddan by trying to explain the high amount of variance in the exam score found in Roddan's study. Both studies focused on the student's own ability to judge how well he or she understood the course material through questionnaires administered during the CS1 course. The Black study found that the student's self-evaluation by week four of CS1 had a reasonably strong correlation of 0.586 to exam scores, but these results should be used with caution because of the modesty of the students' self-evaluations. Another factor that had a significant correlation was academic and social integration. Black used Tinto's 1975 theory on student attrition that examined a number of independent variables to determine the likelihood that a student would be successful in college. The theory focused on how well the student integrated into the university during the 'separation' stage. Black also used the students' tutor's predictions. The students' tutor's predictions correctly predicted a pass or failure over 60% of the time. Prior programming experience and study habits were found to be insignificant factors at predicting the exam scores.

2.3 Predictors available after a CS1

Ford and Venema focused on the students that passed CS1, but were later unable to correctly answer multiple choice questions on programming fundamentals. Ford and Venema administered the Dehandi Test to 111 students (98 males, 13 females) that had all passed a CS1 equivalent course. The Dehandi Test examined both correctness of answers and consistency of answers given. The CS1 course was taught in English and the study found that students who spoke another language more fluently than they spoke English scored worse in CS1 than did other students. This suggests a fluency in the language of instruction is a critical component to a student's success rate in CS1.

2.4 The University of Wisconsin-Platteville CS1 and CS 0.5 Courses

At the University of Wisconsin-Platteville there are two introductory computer science courses. The CS1 course equivalent is *CS 1430: Programming in C++* and the slower paced CS 0.5 course equivalent is *CS 1130: Introduction to Programming.* Most students majoring in computer science are placed into CS 1430, but CS 1130 is available for students that are perceived to not be ready for CS 1430.

CS 1430 covers core programming skills including procedural programming in C++, algorithms, modularity, and abstraction. There are ten graded labs spread evenly throughout the course that reinforce lecture materials. There are a number of out-of-class assignments including programming assignments. These give the student a chance to tie the concepts together and to learn basic programming style, documentation, and development skills necessary for working in a team environment. Topics covered in CS 1430 include expressions, control constructs, functions, arrays, and simple objects. (University of Wisconsin-Platteville Computer Science Department)

The slower paced *CS 1130: Introduction to Programming* is for students without any previous computer programming experience. The course has no prerequisites. It is split into two parts. Part 1 is taught through programming in Karel the Robot. Karel is an educational programming language for beginners. (Pattis) Topics covered in this part include basic instructions, program control and problem solving. The second part is a very basic introduction to C++. Topics covered include: C++ program structure, variables, control statements and simple graphics programming. (University of Wisconsin-Platteville Computer Science Department) The course work is less rigorous than CS 1430 with an emphasis on using in-class labs to ensure the student is building a solid foundation for future CS courses.

CS 1430 has an approximate success rate of 70%. Success is defined as obtaining a sufficient grade to satisfy the requirements for majors and to advance to the data structures class; that is, passing CS 1430 with a 70% or better. Unsuccessful numbers include both students earning below 70% and students who withdraw from the class after the second week of the semester. An unpublished, informal study by Parsons and Hasker found that prior mathematical academic performance was a good predictor of success in CS 1430. The study presented in this thesis attempted to improve the prediction rate and focused on a prediction model that included language ability.

This thesis continued the research of developing an accurate prediction model by using a different set of predictors than the aforementioned papers in an attempt to improve on student attrition rates in CS1. College entrance data, available for most college students in Wisconsin, was used to develop a predictive model of CS 1430 success as measured by scores on the CS 1430 final exam.

3. Method

Typically, there are four to six sections of the CS 1430 course taught by the Computer Science Software Engineering (CSSE) department per semester at the University of Wisconsin-Platteville. The CSSE department used a department-wide final exam across all sections of CS

1430. The exam has remained relatively constant for many years. Half of the exam consists of 50 multiple-choice and true/false questions. This part of the final exam provides an objective measure to use as the dependent measure in a regression study. For this study, the students from the fall semester of 2008 through the fall semester of 2010 were evaluated. The sample consisted of a total of 540 students.

3.1 Data

The independent and dependent variables that were used in this thesis can be found in Table 1. Due to missing data, the final sample contained complete data for only 102 students enrolled in CS 1430 from the fall semester of 2008 through the fall semester of 2010. The data for this thesis was obtained from a variety of sources including the university's institutional research department, the Mathematics department and the admissions office. There were a number of difficulties in obtaining complete data from the institutional research department due to understaffing, department restructuring and data maintenance issues.

Table 1: Dependent and Independent Variables Used For the Study

Dependent Variable:
CS 1430 Final Exam Score
Independent Variables:
High School Percentile – Reported by the student's high school
ACT Composite - Reported by the ACT
ACT Math - Reported by the ACT
ACT English - Reported by the ACT
ACT Science - Reported by the ACT
ACT Reading - Reported by the ACT
WPT Math - Reported by Wisconsin Testing Centers*
WPT Math Sub Score 1 - Reported by Wisconsin Testing Centers*
WPT Math Sub Score 2 - Reported by Wisconsin Testing Centers*
WPT Math Sub Score 3 - Reported by Wisconsin Testing Centers*
WPT English – Reported by Wisconsin Testing Centers*
*University of Wisconsin Placement Test (WPT)

University of Wisconsin Placement Test is used by all universities in the UW System for placing students into the appropriate courses based on scoring levels. During the fall semester of 2008 and the spring semester of 2009, 175,056 students were enrolled at the 14 universities in the UW System. This study used the University of Wisconsin Placement Test English (WPT English) score and the Wisconsin Placement Test Math (WPT Math) score. Not all students are required to take these placement tests. For example, if a student has already earned college credit in Mathematics or English, they are generally exempt from taking the placement tests.

The WPT English consists of three subtests: English usage, sentence correction and reading comprehension. English usage items require a student to identify deviations from standard written American English. Sentence correction items require a student to select the most effective expression from among five choices. The reading comprehension section evaluates how well a student can interpret and understand prose passages that will be encountered in college level reading. The WPT English test helps place students in an appropriate English courses based on their functional ability. Specifically, it is used to place students in remedial English, freshmen composition 1, freshmen composition 2, as well as granting college credit for freshmen composition 1 and/or 2 at the University of Wisconsin-Platteville and some of the other University of Wisconsin campuses. (University of Wisconsin Testing and Evaluation Services)

The WPT Math consists of three broad categories of items: WPT Math Sub Score 1 (mathematics basics), WPT Math Sub Score 2 (algebra), and WPT Math Sub Score 3 (trigonometry). The three sub scores are used to calculate a composite score that ranges from 0 to 40. The test is scored as the number of correct answers, with no penalty for guessing. The WPT Math and WPT Math Sub Scores help in placing students in appropriate entry level mathematics courses. Specifically, the test is used to place students into one of two levels of remedial mathematics courses, College Algebra, Finite Mathematics, Elementary Statistics, Pre-calculus, Trigonometry and Analytic Geometry, and Calculus 1. This test is not used to grant Mathematics credit (University of Wisconsin Testing and Evaluation Services) .

The WPT English and WPT Math exams are designed as a test of skill and not speed. The University of Wisconsin Testing and Evaluation services consider 90 minutes to be ample time for most students to answer all questions. Each year, a new form of the exams is published. (University of Wisconsin Testing and Evaluation Services).

The CS 1430 course has a final examination that is comprehensive and is split into two portions. This study examined the 50-point multiple choice and true false portion which is standardized across all sections of 1430. Table 2 provides descriptive statistics for 102 of the 540 students' data used.

On all measures, the sample group had higher scores than the general population of University of Wisconsin-Platteville students. There are a number of notable observations in Table 2. One notable observation is that students in the sample group achieved, on average, 3.91 points higher on the ACT Math and 3.16 points, on average, higher on the ACT Composite than the general population at University of Wisconsin-Platteville. Even more notable is the difference in WPT Math and WPT English scores between the sample group and the campus wide average. The sample group achieved a score of more than 10 points higher than the campus average on the WPT Math and over 45 points higher on the WPT English. The academic caliber of the students within the sample group, on average, was higher than the average student at the University of Wisconsin-Platteville.

Table 2: Descriptive Statistics of Independent and Dependent Variables

Sample size N = 102	Mean (CS1430 Students)	Standard Deviation (CS1430)	Mean (Campus Wide)	Possible Max Score
CS 1430 Final Exam Score	35.73	6.816	n/a	50
High School Percentile	71.75	18.739	64.35	100
ACT Composite	25.75	3.349	22.59	36
ACT English	23.32	4.344	21.14	36
ACT Math	27.16	3.929	23.25	36
ACT Science	26.44	3.916	22.86	36
ACT Reading	25.50	4.520	22.15	36
WPT Math	28.87	12.103	17.72	40
WPT Math Sub Score 1	637.75	129.790	534.79	850
WPT Math Sub Score 2	620.10	119.616	508.20	850
WPT Math Sub Score 3	629.12	130.582	511.16	850
WPT English	495.98	85.159	449.19	800

4. Results

The results revealed a number of interesting correlations. These correlations were the foundation for developing the prediction model. The model was then used to determine the students' predicted success in the CS 1430 course.

4.1 Tests Performed

Multiple correlation analysis was performed on the data in this study. The intercorrelation matrix can be found in Table 3. Table 3 lists the correlation coefficients (R) between all pairs of variables used in this study. The closer the R value is to 1.0 or -1.0 the stronger the correlation of the two variables. An R value of 1.0 would mean all data points fall on the same line with a positive slope. An example of this is correlating a variable to itself. CS 1430 Final Exam has an R value of 1.000 with itself. Conversely, an R value of -1.0 would mean all data points fall on the same line, but with a negative slope.

There are a number of interesting observations that can be seen in Table 3**.** For example, we see that WPT Math, WPT English, WPT Math Sub Score 1 and WPT Math Sub Score 2 all had high correlation values with the CS 143 Final Exam Score. However, because some of them also had high correlations to each other they may not offer significantly unique predictive value in the multiple correlation. For example, WPT Math and WPT Math Sub Score 1 had a high correlation with an R value of 0.627. When the WPT Math score was added into the model, the

Table 3: Intercorrelation Matrix of Coefficients for Independent and Dependent Variables

	CS 1430 Final Exam Score	High School Percentile	ACT Comp.	ACT Eng.	ACT Math	ACT Science	ACT Reading	WPT Math	WPT Math Sub Score 1	WPT Math Sub Score 2	WPT Math Sub Score 3	WPT Eng.
CS 1430 Final Exam Score	1.000	.239	.355	.274	.320	.287	.308	.501	.400	.489	.374	.495
High School Percentile	.239	1.000	.441	.295	.500	.276	.331	.525	.360	.497	.465	.349
ACT Composite	.355	.441	1.000	.821	.751	.810	.843	.578	.461	.661	.628	.782
ACT English	.274	.295	.821	1.000	.486	.514	.639	.438	.320	.498	.432	.747
ACT Math	.320	.500	.751	.486	1.000	.571	.452	.658	.566	.715	.690	.545
ACT Science	.287	.276	.810	.514	.571	1.000	.588	.486	.412	.552	.553	.567
ACT Reading	.308	.331	.843	.639	.452	.588	1.000	.351	.263	.429	.419	.690
WPT Math	.501	.525	.578	.438	.658	.486	.351	1.000	.627	.832	.801	.529
WPT Math Sub Score 1	.400	.360	.461	.320	.566	.412	.263	.627	1.000	.687	.689	.497
WPT Math Sub Score 2	.489	.497	.661	.498	.715	.552	.429	.832	.687	1.000	.828	.595
WPT Math Sub Score 3	.374	.465	.628	.432	.690	.553	.419	.801	.689	.828	1.000	.546
WPT English	.495	.349	.782	.747	.545	.567	.690	.529	.497	.595	.546	1.000

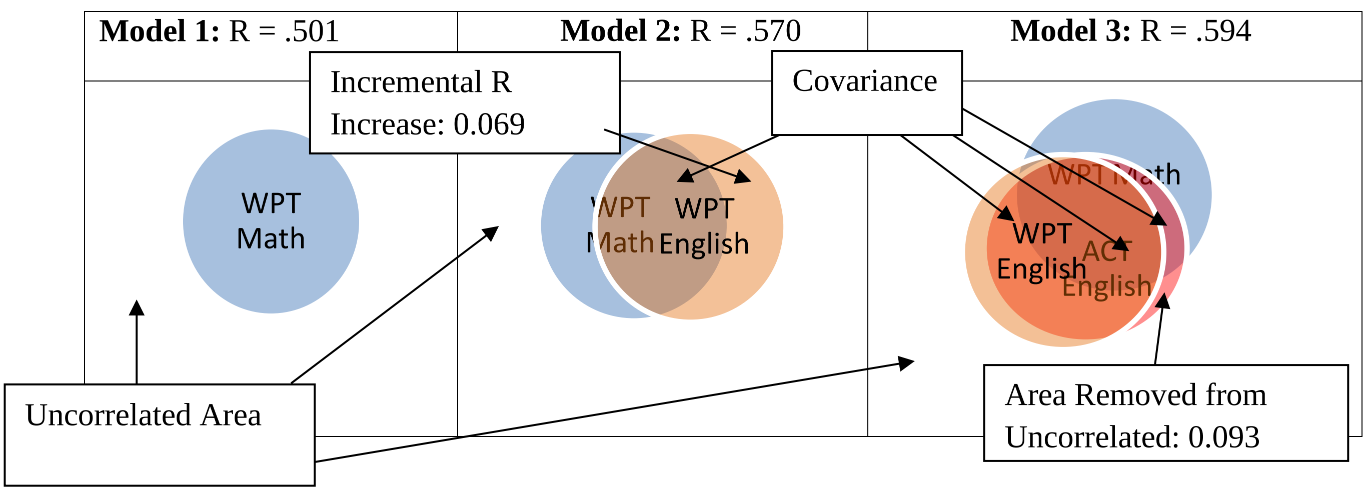
WPT Math Sub Score 1 may not have offered much unique additional predictive power for the model. Similarly, High School Percentile and CS 1430 Final Exam have a relatively low correlation with an R value of 0.239. Therefore, High School Percentile may not offer any added predictive power with the independent variables that were already in the model.

Table 3 also shows that WPT Math Sub Score 2 had a much higher correlation to the CS 1430 Final Exam Score than WPT Math Sub Score 1 or WPT Math Sub Score 3. The subject matter for WPT Math Sub Score 2 is Algebra. It is not clear why the correlation to WPT Math Sub Score 2 is greater than the other two Sub Scores. This large disparity warrants further study into the correlation to WPT Math Sub Score 2.

Another notable observation is the correlation between CS 1430 Final Exam Score and ACT English in comparison to the correlation between CS 1430 Final Exam Score and WPT English. Both the ACT English and the WPT English test the student's ability on very similar subject matter. This is evident with a very high correlation value of R = .747, but ACT English has a much lower correlation to CS 1430 Final Exam Score with R = .274 as compared to the WPT English correlation to CS 1430 Final Exam Score with R = .495. Although the ACT English and the WPT English have a high correlation to each other, the part of the each variable that is not correlated provides unique prediction to the step-wise multiple correlation.

Using the data from Table 3, a prediction model was produced using step-wise multiple linear regression analysis. The steps are shown in Table 4. Multiple linear regression analysis assesses additive effects of the independent variables as they affect the proportion of variance, R^2, in the dependent variable. Using step-wise linear regression analysis we can see that the independent variable that added the most unique predictive power was WPT Math. In Table 4, the first iteration of the prediction model, Model 1, the independent variable WPT Math was added yielding an R of .501. The R value for this first step is equivalent to the correlation between WPT Math and CS 1430 Final Exam Score found in Table 3. In Diagram 1: Model 1, this is illustrated by a Venn diagram depicting WPT Math as the single best predictive variable.

Diagram 1: Step-Wise Multiple Linear Regression Venn Diagram Representation



After step 1, the stepwise multiple linear regression determines which independent variable has the most unique predictive power. An independent variable has the most unique predictive power if it has the least covariance with the independent variables currently in the model and has the highest additive effect on the multiple correlation.

The second model, Model 2, found that the independent variable that added the most unique predictive power of CS 1430 Final Exam Score and the least covariance with WPT Math was WPT English. WPT English had an additive effect on R^2 of 0.074 which increased the multiple correlation value by 0.069 to 0.570. This is illustrated in Diagram 1: Model 2. The covariance between WPT Math and WPT English is represented by the overlapped area and was subtracted to leave the incremental R increase. The incremental R increase of 0.069 is represented by the additional area.

The third model, Model 3, found that the independent variable that added the most unique predictive power was ACT English. ACT English had a negative coefficient causing its relationship to have inverse effect. In effect, by adding the ACT English score the total area was reduced by 0.093 which increased the predictive power of the first two predictors, WPT Math and WPT English, had on the R value.

The stepwise regression halted when the next added variable did not have an additive effect on the statistical significance of the R^2 value. The R^2 value could not be improved upon any further than the three predictor variables found in Model 3. If a fourth predictor was added, the standard error of estimate would have increased, causing the significance of the fourth model to have been lower than that of Model 3. Using Model 3, we explained 35.3% (R^2) of the variance in the CS 1430 Final Exam Scores when the student's WPT Math Score, WPT English Score and ACT English were available. The final model summary can be found in Table 4.

Table 4: Stepwise Multiple Regression

Model Step	R	R^2	Std. Error of the Estimate	Variables
1	.501	.251	5.926	Predictors: WPT Math
2	.570	.325	5.656	Predictors: WPT Math, WPT English
3	.594	.353	5.567	Predictors: WPT Math, WPT English, ACT English

We also applied analysis of variance to the three models. This is illustrated in Table 5. We found that each model had a statistically significant relationship (beyond the p = 0.001 level). That is, there is less than one chance in a thousand that this could have happened by coincidence. For instance, the significance of the F-test in Model 1 was 0.0000152612 which is considered significant. In Table 5, we can see that each model added additional precision up to Model 3. This is why step-wise regression did not halt after Model 1. Step-wise regression was halted after Model 3 because adding an additional independent variable would have decreased the significance.

Table 5: Analysis of Variance of Dependent Variable: CS 1430 Final Exam Score

Model		Sum of Squares	Df	Mean Square	F	Significance
1	Regression	1180.053	1	1180.053	33.598	0.0000152612
	Residual	3512.261	100	35.123		
	Total	4692.314	101			
2	Regression	1525.612	2	762.806	23.847	0.0000000035
	Residual	3166.702	99	31.987		
	Total	4692.314	101			
3	Regression	1655.554	3	551.851	17.809	0.0000000026
	Residual	3036.760	98	30.987		
	Total	4692.314	101			

4.2 Prediction Model

The regression equation is a linear equation written in slope intercept form, $y = b_0 + b_1x_1 + \ldots + b_nx_n$. Linear regression, using a linear equation, is used to find the variable or variable sets that best predict a particular variable. In our prediction model, b_0 represents the zero intercept of the predicted value, $x_1 \ldots x_n$ represent the values of the predictor variables, and $b_1 \ldots b_n$ represents the coefficients of the respective predictor variables. One has to solve for b_0 through b_n to find the best predictor linear equation that minimizes the squared deviations between the sets of predicted and actual dependent values. The linear equation for each model can be found in Table 6.

Table 6: Prediction Models

Model		Regression Equations
1	WPT Math	CS 1430 Final Exam Predicted Score = 27.571 + 0.282 * WPT Math
2	WPT English WPT Math	CS 1430 Final Exam Predicted Score = 17.628 + 0.187 * WPT Math + 0.026 * WPT English
3	ACT English WPT English WPT Math	CS 1430 Final Exam Predicted Score = 19.411 + 0.196 * WPT Math + 0.040 * WPT English + (-0.394) * ACT English

`

As an example, we can use the Model 3 equation to predict a CS 1430 final exam score. Substituting the average scores from Table 2 (WPT Math: 28.87, WPT English: 495.98, ACT English: 23.32), we obtain

CS 1430 Final Exam Predicted Score
= 19.411 + (0.196 * WPT Math) + (0.040* WPT English) + (-0.394*ACT English)
= 19.411 + (0.196 * 28.87) + (0.040* 495.98) + (-0.394*23.32)
= 19.411 + 5.65852 + 19.8392 + (-9.18808)
= 35.72064

The result of 35.72 is within 0.01 of the actual CS 1430 Final Exam score average shown in Table 2. The WPT Math and WPT English had positive coefficients, while ACT English had a negative coefficient. This meant that a lower ACT English score would have resulted in a higher CS 1430 Final Exam score. To illustrate this we can use the values one standard deviation from the means: 40.973 for WPT Math, 581.139 for WPT English and 18.976 ACT English. Using these values would result in a predicted CS 1430 Final Exam score of 43.210724 which is within 0.665 of one standard deviation of the actual CS 1430 Final Exam score.

5. Discussion and Future Research

This study is consistent with previous studies examining success rate in CS1 courses in finding that mathematical ability is the single best predictor of success as defined by the given study. Such results have lead to experimenting with math requirements for CS1. However, this study shows that using other predictors in addition to mathematical ability better predicts the success on the final exam in the CS 1430 course at University of Wisconsin-Platteville. In addition to mathematical ability, language ability, as measured by the WPT English and ACT English exams, had a significant correlation to the student's score on the CS 1430 Final Exam. One possible explanation may be due to the problem solving process that a student goes through when answering a question on the CS 1430 Final Exam. The student must first read the exam question and fully comprehend what the question is asking in order to have a chance at answering the question correctly. This part of the process relies on the student's language comprehension ability. Once the student establishes an understanding of the question, they can then start the mathematical, analytical, thought process of actually solving the problem. Our prediction model suggests that a student with both mathematical and language ability will have the highest potential for success on the CS 1430 Final Exam.

Another factor to consider is that C++ is a language with a vocabulary and grammar. Perhaps if one possesses an aptitude and skill with a language, such as English as measured by the WPT English exam, one might also have an aptitude and skill with other languages like C++.

It was very difficult to collect complete data for all students. Due to this difficulty we were left with only subsets of data from each semester. This may have introduced some bias into the data as we have no evidence that this data was from a truly random sample of students. The goal of future work will be to somehow control the volume of missing data to determine whether the results of this study are valid and not partially impacted by the large volume of missing data.

If further research shows the sample was effectively random, the prediction model presented here effectively predicts a student's CS 1430 Final Exam score (+/- 5.567 points) 68.2% of the time. This suggests the model can be used to place students in the appropriate introductory computer science course. In addition to considering just a student's mathematics maturity, advisors should also consider a student's English skills. Perhaps freshmen English should also be considered as a prerequisite for introductory computer programming courses. This work should help to the lower the attrition rate by first qualifying students that are entering the CS 1430 course. This will allow the professors of the CS 1430 course to have a better understanding of why their students are performing the way they are and possibly adjust teaching methods to address that variance.

6. References

Black, Sarah. "Predictors of first year computing science student failure." *Maxi Project* (2003).

Bray, Nathaniel J, John M Braxton and Anna S Sullivan. "The Influence of Stress-Coping Strategies on College Student Departure Decisions." *Journal of College Student Development* (1999): 645-657.

Bruce, Christine, et al. "Ways of Experiencing the Act of Learning to Program: A Phenomenographic Study of Introductory Programming Students at University." 2004. *Journal of Information Technology Education.* 7 2011.

Cohoon, J. M. and L. Y. Chen. "Migrating out of computer science." *Computing Research News* (2003): 15.

Ford, Marilyn and Sven Venema. "Assessing the Success of an Introductory Programming Course." *Journal of Information Technology Education* (2010): 133-145.

Hagan, D and S Markham. "Does it help to have some programming experience before beginning a computing degree program?" *AC SIGCSE Bulletin* (2000): 25-28.

Hostetler, Terry R. "Predicting Student Success in an Introductory Programming Course." *ACM SIGCSE Bulletin* (1983): 40-49.

Hostetler, Terry R. "Predicting student success in an introductory programming course." *ACM SIGCSE Bulletin* 15.3 (1983).

Parsons, Ted and Rob Hasker. *CS 1430 Predictor Analysis Fall 1997*. Unpublised Study. Platteville, WI, 1998.

Pattis, Richard E. *Karel The Robot: A Gentle Introduction to the Art of Programming.* John Wiley & Sons, 1981.

Rauchas, Sarah, et al. "Language Performance at High Scool and Success in First Year Computer Science." *SIGCSE* (2006): 398-402.

Roddan, Matt. "The Determinants of Student Failure and Attritiion in First Year Computing Science." 8 June 2002. *University of Glasgow.* 9 November 2011. <http://www.psy.gla.ac.uk/~steve/localed/mrodd.html>.

Simon, et al. "Predictors of Success in a First Programming Course." 2006. *Australian Computer Society.* 2011.

Tinto, Vincent. "Dropout from Higher Education: A Theoretical Synthesis of Recent Research." *Review of Educational Research* (1975): 89-125.

United States Census Bureau. "Computer and Internet Use." 7 June 2011. *U.S. Census Bureau.* 12 Febraury 2012. <http://www.census.gov/hhes/computer/>.

University of Wisconsin Testing and Evaluation Services. *Contents of Placement Tests*. 30 July 2011. 20 November 2012. <http://testing.wisc.edu/centerpages/contentsofplacementtests.html>.

University of Wisconsin-Platteville Computer Science Department. "CS 113:Introduction to Programming Syllabus." 23 May 2012. *University of Wisconsin-Platteville.* Syllabus. 23 May 2012.

—. "CS 143: Programming in C++." 25 April 2012. *University of Wisconsin-Platteville.* 25 April 2012. <http://www.uwplatt.edu/csse/Courses/CS143/syllabus.html>.