COMPARING MULTIPLE MEASURE PLACEMENT MODELS IN MATHEMATICS

A DISSERTATION IN

Curriculum and Instruction

and

Mathematics

Presented to the Faculty of the University of
Missouri-Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by

WILLIAM PARKER MORGAN IV

B.S., Missouri State University, 2003

M.S., University of Arkansas, Fayetteville 2005

Kansas City, Missouri

2020

COMPARING MULTIPLE MEASURE PLACEMENT MODELS IN MATHEMATICS

William Parker Morgan IV, Candidate for the Doctor of Philosophy Degree

University of Missouri – Kansas City, 2020

ABSTRACT

Accurate placement into an initial college mathematics course is a key step toward the successful completion of college mathematics and, eventually, a college degree. Conversely, misplacement in mathematics may lead to a reduced likelihood of course completion and degree attainment. This study investigated the ability of two placement models to predict course success in mathematics: a hierarchical placement model utilizing cut point values determined by CHAID decision trees and an algorithmic placement model utilizing the regression equations from binary logistic regression analysis. Both models used the placement measures of high school grade point average (HSGPA) and ACT-Mathematics (ACTM) scores to predict the placement of students from a large, Midwest urban/suburban community college. The accuracy of the two placement models was compared to determine which, if either, placement model more accurately placed students in their initial college mathematics course. The ability of each placement model to discern between a successful student and an unsuccessful student was also compared. The analysis showed that the two placement models perform equally well in predicting course success and in discerning between a successful and unsuccessful student. The algorithmic model tended to underplace students by placing substantially more students in the lowest mathematics course level than

the hierarchical placement model with only minimal improvement in predicting success. These results, along with the likelihood that the placement rules of the hierarchical model would be more easily understood by students and advisors, make the hierarchical model more useful for college mathematics placement.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies, have examined a dissertation titled "Comparing Multiple Measure Placement Models in Mathematics," presented by William Parker Morgan IV, candidate for the Doctor of Philosophy degree, and certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Rita Barger, Ph.D., Committee Chair
Division of Teacher Education and Curriculum Studies

Noah Rhee, Ph.D.
Department of Mathematics and Statistics

Eric Hall, Ph.D.
Department of Mathematics and Statistics

Raol J. Taft, Ph.D.
Division of Teacher Education and Curriculum Studies

Seung-Lark Lim, Ph.D.
Department of Psychology

CONTENTS

TABLES

ILLUSTRATIONS

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my committee members, Rita Barger, Noah Rhee, Eric Hall, Raol Taft, and Seung-Lark Lim for their time and support. I would especially like to thank my committee chair, Rita Barger, for her countless hours of support, thoughtful suggestions, and continued belief and confidence in me. Dr. Barger's endless enthusiasm for mathematics education has been an inspiration throughout this process. Thank you to Metropolitan Community College for providing financial support and to my colleagues at Metropolitan Community College for the encouragement to pursue my dream.

Most importantly, I would like to thank my family, without whom I would not have aspired to complete this degree. To my father, Parker, and mother, Donna, thank you for the passion for education and your fervent belief in me. To my sister, Kelly, thank you for always being proud of me. To my wife, Crysten, there are not enough words to thank you for your unwavering love and support during this arduous journey. We made it through together, as we will always be. To my daughters, Avery, Allie, and Annabelle, I love you and I hope that you are inspired to do great things. Finally, thank you to my Lord and Savior, Jesus Christ, who has given me strength.

CHAPTER ONE

INTRODUCTION

Background Information

Embedded in a larger effort to increase degree attainment, retention rates, and success in college mathematics courses, alternative methods to place students into their initial mathematics course are being recommended by researchers, state agencies, and national organizations (Belfield & Crosta, 2012; Reeves Bracco et al., 2014; Burdman, 2012; Hughes & Scott-Clayton, 2011; Maruyama, 2012; Missouri Department of Higher Education, n.d.; Ngo & Kwon, 2015; Scott-Clayton, Crosta, & Belfield, 2014). The national organization, WestEd, for example, asserts that "states must examine the predictive validity of current placement processes and determine whether those processes are providing results that help their students succeed" (Reeves Bracco et al., 2014, p. 39). The report recommends using multiple measures to place students (p. 6) and defines multiple measures placement as "the use of more than one measure to determine student placement into college-level courses. Common multiple measures include, but are not limited to, additional test scores (beyond a single standardized test score), high school grade point average (GPA), high school grades in specific classes, life experiences, and counselor input and referrals" (Reeves Bracco et al., 2014, p. ii). A joint statement by the Charles A. Dana Center, Achieving the Dream, and Jobs for the Future recommend using "multiple factors—such as a combination of career and academic goals, non-cognitive assessments, high school transcripts, and assessment scores—to determine whether students are placed into developmental courses and to determine which developmental or gateway courses are most appropriate" (Couturier & Cullinane, 2015, p. 9).

Researchers have found that utilizing a multiple measure placement model with cumulative high school grade point average (HSGPA) as one of the measures more accurately places students and increases the likelihood of completing the Gateway mathematics course and, ultimately, a college degree (Ngo & Kwon, 2015; Scott-Clayton, 2012; Scott-Clayton et al., 2014). Even the companies that own the commonly used standardized placement tests recommend using multiple measures to place students (ACT, 2014, p. 100; College Board, 2016b, p. 93).

A few multiple measure placement models have been implemented by colleges and universities in the United States including hierarchical models and algorithmic models (Barnett et al., 2016; Reeves Bracco et al., 2014). However, little research has compared the different models to determine which is more accurate or places more students in higher level mathematics courses. By reproducing the multiple measure placement models currently in use so that placement accuracy is maximized for each model according to a specific sample, placement models can be compared and analyzed. Knowing which placement model more accurately places students and the percent of students placed at each course level may help policymakers determine which model most closely aligns with the goals of the college. Insights into practical considerations demonstrated in the study such as ease or difficulty of implementation will be included in the discussion.

The Problem

Most community colleges use either the Compass or ACCUPLACER standardized placement test as the sole means of placing non-exempt students into their initial mathematics course (Fields & Parsad, 2012; Hughes & Scott-Clayton, 2011; Parsad et al., 2003). While research by the ACT Corporation and the College Board, owners of the Compass and ACCUPLACER, respectively, show that the standardized placement tests are accurate between 60 and 80 percent of the time (Belfield & Crosta, 2012; Scott-Clayton, 2012), other research shows that these standardized placement tests have little or no correlation to success in college (Roska et al., 2009; Mattern & Packman, 2009; Armstrong, 2000) or are redundant measures to the ACT or SAT tests. Medhanie et al. (2012) found that ACCUPLACER added no significant value to the predictive placement design once a student's ACT score was statistically controlled. While disagreement about the predictive validity of standardized placement tests exists, most research has shown that success in high school, as given by a student's cumulative high school grade point average, is a good predictor of success in college (Bahr et al., 2014; Geiser & Santelices, 2007; Medhanie et al., 2012; Ngo & Kwon, 2015; Rothstein, 2004; Scott-Clayton, 2012).

Research has also shown that using multiple measures such as data from high school transcripts in combination with an assessment such as the ACT or SAT or a standardized placement test such as ACCUPLACER or Compass may be the best predictor for college success (Bahr et al., 2014; Belfield & Crosta, 2012; Lewallen, 1994; Ngo & Kwon, 2015; Scott-Clayton, et al., 2014). Working for the ACT Corporation, Westrick and Allen (2014) found that multiple measures, including the use of cumulative high school grade point average, better predicted the success of first-year college courses than any measure used alone. Recently, the ACT Corporation took matters one step further by announcing that they eliminated the Compass placement exam on December 31, 2016 due to a lack of correlation with college success. According to the ACT website (ACT, n.d.),

> A thorough analysis of customer feedback, empirical evidence and postsecondary trends led us to conclude that ACT Compass is not contributing as effectively to student placement and success as it had in the past. Based on this analysis, as part of ACT's commitment to continuous improvement, we have made the difficult decision to phase out all the ACT Compass products (Para. 1).

Empirical data on the low predictability of standardized placement exams and the greater predictability of multiple measures has caused state agencies and national organizations to recommend changes to placement policy. The Missouri Department of Higher Education's Principles of Best Practices states that "placement of students into appropriate college-level courses must be based on multiple assessment measures, which provide a more precise measurement of a student's ability to succeed in college-level coursework" (Missouri Department of Higher Education, n.d., p. 7). The national organizations WestEd, Achieving the Dream, The Charles A. Dana Center, and Jobs for the Future endorse the use of multiple measures for placement to provide a complete understanding of a student's ability (Reeves Bracco et al., 2014; Couturier & Cullinane, 2015). However, colleges appear to be slow in adopting multiple measure placement models (Fields & Parsad, 2012; Hughes & Scott-Clayton, 2011).

Scott-Clayton (2012), in summarizing her 2011 work with Hughes, states that "for non-exempt students, few schools nationally appear to use multiple measures in a systematic way, perhaps because of uncertainty regarding how to collect this information efficiently or how to combine it into a simple and scalable placement algorithm" (p. 12). Non-exempt students are defined in the study as students placed using placement tests due to not achieving a high enough score on another measure such as ACT or SAT to be placed in credit-bearing courses. Without knowledge of what effect a change in placement will cause in terms of the number of students placing into each course and the success rates in those courses, enacting change may be difficult, especially when the basis of the placement design changes from a model based on a standardized placement test to a multiple measures model. Knowing the approximate effects of a new placement model may give insight to those charged with making placement decisions and allow them to make data-driven decisions on whether a change in placement will benefit the college and align with the goals of the college.

The Purpose of the Study

This study attempts to compare two multiple measure placement models, an algorithmic placement model and a hierarchical placement model, to determine which model has greater predictive ability. The goal is to provide colleges and universities considering implementing or amending multiple measure placement models a basis of research on the performance of two such models. Each will be created for five college mathematics course levels using high school grade point average (HSGPA) and ACT-Math scores as placement measures based on the historical data (initial sample) from one large community college. An algorithmic placement model is a collection of regression equations, one for each course level, that statistically determine whether a student should be placed into that course based on whether the value of the dependent variable is above or below a predetermined cut point. The independent variables of each regression equation are the placement measures (ACT-Math and HSGPA) and the coefficient of each measure is statistically determined according to the fit of the initial data sample. The algorithmic model for this study will be created using binary logistic regression equations for each course level. Binary logistic regression allows for the dependent variable to be binary (predicted success or predicted failure) instead of scaled.

A hierarchical placement model places students into a mathematics course level by comparing the score on a placement measure or a combination of scores on multiple placement measures to predetermined cut points. The cut points for the hierarchical placement model will be statistically determined using the CHAID method of decision tree modeling using data from the initial sample. The CHAID method of decision tree modeling groups elements of an independent variable (a placement measure) based on similarity of the dependent variable (success in a mathematics course).

Once the algorithmic and hierarchical placement models are designed, the predictive ability of the models will be compared based on accuracy, sensitivity, specificity, and discrimination for both the initial historical sample and the validation sample. Accuracy is the percent of correctly predicted outcomes (True Positives + True Negatives) / (Total Outcomes). In terms of the study this is (Correctly Predicted Successes + Correctly Predicted Failures) / (Total Outcomes). Sensitivity is the percent of correctly predicted positive outcomes (True Positives) / (Total Positives). In terms of the study this is (Correctly Predicted Successes) / (Correctly Predicted Successes + Predicted Failures but Actual Successes) or (Correctly Predicted Successes) / (Total Actual Successes). Specificity is the percent of correctly predicted negative outcomes (True Negatives) / (Total Negatives). In terms of the study this is (Correctly Predicted Failures) / (Correctly Predicted Failures + Predicted Successes but Actual Failures) or (Correctly Predicted Failures) / (Total Actual Failures). Discrimination is given by the area under the Receiver Operator Characteristic curve (ROC curve) and interpreted as the ability of a model to differentiate between the success and failure of a pair of opposite outcomes, or in terms of this study, opposite performing students. The Receiver Operator Characteristic (ROC) curve plots the Sensitivity versus 1 – Specificity for each value of the probability cut point. In other words, it plots True Positives versus False Positives (also known as Type I error).

In addition to analyzing the predictive ability of the placement models, placement into each course will be determined for each model and compared to determine how much the models agree on the placement of students in the validation sample. A difference in placement will be statistically evaluated to determine if the amount of disagreement is statistically significant. Finally, the strength of HSGPA and ACT-Math on course success in the algorithmic model will be determined to add to the existing research on HSGPA, ACT-Math, and course success.

This study follows up on Scott-Clayton's (2012) suggestion that colleges need further research on simple, scalable placement algorithms. It also addresses Maruyama's (2012) call that "future research should examine in greater detail whether or not grades can be made a stronger determinant of college readiness" (p. 259) with an emphasis on using high school grades as one of the college readiness indicators. The specific goals of the study are:

- To create a hierarchical multiple measure placement model for five levels of mathematics courses based on historical data using the CHAID decision tree method for determining cut point values for each measure
- To create an algorithmic multiple measure placement model for five levels of mathematics courses based on historical data using binary logistic regression

- To determine the fit, accuracy, specificity, sensitivity, and discrimination of the algorithmic model of course success using the initial and validation samples
- To determine the accuracy, specificity, sensitivity, and discrimination of the hierarchical model of course success using the initial and validation samples
- To compare the predictive ability of the hierarchical and algorithmic model of course success using the initial and validation samples
- To compare the level of agreement of course placement of students between the models using the validation sample
- To determine the significance of each predictor variable (HSGPA and ACTM) on course success of the algorithmic model at each course level

Names of mathematics courses are not consistent among colleges and multiple courses exist at the same placement level. Thus, generic names are given for each level of mathematics placement. The credit-bearing mathematics course levels for which data will be analyzed are Gateway, Trigonometry, and Calculus. Gateway courses often include College Algebra, Statistics, and Quantitative Reasoning courses. The developmental mathematics courses for which data will be analyzed are Developmental Math 1 (DM1) and Developmental Math 2 (DM2) in which DM1 is one level below the Gateway course level and Developmental Math 2 (DM2) is two levels below the Gateway course level.

Placement percentages are defined as the percent of students placing into each level of mathematics. Course success is defined as a student achieving a C or better in the course or a satisfactory grade for courses graded as satisfactory/unsatisfactory. A grade of C or better was chosen instead of a grade of D or better because some courses in the study require a C or better to progress to the next course in the sequence. Thus, a D in this study is considered unsuccessful, even though it may satisfy the requirement for a terminal course. When analyzing placement designs, both the success rate and the percentage of students placing at each level of mathematics are of practical importance in assessing effectiveness. For example, if the success rate for Gateway courses is high but the percentage of students placing in Gateway courses as compared to developmental courses is low then the placement model may not be of great practical use.

This study is unique in that it analyzes multiple measure placement models for several levels of mathematics courses. Current research is often limited to placement into the initial college credit course only which we have designated the Gateway level. As an example, the North Carolina Community College System (NCCCS) uses a hierarchical multiple measure placement model for placement into college level courses but relies on a modified version of the ACCUPLACER exam for placement into developmental mathematics courses and higher-level mathematics courses (North Carolina Community College System, 2016). Similarly, Long Beach City College (LBCC) in California is piloting an algorithmic multiple measure placement model for a certain population of incoming students for placement into their Gateway mathematics course. Once again, placement into other courses uses standardized test scores (Long Beach City College, 2016). The scalability of multiple measure placement models may benefit from using the model for all levels of mathematics courses. This study is also unique in that it statistically compares two placement models.

The research questions follow the goals of the study in analyzing and comparing placement and success for the algorithmic and hierarchical models and evaluating the predictive ability of the models and the strength of the predictor variables.

Research Questions

1. What is the predictive ability of the hierarchical multiple measures placement model in terms of accuracy, sensitivity, specificity, and discrimination using the initial and validation samples?
2. What is the predictive ability of the algorithmic multiple measures placement model in terms of measures of fit, accuracy, sensitivity, specificity, and discrimination using the initial and validation samples?
3. What are the differences in predictive ability between the hierarchical and algorithmic placement models using the initial and validation samples?
4. What are the differences in course placement results between the hierarchical and algorithmic placement models using the validation sample?
5. What are the predictive strengths of HSGPA and ACTM scores on student success at each course level for the algorithmic model using the initial sample?

The choice to use ACTM and HSGPA as the measures of the placement design to be researched is due to both measures showing at least some predictive ability to course success in prior research (ACT, 2008; Barnett et al., 2016; Bettinger et al., 2011; Ngo & Kwon, 2015; Noble & Sawyer, 2004; Sawyer, 2013; Scott-Clayton et al., 2014) and because both measures are common data points on students' records. Nationally, 78% of college students enter college at age 20 or less indicating that they are either recent high school graduates and thus have a HSGPA or have recently completed a high school equivalency exam (Shapiro et al., 2013). In Missouri, almost 8500 people successfully completed a high school equivalency test in 2014 (Educational Testing Service, 2015, p. 11) compared to almost 65,000 students graduating high school (Ruffalo Noel Levitz, 2014, p. 9), thus demonstrating that most entering college students gained post-secondary access by completing high school and have a recent HSGPA. In addition, in 2014, 76% of Missouri high school graduates took the ACT exam (ACT, 2016). The research concerning the predictive ability of HSGPA and ACT scores, along with other measures commonly used to place students, is presented in the first section of the theoretical framework.

Theoretical Framework

The theoretical framework will support the study by drawing attention to empirical scholarship on the relationship between measures of placement and student success in college. It will also review research on the underlying non-cognitive, or affective, aspect of HSGPA as a possible reason behind its predictive strength, developmental coursework as a benefit or barrier to success in college, and multiple measure placement models currently used at colleges in the United States. The statistical aspects of placement, the correlation among measures and the odds ratios of measures, are intertwined with the social aspect of placement including barriers to success and motivation. Jabareen (2009) defines a theoretical, or conceptual, framework as “a network, or ‘a plane,’ of interlinked concepts that together provide a comprehensive understanding of a phenomenon or phenomena” (p. 51). The researcher hopes to provide a comprehensive understanding of the components of a multiple measure placement design in the theoretical framework by interweaving the statistical aspects with the social aspects.

The first theme of the theoretical framework addresses the research surrounding the validity of standardized placement exams and HSGPA as predictors of different measures of college success such as course completion, first year grade point average (FYGPA), and graduation rates. This strand looks at prominent studies that have analyzed the correlation or prediction of these measures to college success and compares these two placement measures to other commonly used placement measures. The first theme concludes with an analysis of the research suggesting that a non-cognitive, or affective, aspect is inherent in HSGPA and may explain some of its predictability for success in college. If high school grade point average is correlated to affective measures, including it in the placement model may bring in effects of personality such as academic motivation (Duckworth et al., 2007; Duckworth et al., 2012; Robbins et al., 2004; Sedlacek, 2004).

Since placement design influences the percent of students placing into developmental courses and at which course-level within the developmental sequence a student begins, the second theme of the theoretical framework analyzes the benefits and barriers of developmental education to success in college. While developmental coursework is intended to prepare students to be successful in a credit-bearing course, the process of taking additional coursework may unintentionally hinder student progress (Attewell et al., 2006; Bailey et al., 2010; Complete College America, 2012; Martorell & McFarland, 2011; Scott-Clayton & Rodriguez, 2012). The third theme of the literature review investigates multiple measure placement models currently used at colleges in the United States, specifically, algorithmic and hierarchical models. Included in the theme is the application and functionality of the CHAID method of decision tree modeling and the application and functionality of binary logistic regression as a predictive algorithm.

The Predictive Validity of the ACT Test, High School Grade Point Average, and Other Measures of Success in College

Cumulative high school grade point average (HSGPA) and scores on the ACT-Mathematics test (ACTM) are the foundational elements of the placement models to be studied. It follows that an analysis of the predictive validity of these two measures along with the predictive validity of other measures that could have been used instead of or in addition to these two measures should occur. The measures upon which the placement models will be constructed were chosen based on the literature presented in the literature review. Highlighting that literature is a validity report by the ACT Corporation comparing the predictive validity of ACT scores and high school grade point average to several outcomes of college success including first year grade point average (FYGPA), grades in college mathematics courses, final college grade point average, retention status, and level of degree attainment. The validity report shows that HSGPA is a better predictor than ACT scores of FYGPA, $r = .48$ and $r = .42$, respectively, and final college grade point average, $r = .35$ and $r = .19$, respectively. The ACT test, however, was a better predictor than HSGPA for retention status ($r = .25$ to $r = .22$), collegiate mathematics proficiency ($r = .57$ to $r = .14$), and level of degree attained ($r = .16$ to $r = .08$) (ACT, 2008). Thus, while both HSGPA and ACT moderately predict multiple outcomes of college success, each has predictive strengths that the other lacks. The individual predictive validity of ACT and HSGPA found in the ACT study agrees with other empirical studies (Barnett et al., 2016; Bettinger et al., 2011; Ngo &

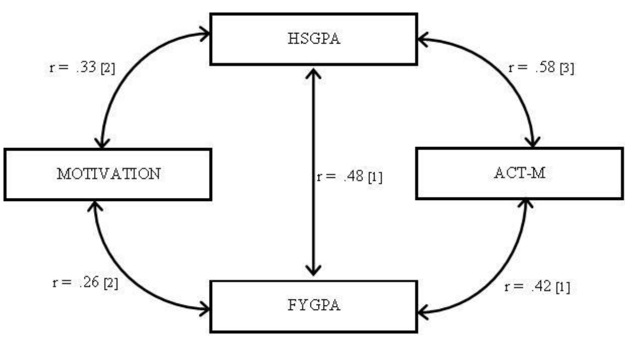
Kwon, 2015; Noble & Sawyer, 2004; Sawyer, 2013; Scott-Clayton et al., 2014). The study by the ACT Corporation is highlighted here because it demonstrates the correlation of HSGPA and the ACT test to several outcomes of college success. Since the ACT test and HSGPA have differing strengths of predictability for various measures of college success, combining the two measures in a multiple measure approach to placement would be assumed to better predict success than either measure alone. This assumption is confirmed by the same research and by other research showing additional predictive ability of both measures combined over the use of HSGPA or ACT alone (ACT, 2008; DesJardins & Lindsay, 2008; Ngo & Kwon, 2015).

When considering the use of multiple measures for placement, high school grades are usually the first suggested measure (Burdman, 2012) and are the single greatest measure of predicting college success in math and English (Barnett et al., 2016). According to Scott-Clayton et al. (2014), placement using HSGPA instead of standardized tests significantly reduces the rate of severe placement errors in developmental mathematics. Standardized test scores and HSGPA used jointly may increase placement accuracy by better determining a student’s preparedness for college (DesJardins & Lindsay, 2008; Ngo & Kwon, 2015; Noble and Sawyer, 2004). However, using multiple standardized tests for placement is likely a redundant action. Medhanie et al. (2012) found little or no significant predictive value to using the ACCUPLACER placement test above the predictive ability of the ACT test on its own. For these reasons a standardized placement exam, such as ACCUPLACER, was not included as a measure in the placement models to be studied. Other measures such as the highest level of mathematics completed in high school were considered for the study but were not included as measures because a cursory look at historical data at the college in which the study is to occur revealed that information on high school courses completed was not available for a large portion of students. Note that at Long Beach City College, the level of last high school mathematics course completed was moderately predictive of college course success in the logistic regression models for seven of the eleven colleges. For comparison, HSGPA was moderately predictive of college success in the logistic regression models for five colleges, highly predictive for one college, and weakly predictive for one college (Willett, 2013). Thus, the level of last high school mathematics course completed might be a significant measure for placing students. However, only HSGPA and ACTM were both rich enough in available data and shown to be effective predictors in prior literature and thus only these two measures were used in the creation and analysis of the algorithmic and hierarchical multiple measure placement models.

The Non-Cognitive Aspect of High School Grade Point Average

One theory as to why high school grade point average has been shown to be a more accurate predictor of first year college success than other measures (Ngo & Kwon, 2015; Scott-Clayton et al., 2014) is the possibility of non-cognitive or affective aspects inherent within HSGPA such as motivation, persistence, time-management, and social support (Duckworth et al., 2007; Duckworth et al., 2012; Sedlacek, 2004). Porchea et al. (2010) studied the influence of six psychosocial predictors on college degree attainment for community college students: Academic Discipline, Academic Self-Confidence, Commitment to College, Steadiness, Social Activity, and Social Connection. Only Academic Discipline and Commitment to College were positive indicators in predicting an earned degree or certificate. "Our expectation that students with greater motivation (i.e., Academic Discipline and Commitment to College) would be more likely to obtain a degree and transfer was met" (p. 700). Robbins et al. (2004) found that motivation is more correlated to HSGPA, r = .33, than ACT/SAT scores, r = .14. Thus, students who have a higher HSGPA may have more intrinsic motivation compared to students scoring similarly on a standardized test with a lower HSGPA. This noncognitive trait may help students achieve greater success in college (see Figure 1). As such, using HSGPA as a placement measure may show evidence of a student's non-cognitive abilities or support structures and better predict future success (ACT, 2014; Allen et al., 2008; Camara & Echternacht, 2000; Geiser & Studley, 2002; Maruyama, 2012; Noble & Sawyer, 2004; Robbins et al., 2006; Sanchez, 2013; Sawyer, 2013; Schmitt, 2012; Westrick & Allen, 2014). Additionally, using HSGPA to place students will allow students with higher HSGPA's to gain access to Gateway mathematics courses or higher-level developmental mathematics courses so that they have fewer courses to complete in the developmental sequence.

Figure 1: Correlations Among Academic and Noncognitive Measures



[1] ACT (2008)
[2] Robbins et al. (2004)
[3] Westrick et al. (2015)

Developmental Education as a Benefit or Barrier to College Success

The accuracy of a placement design is important to community colleges for many reasons including the relationship of placement to developmental education. Unnecessarily placing students into developmental education can be costly for students and the college if developmental education becomes a barrier to students' college success. The most underprepared students (and hence those who test into developmental coursework) achieve a college degree at a lower rate than those not needing developmental courses or those needing fewer developmental courses (Attewell et al., 2006; Bailey et al., 2010; Edgecombe, 2011; Perry et al., 2010; Scott-Clayton & Rodriguez, 2012; Scott-Clayton et al., 2014), but the reasons why might not be obvious. While the success rates for developmental courses are slightly greater than the success rates for credit-level courses at community colleges (Gerlaugh et al., 2007), many students who pass developmental courses do not return for the next course in the sequence. This is known as the stop-out effect because students stop attending or the exit-point effect because the end of each semester is a possible point for exiting the sequence of courses. About 50% of students who complete their developmental mathematics sequence and enroll in College Algebra pass the course. However, only about 65% of students who complete their developmental education sequence ever enroll in College Algebra (Bailey et al., 2010). Thus, over a third of developmental students stop-out, or exit, the sequence just prior to the Gateway course despite successfully completing the required developmental coursework.

Secondly, there is a compounding effect of success rates in the developmental course sequence that makes the probability of a student completing all courses in the sequence plus the Gateway course minute, especially for students needing to complete several courses in the developmental education sequence. Baker et al. (2014) found success rates of 65%, 55%, and 45% for DM 2, DM1, and Gateway courses, respectfully. Thus, for a student beginning in a course two levels below Gateway, the probability of completing the Gateway course is about 16% if the student does not stop-out. Students who begin their collegiate mathematics journey at a course three levels below the Gateway course have even less of a chance of completing the entire sequence. MCC data for 2012 showed that 2% of students who began mathematics coursework three levels below Gateway completed a Gateway course within three years and 12% of students who began two levels below Gateway completed a Gateway course within three years (Metropolitan Community College, 2016).

Developmental courses may be useful for some students to catch up on mathematical knowledge and lead some students on a successful path to degree completion (Bettinger & Long, 2009), but the barriers inherent within the developmental sequence should be considered when placing students. Scott-Clayton et al. (2014) found that about 25% of students were severely misplaced using standardized placement tests alone with about 20% of students severely underplaced. According to this research, many more students could and should be placed into higher level coursework. The study further states that utilizing high school grade point average as one of the placement measures increases the accuracy of placement and, depending on the chosen design, could place more students in higher-level mathematics courses without decreasing the rate of success.

Current Multiple-Measure Placement Models

The most common method of placement in mathematics for community colleges is to use an SAT or ACT score as a determination of college readiness and then to give students the Compass, ACCUPLACER, ALEKS, or institution-created test (Parsad et al., 2003) to determine placement above the Gateway level and below the Gateway level. Models using more than just a standardized test or a series of standardized tests, however, are being used at some colleges including the North Carolina Community College System (NCCCS), New Mexico State University, and Long Beach City College (Barnett et al., 2016). The North Carolina Community College System consists of all community colleges in the state of North Carolina. Every college within the system uses a hierarchical system of determining placement into their Gateway math course. A student will be placed into the Gateway course if any of the following occur: 1) A student has prior college credit for a prerequisite course, 2) A student has a HSGPA of 2.6 or greater and has passed a college prep math in high school as the fourth math credit, 3) A student has an ACT/SAT at national benchmark, or 4) A student scored at benchmark on the North Carolina Diagnostic and Placement test (Barnett et al., 2016).

New Mexico State University uses a matrix placement model in which a combination of high school grade point average and ACTM score determines their initial placement in mathematics (New Mexico State University, n.d.). Students can challenge their placement by taking an institution-created exam. The placement matrix was created using historical data from New Mexico State University and was the inspiration for the organizational design of Figure 2. The placement matrix is a useful organizational and visual aid for data but is essentially a version of the hierarchical placement model using “and” and “or” logic statements.

Figure 2: A Matrix of ACTM scores and HSGPA

	High School Grade Point Average								
ACT-Math	<2.00	[2.00, 2.25)	[2.25, 2.50)	[2.50, 2.75)	[2.75, 3.00)	[3.00, 3.25)	[3.25, 3.50)	[3.50, 3.75)	>= 3.75
0-13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									
28-36									

Long Beach City College (LBCC) uses a binary logistic regression algorithm to place students into the Gateway mathematics course using overall HSGPA, high school math GPA, ACT/SAT-Math score, score on the California State Test (for high school juniors), and highest high school math course as predictor variables (Barnett et al., 2016). LBCC and NCCCS only use their respective multiple measure placement models for Gateway course placement. Other course placement is accomplished through more traditional methods (Barnett et al., 2016; Long Beach City College, 2016). The placement process into mathematics courses other than the Gateway (or transfer-level) course was not evident for Long Beach City College or the North Carolina Community College System. Additionally, LBCC only uses the multiple measure placement algorithm for about 1500 students each fall semester from certain school districts (Long Beach City College, 2016).

The CHAID Method of Decision Tree Modeling

Decision tree modeling offers "a non-algebraic method for partitioning data that lends itself to graphical displays" (Wilkinson, 1992, p. 10) by identifying and grouping elements of an independent variable based on similarity of the dependent variable (Hastie et al., 2009). Categorical or ordinal data groups that are similar on the dependent variable are merged together through an algorithmic process. These "optimally" merged groups are significantly different from one another on the dependent variable and referred to as nodes (Ramaswami & Bhaskaran, 2010). The process is repeated within each node to form a collection of nodes that are significantly different from their sister nodes on the dependent variable (Wilkinson, 1992). The visual representation of this process resembles the branches of a tree. The CHAID method of decision tree modeling uses the Pearson Chi squared statistic as the mechanism for merging and separating data groups to form the "optimally" merged groups. The break points between the "optimally" merged groups will be the cut points for the Hierarchical model that determine course placement. The "and" and "or" logic statements of the Hierarchical model are created naturally from the decision tree model with sister nodes forming the "or" statements and parent/child node pairs forming the "and" statements. Decision trees are "conceptually simple yet powerful" (Hastie et al., 2009, p. 305) predictive models that "have become popular alternatives to regression, discriminant analysis, and other procedures based on algebraic models" (Wilkinson, 1992, p. 2).

Binary Logistic Regression

Binary logistic regression, like other regression techniques, is used to obtain a model that "best describes the relationship between the outcome variable and a set of independent variables" (Hosmer et al., 2013, p. 1). However, other regression techniques such as linear regression and discriminant analysis make assumptions about the linear relationship between each outcome variable and covariate and about the normal distribution, constant variance, and mean zero of the error from the conditional mean (Hosmer et al., 2013). When the outcome is dichotomous, the assumptions about linearity and error are violated. These violations make regression methods using ordinary least squares, such as linear regression or discriminant analysis, unsatisfactory for predicting dichotomous or categorical outcomes.

Binary logistic regression is a maximum likelihood method "designed to maximize the likelihood of reproducing the data given the parameter estimates" (Peng et al., 2002, p. 5). It is specifically useful for predicting a binary or categorical outcome variable because the outcome is transformed through the logarithm to be linearly related to each covariate. In general, binary outcomes can be best estimated by a sigmoidal, or s-curve, which cannot easily be modeled using linear equations but can be transformed into a linear relationship using logarithm functions (Hosmer et al., 2013; Peng, Lee, et al., 2002).

The actual outcome of a binary event (e.g. no/yes or fail/pass) is 0 or 1. The probability of an event occurring has a range between 0 and 1 with 0 meaning that there is no probability of the event occurring and 1 meaning that the event is a probable certainty to occur. The odds (the probability of an event occurring divided by the probability of an event not occurring) are created to extend the range of probable values to be between 0 and infinity instead of 0 to 1. The logarithm, which has a domain of 0 to infinity and a range of negative infinity to infinity, transforms the odds into values that allow for linearity to exist between the outcome variable and each covariate. Additionally, the binomial distribution of the outcomes matches the binomial distribution of the errors (Hosmer et al., 2013) thus eliminating the need to assume that the errors are normally distributed with constant variance as in ordinary least squares methods. Overcoming these assumptions makes binary logistic regression an "extremely flexible and easily used function" (Hosmer et al., 2013, p. 7). The ability to predict binary and categorical outcomes has led to its increased use in social sciences and higher education research (Peng & So, 2002b; Peng et al., 2002).

Research Methods

This study will use a non-experimental design defined as a study in which the researcher does not randomly select and group the participants in the study and does not have a well-defined control group among the sample (Warner, 2012). As is often the case in educational research, an experimental design is not suited to the real-world nature of studying the lives of students (Schanzenbach, 2012). An experimental design for this study would require two students entering the college together to be placed using different placement models. Some students would be randomly assigned to one placement model while other students would be randomly assigned to the other placement model. In addition to the possible feeling of unfairness by some students, the logistical considerations of performing such an experiment make the experimental design unfeasible.

The site of the study was Metropolitan Community College (MCC), a large Midwestern community college with five campuses located in urban and suburban parts of a metropolitan city. Because the data was gathered from the entire community college system, it was infused with the diversity of the student population and the culture present on each of the campuses. Total enrollment for MCC is about 18,000 students with over 8000 new students enrolling each year (Metropolitan Community College, 2017). A large site was necessary for the study to allow a large enough sample size to increase the predictive power of the multiple measure placement models. The study used a sample from a single college instead of from multiple colleges due to the different course sequences at different colleges and the college-specific nature of placement.

Student record data was gathered via an internal query on the MCC operating system and analyzed to answer the research questions. Specifically, data gathered included high school cumulative grade point average, ACT-Math scores, mathematics courses taken at MCC and the semester in which the courses were taken to identify the initially placed course, and the grade earned in each mathematics course taken at MCC. Only a student's initial mathematics course was used in the analysis since this course corresponds to the course in which the student was placed. The sampling strategy was archival, or cohort, sampling in which all possible data was gathered within a preselected time interval delimited only by minimum requirements such as the availability of test scores, HSGPA, or course grade. Thus, the largest possible sample size within the time frame established that allows the research questions to be answered was gathered and analyzed. Data were gathered for students with a record of placement from 2012 through 2019 and separated into an initial sample (fall 2012 to fall 2017) used to build the placement models and a validation sample (spring 2018 to fall 2019). The data will be collected in Excel, delimited accordingly, and analyzed using SPSS software. Charts and figures representing the data analysis are shown in Chapter 4.

The placement measures HSGPA and ACTM scores as well as success data for into each course level were organized and analyzed for the development of the multiple-measure placement models. A hierarchical placement model was created based on the success rates for a combination of HSGPA and ACTM scores. Cut points for each measure were determined using the CHAID method of decision tree modeling with a maximum of two levels for each course level. A maximum of two levels limited the possible number of AND logic statements to one for each first level cut point. Using decision trees to determine cut points allowed the cut points to be statistically determined based on the data sample. For the algorithmic model, binary logistic regression equations were fitted to the data for each course level with independent variables of ACT-Math scores and cumulative high school grade point average and dependent categorical variable of success rate in each mathematics course-level. The regression equations were used in sequence to determine course placement. The strength of each predictor variable in the logistic regression equations are given in terms of odds ratio. Both models were analyzed for accuracy measures using the initial and validation samples at each course level.

Limitations and Ethical Considerations

Limitations

This section details the limitations of the proposed study and ways to perform ethical research that protects human subjects. To address the limitations of the study, consideration was given to possible weakness and ways to address the weaknesses. The limitations of the study include a lack of generalizability and the inherent limitations of a non-experimental design. Several validity issues also limit the study including data existing on only a subsample of the sample of MCC students, possibly inaccurate data, and models created using only two pre-chosen measures.

The study may not be generalizable to other colleges or universities since the sample consists of data from one community college. Peculiarities or characteristics of the college or student population sampled may influence the data in ways that do not exist at other colleges or for other student populations. However, the community college is a system of five campuses, each with a diverse population and culture. Each campus has between two and five thousand students enrolled with campuses located in urban and suburban locations giving the possibility of generalization to many colleges in the United States. Regardless, the goal of the study is not to generalize the exact placement outcomes but rather to study the effectiveness of the models. Certain scores on measures that place students into specific courses are not likely to be generalizable to other colleges, but the predictive ability of the models as a whole may be generalizable. Furthermore, this study may be useful as a reproducible model. This study could be repeated at other colleges interested in creating and comparing data-driven multiple measure placement models using institutional data. Using a sample from a single college aligns with the college-specific nature of placement. Future research is encouraged using datasets from other colleges and using samples from multiple colleges.

The validity of the data is another possible limitation of the study. Initial efforts at data screening showed that placement records are not available on all students. Of the approximately 40,000 unique student data records for placement between January 2012 and September 2015, about 40% of unique students had a recorded high school grade point average in the MCC system. Furthermore, less than 20% of unique students had record of high school grade point average, an ACT-Math score, and a grade for a mathematics course at MCC and, therefore, able to be part of the sample. One possibility for the small percentage of students with placement records is that students requiring specific financial aid or who are participating in the A+ Scholarship program were required to submit high school grade point average and ACT scores while the general population were not. If this is the case, then the sample used may have a confounding effect. For example, students participating in the A+ Scholarship program may succeed at lower (or higher) rates than the mean of MCC students thereby lowering (or raising) the scores from what they would have been if there were placement data for all students. This confounding effect would cause the placement model to underpredict where students should be placed. The A+ Scholarship program is a state-funded program that provides scholarship funds in the form of tuition and fees reimbursement to eligible high school graduates who attend a participating public community college or other two-year school (Missouri Code of State Regulations 6 CSR 10-2, 2017.).

The second limitation is that only two, pre-chosen measures, HSGPA and ACTM, were used to build the placement models. If data was available on more measures, such as last high school mathematics course taken, statistical tests could determine which measures should be included in the models. While research has shown the two chosen measures to be predictive of college success, some predictive strength, and therefore model accuracy, may have been lost by not including more measures in the model building process. Future studies should include additional measures in the design of the placement models.

Finally, this study was limited in its application because it was not an experimental study that can statistically compare the success of students actually placed using the two models. However, using different placement models for students at one college creates ethical issues and the researcher does not have the authority at the college to perform such an experimental study. Therefore, an experimental study of the placement models was not feasible and, instead, theoretical placement was compared utilizing historical data. Utilizing a validation sample to compare the two models adds to the validity of the comparisons over using an initial sample only, but not to the extent of an experimental study.

Ethical Considerations

No ethical considerations arose from this study since it was an analysis of archived, aggregated data. Student identification numbers from the original data set were transformed into nonrelated numbers to protect the privacy and confidentiality of students' personal and academic information. All data gathered were stored securely under password in the office of the principle researcher. The proposal for this study received approval from the Institutional Review Boards (IRB) at University of Missouri – Kansas City (UMCK) and Metropolitan Community College (MCC).

Significance of the Study

The placement of college freshmen into their first mathematics course has been drawn to the fore of conversation in higher education on the wave of discussion regarding college completion, by the announcement by the ACT Corporation that the Compass placement test will be eliminated, and by recent research on the possible ineffectiveness of standardized placement tests as predictors of college success. Utilizing multiple measures for placement is being pushed by national organizations and state agencies but without full knowledge of the effectiveness of the placement models. This study attempted to create two multiple measure placement models that reflect current model structures at U.S. colleges and to compare the models to determine which is more accurate and if the difference in accuracy measures and predicted placement are significant. As stated before, much research exists documenting the correlation of predictive measures with success in college, but little research has been done comparing the effectiveness of multiple-measure placement models. As a result, little has been written on the practical considerations of implementing different models such as ease or difficulty of implementation. This study gives insights into the practical considerations of each model. The study is also unique in that it uses decision tree models to statistically determine cut points based on the sample data used to create the hierarchical placement model. While decision trees have been used in educational research for college admissions (Wilkinson, 1992) and in determining at-risk students (Campbell et al., 2007), minimal research has been performed using decision tree modeling to determine cut scores for placement.

Definition of Terms

1. **Mathematics pathways** are mathematics courses that fulfill the mathematics requirement for an associate's or bachelor's degree. Math Pathways often include College Algebra, Statistics, Quantitative Reasoning, and Mathematics for Educators.
2. **Acceleration** is a design in which students have the ability to complete a course past the course in which they place within a single semester. Examples of acceleration designs include co-requisite models and a fusion of two courses in sequence into one course.

3. A **co-requisite** model is a design in which a student placing in a lower-level course enrolls in an upper level course and a supplemental course that offers the remediation necessary to pass the upper level course. It is a form of acceleration.
4. **Gateway courses** satisfy the degree requirement or courses for which placement is on par with the degree-satisfying course.
5. **Non-exempt students** are defined in the Scott-Clayton's (2012) study as students placed using placement tests due to not achieving a high enough score on another measure such as ACT or SAT to be placed in credit-bearing courses.
6. **Binary Logistic Regression** is a regression model that predicts a binary outcome given one or more independent predictor variables. It also gives the strength of the relationship between predictor variables and the dependent variable in the form of odds ratios.
7. **Accuracy** is the percent of correctly predicted outcomes (True Positives + True Negatives) / (Total Outcomes). In terms of the study this is (Correctly Predicted Successes + Correctly Predicted Failures) / (Total Outcomes)
8. **Sensitivity** is the percent of correctly predicted positive outcomes (True Positives) / (Total Positives). In terms of the study this is (Correctly Predicted Successes) / (Correctly Predicted Successes + Predicted Failures but Actual Successes) or (Correctly Predicted Successes) / (Total Actual Successes).
9. **Specificity** is the percent of correctly predicted negative outcomes (True Negatives) / (Total Negatives). In terms of the study this is (Correctly Predicted Failures) / (Correctly Predicted Failures + Predicted Successes but Actual Failures) or (Correctly Predicted Failures) / (Total Actual Failures).

10. **Discrimination** is given by the area under the ROC curve and interpreted as the ability of a model to differentiate between the success and failure of a pair of opposite performing students.
11. **Receiver Operator Characteristic (ROC) curve** plots the Sensitivity versus 1 – Specificity for each value of the probability cut point. In other words, it plots True Positives versus False Positives (also known as Type I error). The Sensitivity (or True Positives) is traditionally graphed on the y-axis. The area under the ROC curve is the discrimination.
12. The **probability cut point** is the chosen probability at which a positive outcome or negative outcome is determined for each case. If the resulting probability from the binary logistic regression is at or above the probability cut point after entering specific values for each covariate then the case is predicted to have a positive outcome.
13. The **logit**, or log-odds, is the natural logarithm of the odds a positive outcome. In terms of the study this is the natural logarithm of the odds of success in the mathematics course. The logit is the result of the binary logistic regression equation after entering values for the covariates.
14. The **odds** of an event occurring, such as success in a mathematics course, is the probability that it will occur divided by the probability that it will not occur. Thus, an odds of 1 means that there is a 50% probability that the event will occur and a 50% probability that the event will not occur. Odds greater than one indicates an increased likelihood that the event will occur.

15. An **odds ratio** is the ratio of the odds of one event to the odds of another event. The number $e \approx 2.71828$ raised to the Beta value (or coefficient) of the covariates in a binary logistic equation represents the odds ratio for that covariate. If the covariate is continuous, the odds ratio is the odds of the event at $(x + c)$ to the odds of the event at x. In other words, the odds ratio represents the change in likelihood of a positive outcome for every increase of c in the covariate. If the covariate is binary (Success = 1 and Not Success = 0, for example) then the odds ratio is the ratio of the odds of the event occurring (Success) to the odds of the event not occurring (Not Success).

CHAPTER TWO

LITERATURE REVIEW

Overview of the Literature

This literature review contains ten conceptual strands divided into three themes: placement measures, placement into developmental education as a benefit or barrier to success, and multiple measure placement models (Table 1). The first theme, placement measures, reviews the literature on the validity of measures used to place students into initial mathematics courses, specifically, ACT-Math scores (ACTM) and cumulative high school grade point average (HSGPA). The theme also includes research comparing scores on national standardized placement tests such as Compass and ACCUPLACER to ACT scores and HSGPA. Finally, the placement measure theme contains a strand on the research connecting non-cognitive measures such as motivation to HSGPA and measures of college success. The second theme has but one strand, a review of the benefits and barriers that may arise from students placing into developmental education courses. The third and final theme reviews the literature on multiple measures placement models. It presents the literature regarding two multiple measure placement models currently used at higher education institutions: the hierarchical placement model and the algorithmic placement model. Since the hierarchical model relies on the CHAID method of decision tree modeling to determine cut points for the placement measures, the theme details the functionality of the CHAID method. Since the algorithmic model relies on binary logistic regression, the theme also details the functionality of binary logistic regression as the statistical tool for predicting a binary outcome such as success or failure in an initial college mathematics course. In general, the literature details the effectiveness of the placement measures used in the study compared to other placement measures and insight into how the algorithmic and hierarchical placement models have functioned at higher education institutions. The research presented also brings to light the need for accuracy in placement by showing the ill-effects of misplacement, specifically, misplacement that leads to developmental education and by showing the lack of accuracy by current single-measure placement models.

Table 1

Themes and Conceptual Strands of the Literature Review

Theme 1: Placement Measures	
	Strand 1: The Validity of the ACT Test in Predicting College Success
	Strand 2: The Validity of HSGPA in Predicting College Success
	Strand 3: Comparing the ACT Test and HSGPA
	Strand 4: Comparing Standardized Placement Tests to the ACT Test and HSGPA
	Strand 5: Noncognitive Measures Inherent in HSGPA
Theme 2: Developmental Education as a Benefit or Barrier to Success	
	Strand 6: Developmental Education as a Benefit or Barrier to Success
Theme 3: Multiple Measure Placement Models	
	Strand 7: The Hierarchical Placement Model
	Strand 8: The Algorithmic Placement Model
	Strand 9: CHAID Method of Decision Tree Modeling in the Hierarchical Model
	Strand 10: Binary Logistic Regression in the Algorithmic Model

Placement Measures

The ACTM test and HSGPA were chosen as measures to predict the success of students in an initial college mathematics course, and therefore appropriate placement, due to significant correlation of the measures to initial success in college (ACT, 2008; ACT, 2014; Allen et al., 2008; Camara & Echternacht, 2000; Medhanie et al., 2012; Maruyama, 2012; Noble & Sawyer, 2004; Sanchez, 2013; Sawyer, 2013; Schmitt, 2012). This first theme of this literature review details a few prominent studies to bring to light the predictive relationship between the ACT test and college success and between HSGPA and college success. Additionally, the literature analyzes the relationship of HSGPA to noncognitive measures such as motivation and determination.

The Validity of the ACT Test in Predicting College Success

The ACT-Mathematics test is a 60-minute 60-item multiple choice test designed to assess the academic preparation for entry level college mathematics courses. The test requires "students to use their mathematical reasoning skills to solve practical problems in mathematics" (ACT, 2014, p. 6) over six content areas: pre-algebra, elementary algebra, intermediate algebra, coordinate geometry, plane geometry, and trigonometry. The range of scores for an ACTM test is between 0 and 36, although approximately 96% of ACTM scores and 93% of ACT-Composite (ACT-C) scores are between 12 and 32 (ACT, 2017). The ACT test is a widely used assessment with more than 1.9 million members of the Unites States high school class of 2015 having taken the ACT test (ACT, 2016).

In 2015, almost 50,000 Missouri high school students took the ACT exam (ACT, 2016) making it a highly available measure to use for college placement. Like the ACT, the

SAT test also attempts to measure academic preparedness in English, Math, and Reading (only the ACT has a Science section) and has similar correlations to early success in college (ACT, 2014; Rothstein, 2004). The two academic preparation assessments have also shown to be highly correlated to each other. Dorans (1999) found that the ACT and SAT have a .92 correlation to each other while the ACT-Math and SAT-Math were correlated at an r-value of .89. Thus, while some of the studies highlighted in this literature review use the SAT for their study, the similarities in function and high correlation of the two studies allow the research regarding one to be inferred to the other. Colleges can use a common crosswalk of ACT scores to SAT scores in their admissions and placement decisions (College Board, 2016). While either test could have been used as a measure to be studied, the ACT test was chosen for this study because of its availability in Missouri.

The ACT Corporation released its latest research on the reliability and validity of the ACT test in 2014 in its Technical Manual. Accord to the Technical Manual, the ACTM test has a median reliability coefficient of .91. The reliability coefficient is calculated as 1 minus the quotient of the Standard Error of Measurement squared and the Variance squared. The median Standard Error of Measurement for the ACTM test is 1.50 (ACT, 2014), indicating that, for a normal distribution of scores, the true score for about 2 out of 3 participants would be within 1.50 of their original score.

Several studies have examined the predictive power of HSGPA and ACT-C on First Year Grade Point Average (FYGPA) with most literature showing that ACT-C is moderately correlated to FYGPA with r-values ranging from .42 to .44, (ACT, 2008, ACT, 2014; Noble & Sawyer, 2004; Sawyer, 2013). In addition to correlational values, linear regression models showed that both ACT-C and ACTM coefficients were statistically significant predictors of

FYGPA with $p < .001$ and $p < .01$, respectively (ACT, 2014; Bettinger et al., 2011). The ACTM test explained 14.1% of the variance in FYGPA (Bittinger et al., 2011) while the combination of ACT-C and HSGPA accounted for between 23% and 39% of the variance in FYGPA (Sanchez, 2013; Schmitt, 2012).

The preceding research is evidence that the ACT Test is reliable and is moderately correlation to FYGPA. However, some researchers question whether a student's socioeconomic status (SES) including parents' educational level is an interaction between ACT-C or SAT and FYGPA (Bettinger et al., 2011; Geiser & Studley, 2002; Rothstein, 2004). The ACT-C is correlated with SES at at much higher r-values than HSGPA with r-values for ACT-C ranging from .31 to .34 and r-values for HSGPA ranging from .10 to .20 (Allen et al., 2008; Westrick et al., 2015). In linear regression models with HSGPA and SAT as dependent measures of FYGPA, the coefficients for SAT dropped in value between 12% and 23% when SES variables were added while the coefficient for HSGPA remained stable (Geiser and Studley, 2002; Rothstein, 2004). Thus, SES variables are correlated to ACT and SAT and account for some, but not all, of the ability of these academic preparation measures to predict FYGPA. The ability of HSGPA to predict FYGPA is largely unaffected by the inclusion of SES variables.

Comparing Standardized Placement Tests to the ACT Test and HSGPA

While the ACT attempts to measure the academic preparation or aptitude of a student for admission, placement, and scholarship decisions (Geiser & Santelices, 2007), standardized placement tests such as ACCUPLACER or Compass are specifically designed to place students into a particular level of mathematics, English, or reading course in college.

According to the ACCUPLACER Program Manual (College Board, 2016b) “the primary function of ACCUPLACER placement assessments is to assist with determining if students are prepared for a college-level course or if they would benefit from a developmental course” (p, 5). Some research, however, suggests that for traditional students (students entering college directly from high school), including a standardized placement exam provides little or no predictive value over a model comprised of ACT and HSGPA in determining first year success (Medhanie et al., 2012). Additionally, using only standardized placement exams severely misplaces about one-fourth of mathematics test-takers (Scott-Clayton et al., 2014).

However, success in college mathematics courses and other measures of college success are significantly correlated to scores on standardized placement tests (Medhanie et al., 2012; Robbins et al., 2006; Westrick & Allen, 2014). Westrick and Allen (2014) found that the Compass Pre-Algebra test was predictive of success in an Arithmetic Skills course and Elementary Algebra and that the Compass Algebra test was predictive of success in Intermediate and College Algebra. Medhanie et al. (2012) found that the ACCUPLACER test has a correlation r-value of .23, .25, and .32 for Arithmetic Skills, Elementary Algebra, and College Algebra, respectfully. Robbins et al. (2006) found that Compass tests had a correlation r-value of .25 to first semester GPA for 2-year colleges, specifically. This compared with r-values of .21 and .31 for ACT and HSGPA, respectfully. Medhanie et al. (2012) examined the correlation between ACCUPLACER and ACTM and created a logistic regression model that gave the predictive weights of ACTM and ACCUPLACER on course success in developmental and college credit mathematics courses (C or better). The researchers found that ACCUPLACER and ACTM have a mean correlation among the colleges studied of r equal to .43, which suggests overlap in the information provided by these two test scores.

When Medhanie et al. (2012) included ACTM and ACCUPLACER in a logistic regression model, only the ACTM scores were statistically significant. The odds ratio for ACTM in determining success in developmental mathematics and college level mathematics courses were 1.20 and 1.09, respectfully. Both were significant at the $p < .01$ level. The odds ratio for ACCUPLACER in determining success in developmental mathematics and college level mathematics courses were 1.00 and 1.01, respectfully - neither statistically significant. Medhanie et al. (2012) summarized by stating that “the evidence suggests that little (if any) predictive power is lost in relying on ACT mathematics test scores to place students in their first postsecondary mathematics course (p. 347). A further result showing greater differentiating power between college level mathematics courses and developmental mathematics courses was a comparison of mean scores for each level of mathematics. ACCUPLACER distinguished between the mean values of the developmental mathematics courses and college level mathematics courses with a Cohen’s d value of .28. For ACTM, the Cohen’s d-value was .90, showing a much greater ability to distinguish between college level and developmental level courses than ACCUPLACER (Medhanie et al., 2012).

On the other hand, standardized placement exams have some value in certain instances, specifically when compared to HSGPA (Westrick & Allen, 2014). The Compass Pre-Algebra test was better than HSGPA in predicting success in an Arithmetic Skills course. Thus, the lowest level of students placing into college may be more accurately placed if a standardized placement test is included. The Compass test also held similar values in predicting success for both traditional and nontraditional students whereas the predictive value of HSGPA was lower for nontraditional students than for traditional students. However, the predictive values of HSGPA in predicting course success for nontraditional students dropped to levels basically equivalent to those of the Compass exam. So, although HSGPA is not as strong of a predictor for nontraditional students as traditional students, its predictive ability for nontraditional students is still on par with standardized placement exams. In summary, standardized placement tests are weakly to moderately correlated to college success measures but do not add much predictive ability beyond that of ACT scores and HSGPA.

Comparing the ACT Test and HSGPA

Cumulative high school grade point average (HSGPA), on the other hand, is correlated to college success measures and offers significant predictive ability above the ACT test alone (Camara & Echternacht, 2000; Geiser & Santelices, 2007; Geiser & Studley, 2002; Sawyer 2013). Sawyer (2013) found that HSGPA alone is correlated to FYGPA at an r-value of .48 and HSGPA plus ACT is correlated to FYGPA at an r-value of .54, a change of .06. Camara and Echternacht (2000) found similar added predictive power for SAT over HSGPA alone. When the order of the variables is reversed, HSGPA adds a larger amount of predictive power over SAT than SAT did over HSGPA. According to the ACT Technical Manual (2014), ACT and FYGPA are correlated at an r-value of .42 and ACT plus HSGPA are correlated to FYGPA at an r-value of .53, a change of .11. Taken together, ACT and HSGPA account for a moderate percentage of explained variance in FYGPA with values of 18%, 23%, and 39% depending on the study (Maruyama, 2012; Sanchez, 2013; Schmitt, 2012).

Studies performed by researchers employed by ACT have shown that ACT and HSGPA have different strengths in predicting multiple aspects of college success (Noble & Sawyer, 2004; Sawyer, 2013). A 2008 ACT report summarized these different predictive strengths showing that HSGPA is better than ACT in predicting FYGPA ($r = .48$ vs. .42) and final college grade point average ($r = .35$ vs. .19). ACT, on the other hand is better than HSGPA in predicting academic proficiency in mathematics ($r = .57$ vs. .14) and level of degree attainment ($r = .16$ vs. .08). Academic proficiency is a measure of the grade earned in a course. A high correlation to academic proficiency means that the measure is better able to predict which students will earn A's in the course, for example. Thus, using ACT and HSGPA in combination may give a more accurate prediction of how a student will perform in the first year overall and how well a student will perform in a particular mathematics course than either measure alone. The ACT Corporation encourages colleges to use multiple measures in placement decisions:

> It is unlikely that ACT scores will measure all aspects of students' readiness for all first-year college courses. Therefore, it is advisable to consider using additional measures, such as high school coursework and grades, scores on locally developed placement tests, or noncognitive measures, in addition to ACT scores in making placement decisions (ACT, 2014, p. 116).

Noble and Sawyer (2004) and Sawyer (2013) performed similar studies in which HSGPA and ACT were predictors of different levels of FYGPA. The goals of the studies were to determine which levels of FYGPA were better predicted by ACT and which levels of FYGPA were better predicted by HSGPA. The studies reached similar results – HSGPA better predicted lower levels of FYGPA and ACT better predicted higher levels of FYGPA.

Specifically, Noble and Sawyer (2004) found that HSGPA is a more accurate predictor of FYGPA at levels of 2.00, 2.50, and 3.00 and ACT is a more accurate predictor of FYGPA at levels of 3.50 and 3.75. The predictive power is based on the ability of each measure to differentiate FYGPA within the specified range. Noble and Sawyer (2004) determined the predicted HSGPA and ACT-C score needed to have a 50% probability of achieving each level of FYGPA (2.00, 2.50, 3.00, 3.25, 3.50, and 3.75). The results are presented in Table 2. The HSGPA in this study is unweighted so the maximum value is 4.00. The authors note that no value of HSGPA can predict a FYGPA of 3.75. For many colleges in the study, a HSGPA of 4.00 corresponded to a less than 50% probability of earning a FYGPA of 3.25, 3.50, and 3.75. The lack of ability of HSGPA to predict high levels of FYGPA may be due to the tendency for a student's FYGPA to be between .2 and .6 points lower than that student's HSGPA (Morrissey & Liston, 2012; Noble & Sawyer, 2004; Sanchez, 2013).

Table 2

Predicted HSGPA and ACT-C for Each Level of FYGPA

FYGPA	HSGPA	ACT-C
2.00	2.21	14
2.50	2.78	18
3.00	3.39	22
3.25	3.73	25
3.50	3.91	27
3.75	N/A	30

Noble and Sawyer (2004)

Sawyer's (2013) research adds that HSGPA and ACT-C are both significant predictors at each level of FYGPA (2.00, 3.00, 3.50, and 3.70) in univariate and bivariate logistic regression equations at the $p < .001$ level. However, the standardized slope coefficients for HSGPA were "uniformly higher than those for ACT-C" (p. 102). Both HSGPA and ACT-C had greater slope coefficients at greater values of FYGPA indicating that both measures are "more strongly related to high levels of success than they are to low levels of success" (p. 101). Sawyer (2013) also found that at low levels of success, ACT plus HSGPA does not add much predictive power over HSGPA alone. Only when the value of HSGPA is 3.30 or above does ACT began adding incremental predictive ability.

An interaction effect may be present between ACT and HSGPA in predicting FYGPA. Sawyer (2013) found a significant interaction effect between HSGPA and ACT-C as a result of a greater differentiation among the probability of success at higher HSGPA's than for lower HSGPA's for various ACT-C values. For example, a HSGPA of 2.40 and an ACT-C of 10 yields a .10 probability of achieving a FYGPA of 3.00. A HSGPA of 2.40 and an ACT-C of 35 yields a .35 probability of achieving a FYGPA of 3.00. Thus, at a HSGPA of 2.40, there is not much predictive difference for an ACT-C of 10 versus an ACT-C of 35. However, at a HSGPA of 3.60, the range of probability for a FYGPA of 3.00 increases dramatically. A HSGPA of 3.60 and an ACT-C of 10 yields a .22 probability of achieving a FYGPA of 3.00. A HSGPA of 3.60 and an ACT-C of 35 yields a .95 probability of achieving a FYGPA of 3.00, a much larger difference in predictive ability. However, Sanchez (2013) found the interaction effect of HSGPA and ACT-C to be insignificant with the p-value at approximately .99 for both FYGPA levels of 2.50 and 3.00.

The ACT Technical Manual (2014) asserts that the ACT test is a more consistent predictor over time than HSGPA, which may be susceptible to grade inflation. The mean ACT-C from 1991 to 2003 changed from 20.6 to 21.0, a 2% increase. The mean HSGPA from 1991 to 2003 changed from 2.94 to 3.20, a 9% increase. Thus, the predictive value of a HSGPA of 3.50, for example, might be different now than it was a decade prior. Setting placement cut scores by a specific value of HSGPA might need to be continually adjusted based on recent high school graduates. In summary, HSGPA and ACT together are significantly better at predicting college success measures than either measure alone. HSGPA tends to better differentiate between lower values of FYGPA, overall FYGPA, and final grade point average. The ACT test tends to better differentiate between higher values of FYGPA, earned grade in mathematics courses, and degree attainment. Thus, using both measures to determine placement appears beneficial over one measure alone.

The Validity of HSGPA in Predicting College Success

The literature consistently concludes that HGPA is the best predictor of first year academic success with higher correlation values to FYGPA and greater standardized beta values in linear and logistic regression models than any other measure (ACT, 2014; Allen et al., 2008; Rothstein, 2004; Camara & Echternacht, 2000; Geiser & Studley, 2002; Maruyama, 2012; Noble & Sawyer, 2004; Robbins et al., 2006; Sanchez, 2013; Sawyer, 2013; Schmitt, 2012; Westrick & Allen, 2014). Geiser and Santelices (2007) attribute the high correlation of HSGPA to first year college success to the time period over which each measure is assessed. Standardized tests are “single shot assessments” that measure academic ability or aptitude for learning while HSGPA “accumulates over a period of years in a variety of subjects and reflects intellectual ability as well as motivation, personal discipline, and perseverance" (p. 26). The affective aspects of HSGPA will be discussed in a following strand of the literature review.

In relation to initial college success, HSGPA has a moderate to strong correlation to FYGPA with r-values ranging from .43 to .53 with a median correlation of about .48 (Allen et al., 2008; Camara & Echternacht, 2000; Noble & Sawyer, 2004; Robbins et al., 2006; Sanchez, 2013; Sawyer, 2013; Westrick et al., 2015). Logistic regression equations with FYGPA as the dependent variable have statistically significant positive beta-values for HSGPA (Robbins et al., 2006; Sawyer, 2013). The logistic regression model by Sawyer (2013) showed a significant ability of HSGPA in predicting FYGPA across multiple levels of FYGPA (2.00, 3.00, 3.50, and 3.70). Thus, both correlational and regression analyses have shown a moderate to strong connection between HSGPA and first year college success. In relation to long-term college success, HSGPA is correlated to final college grade point average at r-values ranging from .35 to .41, a moderate to strong correlation (ACT, 2008; Robbins et al., 2004).

In addition to a predicting initial and long-term college success, HSGPA has shown to improve the accuracy of placement models over models that solely rely on standardized tests. Studies by Ngo and Kwon (2015) and Scott-Clayton et al. (2014) show that including a high school transcript component in placement decisions increases the number of students who place into credit-bearing courses as opposed to developmental courses. Additionally, those students who placed higher due to the high school transcript component succeeded in the course at similar rates as their peers and attained similar credits in the long-term. Ngo and Kwon (2015) found that students boosted in placement using HSGPA were as successful as their peers in a similar score range and slightly but significantly more successful than the full sample of students at that mathematics course level. "The findings suggest that community colleges can improve placement accuracy in developmental math and increase access to higher-level courses by considering multiple measures of student preparedness in their placement rules" (Ngo & Kwon, 2015, p. 442).

The aggregate results of the Scott-Clayton et al. (2014) study show that 26% of students were severely misplaced using ACCUPLACER and Compass placement tests alone with 19% being severely underplaced and 7% being severely over-placed when success rates were held constant. When only high school transcript data is used to place students, the percentage of students severely misplaced drops to 23% with 18% severely underplaced and 5% severely over-placed. When placement tests and high school transcript data is used together the severe misplacement rate drops slightly to 22% with 18% remaining severely underplaced and 4% severely over-placed. When remediation rates were held constant, the probability of success in the college-level course using placement tests only was 69%. The success rate increased to 73% when high school transcript data only was used and to 74% when placement tests and high school transcript data was used in combination. The authors conclude that using high school transcript data would increase the success rate in a college level mathematics course over placement by tests alone holding constant the remediation rate.

Noncognitive Measures and Cumulative High School Grade Point Average (HSGPA)

This review of the literature on the relationship of noncognitive measures and HSGPA begins with Robbins et al. (2004) and weaves through the development and utilization of the Student Readiness Indicator (SRI). Robbins et al. combined constructs from educational persistence theory and motivational theory to create a model to predict student success. The individual constructs, or scales, used to build the noncognitive model were correlated to the cognitive measures of college GPA, HSGPA and ACT/SAT as well as retention. In this way, Robbins et al. were able to determine the strength of each scale on future success and retention in college and to bring to light the connection of each variable to cognitive measures. The researchers developed three hierarchical regression models to predict college GPA: one using only SES, HSGPA, and ACT/SAT, one using only noncognitive measures, and the full model using all predictors. The results compare the predicted explained variance of each model in college GPA and the added explained variance of noncognitive measures over traditional cognitive measures in predicting college GPA. It also shows the influence of each measure within the models.

Le et al. (2005) built upon the work of Robbins et al. but redefined some of the scales and included additional scales that were not present in Robbins's study but were present in other literature. The scales used by the researchers became the Student Readiness Inventory (SRI), "an inventory of psychosocial and skill factors" (Le et al., 2005, p. 499). The SRI would be used by others to further the research on noncognitive measures. Robbins et al., (2006) utilized the SRI in a large-scale study to retest the correlations of cognitive measures to the noncognitive scales in SRI and performed a hierarchical linear regression to isolate the effects of each noncognitive scale on college GPA and retention. Lastly, Porchea et al. (2010) followed a sub-sample of two-year college students from the Robbins et al. (2006) study for five years using multinomial logistic regression to measure the effect of each predictor on long-term college success, specifically degree completion and transferring to a four-year college. Table 3 shows a summary of the sequence of literature on noncognitive measures and HSGPA.

Table 3

The Sequence of Literature on Noncognitive Measures and HSGPA

Study	Overview and Findings
Robbins et al. (2004)	Studied the psychosocial and study skills correlation to College GPA • Academic Discipline and Academic Self-Efficacy are most correlated to college GPA • HSGPA is more correlated to Academic Discipline and Academic Self-Efficacy than ACT
Le et al. (2005)	Developed the SRI scale to study noncognitive measures based on the work of Robbins et al. (2004) and correlated them to HSGPA and ACT • HSGPA was moderately correlated to Academic Discipline, Commitment to College, Social Connection, and Academic Self-Confidence • ACT was moderately correlated to only Academic Self-Confidence
Robbins et al. (2006)	Correlated the noncognitive measures in the SRI to college success and retention measures • Only Academic Discipline was a significant positive predictor of FYGPA • Academic Discipline, Commitment to College, and Social Connection (for 4-year schools only) were significant positive predictors of retention
Porchea et al. (2010)	Measured the predictive ability of noncognitive measures on long-term college success measures • Academic Discipline and Commitment to College were positive predictors of earning a two-year degree and transferring while Academic Discipline also positively predicted earning a two-year degree and not transferring.

Robbins et al. (2004) defined and analyzed eight psychosocial and study skill factors plus one institutional factor in their model. Of the psychosocial and study skill factors, only Achievement Motivation (r = .257) and Academic Self-Efficacy (r = .378) were correlated with college GPA at an r-value greater than .20. In the regression model with only noncognitive measures to predict college GPA, only three measures were significant: Achievement Motivation (b = .161), Academic Goals (b = -.027), and Academic Self-Efficacy (b = .334). Thus, Achievement Motivation and Academic Self-Efficacy were the only significant positive influences on college GPA. Achievement Motivation was defined as "one's motivation to achieve success; enjoyment of surmounting obstacles and completing tasks undertaken; the drive to strive for success and excellence" while Academic Self-Efficacy was defined as a "self-evaluation of one's academic ability and/or chances for success in the academic environment" (Robbins et al., 2004, p. 267).

While both HSGPA and ACT were moderately to highly correlated with college GPA, r = .413 and r = .368, respectively, HSGPA had a much stronger correlation to Achievement Motivation and Academic Self-Efficacy, the measures that most strongly predicted success in college. ACT had a correlation of r = .142 to Achievement Motivation and r = .217 to Academic Self-Efficacy while HSGPA had a correlation of r = .326 to Achievement Motivation and r = .518 to Academic Self-Efficacy. In fact, HSGPA had a stronger correlation to every noncognitive measure than ACT except for General Self-Concept, r = .120 and r = .140, respectively. Thus, noncognitive measures that predict college success and HSGPA appear to be intertwined. Robbins et al. summarize that "these findings suggest that study skills are a precursor of positive class performance, which drives later achievement and persistence behavior" (2004, p. 276).

Le et al. (2005) developed the SRI to be a research tool for studying noncognitive measures. Their research also examined the correlations between each noncognitive measure, or scales, and HSGPA and ACT scores. HSGPA was moderately correlated at $r > .20$ with four measures: Academic Discipline ($r = .28$), Commitment to College ($r = .21$), Social Connection ($r = .20$), and Academic Self-Confidence ($r = .32$). HSGPA was also weakly correlated with General Determination ($r = .11$) and Goal Striving ($r = .16$). ACT, on the other hand, was only moderately correlated with Academic Self Confidence ($r = .32$). All other measures had correlations with ACT less than .10. While the study limits the correlations of the SRI scales to HSGPA and ACT, it can be stated with some certainty that HSGPA is more intertwined with noncognitive measures than ACT and might be one reason why HSGPA better predicts college success than other measures.

Robbins et al. (2006) used the SRI to analyze the relationship of the noncognitive scales to measures of college success including first-year college GPA (FYGPA) and first year retention. The researchers used correlational and regression analyses to determine the relationships. Only Academic Discipline was a significant positive predictor of FYGPA in the hierarchical linear regression model with a $b = .185$ ($p < .001$) for two-year colleges and $b = .221$ ($p < .001$) for four-year colleges. Academic Discipline and Commitment to College were the only significant positive predictors of first-year retention in the logistic regression model for two-year schools with odds ratios of 1.14 ($p < .05$) and 1.11 ($p < .05$), respectively. Social Connection was an added positive measure for students at four-year colleges with odds ratios of 1.14 ($p < .01$), 1.19 ($p < .001$), and 1.13 ($p < .01$) for Academic Discipline, Commitment to College, and Social Connection, respectively. Academic Discipline was the highest correlated measure to FYGPA and first-year retention for both 2- and 4-year colleges.

In summary, Robbins et al. (2006) showed that, of the noncognitive measures in the SRI, Academic Discipline was the most predictive of FYGPA while Academic Discipline, Commitment to College and Social Connection were positive predictors of first-year retention. Combining the results from Robbins et al. (2006) with Le et al. (2005) shows that HSPGA is moderately correlated to the one noncognitive measure that positively predicts college success and the three noncognitive measures that positively predict college retention.

A longitudinal study by Porchea et al. (2010) added degree completion and transfer rate of two-year college students to the list of success measures positively associated with Academic Discipline and Commitment to College.

> The probability of obtaining a degree and transferring (rather than dropping out) increased from 0.50 to 0.55 with each standard deviation increase in Academic Discipline score; the probability of obtaining a degree and not transferring increased from 0.50 to 0.61. The probability of obtaining a degree and transferring (rather than dropping out) increased from 0.50 to 0.56 with each standard deviation increase in Commitment to College score (Porchea et al., 2010, p. 700).

An increase in Academic Self-Confidence, the only noncognitive measure positively correlated to ACT above $r = .20$, was negatively associated with degree attainment for two-year college students in the study.

Placement into Developmental Education as a Benefit or a Barrier to Success

Developmental education is a course or sequence of courses designed to help students who do not qualify for college level courses acquire the knowledge and skills needed to be successful in the college level course in that discipline. In general, developmental education, also called remedial education, exists in mathematics, English, and/or reading. For mathematics, the developmental education sequence may be one course or up to four courses and students are usually required to complete the sequence prior to enrollment in college level mathematics. Recently, however, co-enrollment in a college level course and a developmental course in which the developmental course is used to support learning in the college level course has become a popular option (Vandal, 2014) and endorsed by national organizations such as Complete College America (Complete College America, 2013). This arrangement is often referred to as a corequisite model.

Placement is tied to developmental education since about half of entering college students place into developmental education with most of those students placing into developmental mathematics (Attewell et al., 2006; Bailey et al., 2010; Burley et al., 2001). Based on Scott-Clayton et al. (2014) analysis that up to one-fourth of students are severely underplaced into developmental mathematics courses, if students required to take developmental courses succeed, persist, or complete the degree at a rate lower than similarly-prepared peers in college level courses, placement into developmental education may be a barrier to their success. Placement into developmental education might be a benefit to success if those placed in developmental courses succeed, persist, and complete degrees at higher rates than their similar counterparts. This strand of the literature review will focus on success in college for students enrolled in developmental education. The research presented is highlighted by studies using massive databases to show the course-taking patterns of students in developmental mathematics and analyze the benefit or barrier of developmental education and several studies using Regression Discontinuity (RD) to analyze the students on the border of placement.

The percentage of students placing into at least one developmental education course is estimated to be about 40 percent of all college students and more than half of all community college students (Attewell et al., 2006; Bailey et al., 2010; Burley et al., 2001; Scott-Clayton & Rodriguez, 2012). The large numbers of students in developmental education amplifies the possible positive or negative impact of developmental coursework. To determine the impact of developmental education, Bailey et al. (2010) and Attewell et al. (2006) used large-scale databases to determine the enrollment patterns in developmental courses including the completion of developmental coursework, a college level mathematics course, and a degree. Bailey et al. (2010) used the database from Achieving the Dream schools with a sample of over 250,000 students from 37 colleges to study the developmental education sequence. According to the research, 59% of the sample were referred to developmental mathematics with 24% referred to one level below college level mathematics, 16% referred to two levels below college mathematics, and 19% referred to three levels below college level mathematics. Of these students, only 33% finished the developmental mathematics sequence with 45%, 32%, and 17% from one, two, and three levels below college level mathematics, respectively, completing the developmental sequence. Perry et al. (2010) found similar results at California community colleges with 73%, 41%, 24%, and 13% completing the developmental mathematics sequence from one, two, three, and four levels below college level mathematics, respectively. Therefore, many students referred to developmental education never make it out of the developmental sequence and into college level mathematics. In fact, only 20% of students in mathematics remediation completed a college level mathematics course within three years (Bailey et al., 2010, p. 260).

While it might seem like the failure to complete the developmental mathematics sequence and the college level mathematics course might be due to students' lack of mathematical skills, Bailey et al. (2010) found that failure to enroll initially or in the next course in sequence is a greater barrier to success than course failure or withdrawal (p. 260). For mathematics, 11% of students exit the developmental sequence without ever failing a course and 27% of students assigned to developmental mathematics courses never enrolled. Jaggars and Hodara (2011) found similar results using data from the City University of New York, a system of 23 institutions including some community colleges. According to the study, 46% of students stopped the progression through developmental mathematics because they failed the course, 18% stopped because they simply did not enroll in the next course, 7% completed the developmental sequence but did not enroll in the college level course, 6% failed the college level course, and 22% passed the college level course. As with the Bailey et al. (2010) results, the further removed a student begins from the college level course the lower the percentage of students completing a college level course.

Bailey et al. (2010) also analyzed the success rates of students who were placed into developmental courses but ignored their placement and enrolled directly into college level courses. While some selection bias might exist, these direct-enrollers succeeded in the college level course at rates similar to those who enrolled in a college-level course after completing the developmental sequence. Comparing the completion rates of these direct enrollees to those taking the developmental course, however, shows that 72% of direct enrollers completed college level courses compared to 27% of those who complied with their placement and enrolled in developmental coursework (p. 261).

Students who do not enroll in a subsequent semester are referred to as stopouts and if they fail to return to college within a set time period, they are referred to as dropouts. Burley et al. (2001) analyzed the stopout and dropout patterns in Texas community colleges and found that the longer a student is enrolled, the more consecutive the enrollment, and the more frequent the enrollment were correlated to higher college GPA. The dropout and stopout rates in the first two semesters for students in developmental education were ten percentage points higher than students not enrolled in developmental education, 43.1% to 33.1%. Some have claimed that the structure of developmental education, especially developmental mathematics at the community college with two or three courses below college level, is a barrier to success because it has multiple exit points for which students can dropout or stop out. Edgecombe (2011) researched the developmental education sequence and came to the conclusion that "this analysis illuminates a major structural deficiency in the traditional sequence – a multitude of exit points available to and taken by students – that seriously undermines academic achievement" (p. 1). Similarly, Bailey et al. (2010) state that "one interpretation is that the developmental education obstacle course creates barriers to student progress that outweigh the benefits of the additional learning that might accrue to those who enroll in remediation" (p. 261).

Fong et al. (2014) demonstrate the effect of multiple exit points in the developmental education sequence using data from the 2004 National Postsecondary Student Aid Study of California community colleges. Analyzing the enrollment patterns for one cohort in the four developmental mathematics courses, they determined the rates for placement, enrollment, success, and enrollment in the next course in sequence (Figure 1, p. 41). For Arithmetic, 15,106 students placed in the course, 9255 (61%) enrolled in the course, 5961 (64% of the

9255) passed the course, and 4310 (72% of the 5961) enrolled in Pre-Algebra. At each step there are negative effects of not enrolling in the course initially, not passing the course, and not enrolling in the next course even when passing. Only 1004 (6.6%) of the initial 15,106 who placed into Arithmetic completed the developmental sequence. This percentage rises to 10.8% if only the initially enrolled students are considered. The results for Pre-Algebra are similar with 14,879 students placing into Pre-Algebra, 10,035 (67%) enrolling, 6776 (68% of the 10,035) passing, and 5055 (73% of the 6776) enrolling in Elementary Algebra. In total, 1746 (11.7%) of students placing in Pre-Algebra complete the developmental mathematics sequence with that percentage rising to 17.4% when only initially enrolled students are considered.

Attewell et al. (2006) performed a database study similar to the study by Bailey et al. (2010) but analyzed graduation rates instead of the developmental sequence and college-level course success. The study analyzed data from the 1988 National Educational Longitudinal Study (NELS88) tracking student achievement over an 8-year period. Raw data showed that students enrolled in developmental education graduated at much lower rates than students who were not enrolled in developmental education, 28% to 43%, respectfully, for students at two-year colleges. This mirrors the course success results from Bailey et al. (2010). However, when statistical techniques were employed to control for prior academic achievement, SES variables, race, and gender, enrollment in remedial coursework had no impact on graduation rates for students at 2-year colleges and small but negative impact on graduation rates for students at 4-year colleges. The researchers used the statistical technique of counterfactual model of causal inference to match students with similar background characteristics, one enrolled in developmental courses and one not enrolled in developmental courses using logistic regression and a propensity score of less than .01 to match the students. Students with no matching pair are removed from the sample. The two groups of matched students were then compared on measures such as graduation rate.

For two-year colleges, students enrolled in developmental coursework were predicted to graduate at a statistically insignificant rate of 3% more than their peers who were not enrolled in developmental coursework. For four-year colleges, students enrolled in developmental coursework were predicted to graduate at a statistically significant rate of 7% lower than their non-developmental peers and take .2 years longer to graduate, a statistically but not necessarily practically significant amount of time. These results suggest that the developmental sequence itself might not be a barrier to success for two-year colleges. "Taking one or more remedial courses in a two-year college does not, in itself, lower a student's chances of graduation. Causal factors that do reduce one's chances of graduating include low family SES, poor high school preparation, and being Black, but not college remediation *per se*" (Attelwell et al., 2006, p. 905). Scott-Clayton and Rodriguez (2012) argue, however, that the grade on a college level course is a better outcome to measure than graduation, which may be based on much more than the fairly narrow treatment of remediation (p. 5). Attewell et al. (2006) state that there is no evidence that remediation improves graduation chances in any statistically significant sense.

The final portion of the developmental education strand in this literature review summarizes several studies that use regression discontinuity (RD), a statistical mechanism to compare two groups separated by a single measure. For example, the two groups may be separated based on whether they scored above or below the cut score for placing into credit-bearing courses and are compared on any one of several measures including graduation or grade in a college-level mathematics course. Regression discontinuity uses Ordinary Least Squares (OLS) regression on each of the groups separately and analyzes the difference in value (or discontinuity) at the cut point. If the two regression lines (or curves) are continuous or close to continuous at the cut point then the treatment is said to have no effect. The treatment in each of these studies is placement into developmental education. If there is a gap or discontinuity in the regression lines for the treatment and the control (no developmental courses) then, depending on the context, the treatment has a positive or negative effect on the chosen outcome. RD often compares the values of each predicted regression within a narrow band region around the cut point and for the entire sample. The slopes of the regression lines may indicate a positive or negative effect of the treatment and are based on the interaction effects of the variables in the regression equations. In the context of analysis of an outcome based on a predetermined cutoff score, RD is considered a close approximation to an experimental design, especially for students within the narrow band surrounding the cut score.

> RD design is the closest nonexperimental research design to a random assignment experiment in which a portion of entering students would be assigned to one level of math and a portion would be assigned to the next higher level. Finally, by focusing on students who score close to the cut point, RD design will most closely resemble a true randomized experiment as the actual underlying difference in the ability of those taking the tests will vary little within the study sample. In other words, many of the students who test below the cut point might have tested above on a different day or different test form and vice versa. Thus, the assignment to different math levels will

> be determined largely by testing error, not by differences in the actual underlying ability of the students taking the test (Melguizo et al., 2011, p. 176).

Six studies highlighted in this literature review analyzed the effects of developmental education on a host of outcomes. Five of the studies used regression discontinuity and one study used instrumental variable (IV) regression which corrects for bias in OLS regression when one or more of the covariates are correlated with the error in the outcome variable. The results of the six studies were mixed ranging from indicating that developmental education is beneficial to college success to indicating that developmental education is a barrier to success. For most outcomes, however, developmental or remedial education had little to no effect.

Bettinger and Long (2009) used IV regression on data from several colleges in Ohio using distance from college and the remediation rates at the college of choice to control for bias in the OLS regression. In this study, mathematics remediation was predicted to increase the 5-year degree completion rate by 9.0% compared to similar students controlling for prior academic, socioeconomic, and other background variables. Mathematics remediation was also predicted to lower one-year dropout rates by 8.3% and five-year dropout rates by 9.6%.

Calcagno and Long (2008) used regression discontinuity to study the effects of remediation on persistence, credits earned, future math course performance, and degree completion. Students within the margin of the placement cutoff but who entered developmental education were 2% to 3% more likely to persist to the second year and earn about 7 more credits. However, a majority of the earned credits were earned in developmental courses since the total college-level credits earned were roughly equivalent for students in developmental courses and students not in developmental courses. In addition, no discernable difference was found in degree completion, long-term persistence, and future math course performance. “Overall, the results suggest that remediation might promote early persistence in college, but it does not necessarily help students on the margin of passing the cutoff to make progress toward a degree” (p. 22).

Scott-Clayton and Rodriguez (2012) used RD design to analyze similar outcome variables as Calcagno and Long (2008). The researchers used a sample of over 100,000 students from one large community college system. In this study, developmental education showed no effect on degree completion, persistence, or passing the college exit exam. However, developmental education was predicted to lower success in the college level mathematics course. Students who were placed into developmental education were predicted to take college level mathematics at an 8% lower rate, pass at a 5% lower rate, earn a C or better at a 4% lower rate, and earn a B or better at a 2% lower rate. All of these results were statistically significant, $p < .01$. However, since these differences in success rates are relatively low and no other measure was significant, Scott-Clayton and Rodrigues concluded that developmental education neither increases or decreases college success.

Boatman and Long (2010) used RD design on a 2000 cohort of students from two-year and four-year colleges and universities in Tennessee to analyze college credits and degree completion for each level of developmental mathematics. The researchers analyzed the outcome measures for students who began at college level to students who began one level below college, one level below college level to two levels below college level, and two levels below college level to three levels below college level. Students who began one level below college level were predicted to have fewer college credits in three years and a decreased rate of degree completion, especially for students at two-year colleges, compared to peers who began in college-level courses. Students who began two levels below college level earned less college credits than similar students who began one level below college level but had similar degree completion rates. The same holds for students beginning three levels below college level compared to similar students beginning two levels below college level. On a positive side for remediation, students beginning one level below college level had a greater, although not significantly so, predicted grade in the first level college course than students who began in college level mathematics. Thus, developmental education courses decreased overall college success but might help grades in an initial college-level course.

Moss and Yeaton (2014) used RD design and a sample of over 2000 students from one community college as well as an experimental design with sixty-three student volunteers who were placed in developmental mathematics or college level mathematics at random. The sixty-three volunteers had initial placement scores within a ten-point range of the cut point for remediation. The outcome being compared was course grade in the college level mathematics course. Note that the college level mathematics course in the study, Intermediate Algebra, is considered by many colleges to be one level below college level mathematics. Thus, in comparison to the other studies discussed in this literature review, this study analyzed the influence of a developmental mathematics course two levels below college level on the grade of a course one level below college level. Students who placed into the course two levels below college level were predicted to earn a higher grade in the course one level below college level than students who placed into one level below college level thus showing a slight benefit for enrollment in a developmental mathematics course.

The difference was about .3 of a grade point on a 4-point scale and was statistically significant, $p < .05$. The random experiment yielded similar results with students placing in the course two levels below college level earning a grade .29 higher in the course one level below college level than student placing one level below college level thus validating the results of the regression discontinuity design.

Martorell and McFarlin (2011) used RD design to assess the impact of developmental education on academic and labor market outcomes for a large sample of students at two-year colleges in Texas. Students in developmental education were predicted to earn fewer academic credits and complete one year of college at a lower rate. Both differences in these outcomes were significant but small. The differences in degree completion, transferring to a four-year college, and labor market earnings were predicted to be small and statistically insignificant when comparing students in developmental courses to similar students not placed in developmental courses. Martorell and McFarlin (2011) conclude that “marginal students in Texas receive little benefit from remediation, despite the large financial cost of the program” (p. 26). The mixed results of the experiment mirror the mixed results from the collection of RD studies on the benefit or barrier of developmental education.

Placement Models

An analysis of the multiple measure placement models currently being used by colleges in the United States was conducted. Several models were found but only a few unique models existed with literature on their effectiveness including a hierarchical model and an algorithmic model. No journal article could be found that experimentally tested the models against one another in terms of accuracy or other factors. This study attempts to compare of multiple measure placement models as well as determine their individual ability to predict course success. As such, prior research on the hierarchical and algorithmic models is reviewed in this strand of the literature review. The perceived strengths and weaknesses of each model based on the literature will also be discussed.

Multiple Measure Placement Models

Three distinct multiple measure placement models for college placement were found by the researchers in reviewing the literature: a hierarchical placement model, a decision tree model, and an algorithmic placement model (Bahr et al., 2014; Barnett et al., 2016; Reeves Bracco et al., 2014). The hierarchical model uses cut points of different measures to place students into mathematics courses. A student qualifies for a particular course if the student scores at or higher than the cut point on any one of the placement measures. The decision tree model is similar to the hierarchical model in that it uses cut points for each measure but differs in that it may use different cut scores for a particular measure depending on a student’s score on another measure utilizing and/or logical statements and the cut points are statistically determined. The algorithmic model uses an algorithm to place students based on the value of each placement measure using either binary or ordinal logistic regression (Barnett et al., 2016). These placement models are general model types and do not specify which placement measures are in the models. Placement measures vary by college but most often include high school transcript data and one or more standardized tests (Reeves Brocco et al., 2014).

This literature review will focus on two of the distinct models, hierarchical and algorithmic, because no study could be found on student success using a decision tree model and because the number of variables needed to differentiate the decision tree model from the hierarchical model does not exist in this study. Decision trees will be utilized in this study to statistically determine cut points for placement measures in the hierarchical model. Note that, according to Bahr et al. (2014), a decision tree model has some advantages over an algorithmic model including the "interpretability of output, robust handling of scalar and categorical data of any distribution, and inclusion of non-linear relations and interaction effects without additional specifications" (p. 8). Bahr et al. (2014) are currently studying decision trees as a multiple measure placement model for California community colleges. A further description of the design of the hierarchical and algorithmic placement models, an example of how a college uses each model, the research surrounding that college's use of the model, and possible strengths and weaknesses of each model will be discussed.

Hierarchical Model

The hierarchical placement models found in literature utilize an "or" functionality in placement. A student scoring at or above the cut point on any one of the placement measures qualifies to enroll in a particular mathematics course (Barnett et al., 2016; Morrissey & Liston, 2012). For example, if HSGPA, ACTM, and ACCUPLACER Math scores are used as measures of placement, cut points would be set based on historical data, and a student with a HSGPA at or above the cut point would be eligible to enroll in a college level mathematics course. If the student had a HSGPA below the set cut point, then the student's ACTM score would be compared to the cut point for ACTM. Again, if the student's ACTM score is at or above the set cut point, the student would be eligible to enroll in a college level mathematics course. The process would continue for the ACCUPLACER math measure. The hierarchical model in this study is unique in that it allows “or” and “and” logic statements to maximize the accuracy of the model. The CHAID method of decision tree modeling will be the statistical tool used to determine the cut points and the utilization of the “or” and “and” logic statements. The CHAID method of decision tree modeling allows for input measures to be continuous, such as HSGPA, ordinal, or categorical. Note that the CHAID method will create similarly-sized ordinal or categorical groups (or segments) out of the continuous data.

The North Carolina Community College System (NCCCS), a 58-college system serving over 700,000 students (North Carolina Community College System, 2016), utilizes a hierarchical model for placement into college level mathematics. The NCCCS uses four placement measures: prerequisite college credits, a combination of an unweighted cumulative 11-semester HSGPA of 2.60 or greater and a fourth math benchmark high school course, ACT-Math or SAT-Math scores at national benchmarks (22 and 530, respectfully (Central Carolina Community College, n.d.)) and the North Carolina Diagnostic and Placement (NCDAP) test scores at or above course-specific cut-points (Morrissey & Liston, 2012; North Carolina Community College System, 2016). Math benchmark courses are high school courses designated by the State Board of Community Colleges as courses eligible for multiple measures placement and are listed in Appendix A. The NCDAP is a customized version of the ACCUPLACER test (Central Carolina Community College, n.d.). A student who has satisfied any one of these requirements will be placed in a college level mathematics course at a North Carolina community college. It should be noted that colleges may require students to enroll in a supplemental corequisite course if a student’s HSGPA exceeds the cut point by only a small margin (North Carolina Community College System, 2016). Student scores on measures used for placement are valid for five years, up from a previous validity range of two years (Barnett et al., 2016; Central Carolina Community College, n.d.).

At the behest of the NCCCS president, the Community College Research Center (CCRC) studied the historical placement and success data for approximately 20,000 NCCCS students to determine if an improved placement model could be created to increase enrollment and success in college-level courses. Prior to fall 2013, the NCCCS placed students in college-level mathematics using a combination of three standardized placement tests: Compass, ACCUPLACER, and Asset (Morrissey & Liston, 2012; Reeves Brocco et al., 2014). The CCRC study found that HSGPA was “significantly more predictive of a student’s college success than test scores from the existing placement tests” (Morrissey & Liston, 2012, p2). The model created by CCRC researchers based on the historical data suggested that the number of misplaced students would be cut from 30 percent to 15 percent across English and mathematics (Morrissey & Liston, 2012). Based on the results, NCCCS adopted the hierarchical multiple measures placement model described above. Colleges within the district were able to use the placement model beginning fall 2013 and were required to use the placement model by fall 2016. The HSGPA value of 2.6 was chosen based on the results of research that showed that grade point average in the college Gateway mathematics course is 0.6 points lower than the cumulative HSGPA for the average student. Thus, to create an average passing grade point average of 2.0, the NCCCS set the HSGPA level for entry into the Gateway course at 2.6 (Morrissey & Liston, 2012). Research is ongoing to ensure the accuracy of the cut-off points of the measures used to place students and of the placement model overall. Cut points for other measures were chosen based on national standards and were updated for fall 2016 (North Carolina Community College System, 2016).

A hierarchal model may have some strengths and weaknesses compared to the algorithmic model. Possible strengths of the model are that it can incorporate several placement measures with specific values for each measure, can be adjustable using updated data, can be easily understood by policymakers, faculty, advisors, and students, and can be programmed into a college's CRM relatively easily. One weakness of hierarchical models is that cut point values may be subjectively determined, determined by broad-based research such a comparison of mean values of the predictor and dependent variables, or determined using national standards that may or may not reflect the student population. This study mitigates a portion of this weakness by statistically determining cut points for placement measures using decision tree modeling but retains some placement based on national standards to allow the model to be of practical use.

Algorithmic Model

The algorithmic placement model utilizes binary logistic regression or ordinal logistic regression to maximize the likelihood of a binary or ordinal outcome based on historical data. It uses values of input measures and the success or failure of the students with those given values to predict the success or failure of future students. The logistic regression models allow for input measures to be continuous, such as HSGPA, or categorical, such as whether a student completed a fourth year of high school mathematics. A discussion of the basics of binary logistic regression, the method that will be used in this study, appears in the next strand of the literature review.

Long Beach City College (LBCC) piloted an algorithmic multiple measures placement model for high school students at one school district beginning in fall 2012 using logistic regression. Fuenmayor, Hetts, and Rothstein, the lead researchers of the LBCC study, used five measures from high school transcript data in the algorithmic model for placement into a Gateway mathematics course: the last math course taken, grade in last math class, 11th grade California Standards Test (CST) score in mathematics, overall mathematics GPA, and cumulative high school grade point average (HSGPA) (Barnett et al., 2016; Willett, 2013). Specific coefficient values for the regression equation were not found but Willett (2013) and the LBCC research brief (2014) report that college success was strongly correlated to high school success. The measures used for student placement were the result of logistic regression modeling with five years of historical data on approximately 5000 LBCC students (Barnett et al., 2016). Long Beach City College enrollees not in the pilot placement program are placed according to the ACCUPLACER exam with exceptions for showing evidence of college course completion in high school or college with a grade of A or B (Long Beach City College, n.d.). The LBCC placement pilot became a model for the Student Transcript-Enhanced Placement Study (STEPS) to research the predictive ability of high school transcript data on college success across ten colleges in California and multiple K12 school districts.

The Long Beach City College algorithmic placement model placed students in Gateway mathematics courses at a rate of 30% compared to 9% from the 2011 cohort from the same school district that used traditional placement. Actual enrollment of students in Gateway mathematics courses was 16% for those placed via the algorithmic placement and 5% for those placed traditionally, more than three times as many enrollers in a Gateway mathematics course for students placed using the algorithmic model. Success rates, however, remained about the same with 52% of the pilot placement students succeeding compared to 54% success for the baseline cohort (Long Beach City College, 2014; Willett, 2013). LBCC expanded the algorithmic placement in fall 2013 to include three school districts. The 2013 cohort of students placed via the algorithmic model enrolled in Gateway mathematics at a rate of 22% and succeeded at a rate of 51% (Long Beach City College, 2014).

The STEPS research, led by Willett, expanded the number of colleges studied to ten and developed an algorithmic model for each college using its own historical data as opposed to merging the data and determining an overall model for all colleges (Willett, 2013). Measures of placement that most correlated to student success and projected placement percentages in Gateway mathematics were determined for research but not utilized in practice. The predictor variables tested in the algorithm were the most recent CST mathematics score, the number of college preparatory courses completed in high school, the level of the most recent high school mathematics course, the grade in the most recent high school mathematics course, and HSGPA excluding mathematics courses (Willett, 2013).

Ordinal logistic regression was used by two of the ten colleges with binary logistic regression used by the other eight. The Nagelkerke pseudo R^2 value used to estimate the predictive value of the model was .18 (an average of the 10 models) for predicting mathematics success, a moderately weak value for educational research (Willett, 2013). HSGPA excluding mathematics courses, the highest level of mathematics taken in high school, and the grade received in the last high school mathematics course all moderately predicted success in college mathematics. Scores on the CST mathematics test and the grade received in the last high school mathematics course weakly predicted college success with fewer colleges reporting the measures as significant predictors. “These findings suggest that high school transcript data are at least partially predictive of college performance” (Willett, 2013, p. 17).

Possible strengths of the algorithmic model are that it uses a combination of measures, which may increase predictive ability, can incorporate several placement measures, and can be adjusted using updated data. The model is created statistically thus reducing the possibility of human bias or subjectivity in determining cut scores. One possible weakness of the algorithmic model is that policymakers, faculty, advisors, and students may not easily understand it. Students will not necessarily know which level of course they are allowed to enroll in based on their placement measures (as in a hierarchical model). Advisors may not be able to inform a student how much he/she needs to improve on a placement measure to be allowed enrollment into the next higher-level mathematics course.

An alternative algorithmic model to logistic or ordinal regression was created by mathematics/computer science faculty at Mercer University in Georgia. Denny et al. (2012) used a simple algorithm based on “natural breaks” in HSGPA and SAT Mathematics (SATM) data to place students. They began by statistically determining linear regressions for three courses, Calculus, PreCalculus, and Intermediate Algebra, using the placement measures: HSGPA (x_1), MAA1 (x_2), MAA2 (x_3), and SATM (x_4). The MAA1 test and MAA2 test are mathematics placement exams by the Mathematics Association of America used to determine preparedness for PreCalculus and Calculus, respectively. The dependent variable for each linear regression was grade (on a 4-point scale) in the initial mathematics course. Denny et al. (2012) found that HSGPA was the most important variable in each regression equation for determining the grade of the college course with SATM not significant in any of the three equations. The regression equations are as follows:

Intermediate Algebra: $Y_1 = -3.218 + 1.042\ x_1 + 0.242\ x_2$

PreCalculus: $Y_2 = -2.812 + 1.088\ x_1 + 0.134\ x_2$

Calculus: $Y_3 = -4.750 + 1.579\ x_1 + 0.102\ x_3$

Due to a need to know placement test scores prior to orientation, HSGPA and SATM were used as placement measures instead of the MAA placement tests that were given on campus. They then organized the data into a matrix of HSGPA versus SATM and looked for natural breaks in the results (Denny et al., 2012). Natural breaks in the data were found and estimated using a level curve – a mathematical term similar to a curve of constant elevation on a topographical map. Using the formula 100*HSGPA + SATM, they found a common value at which the natural break occurred. In PreCalculus, for example, the level curve occurred at 850 so that students with a score of 850 or above using the formula 100*HSGPA + SATM were placed into PreCalculus. Similarly, the cut point for the level curve score for Calculus occurred at 950. Thus, the researchers set up a placement model based upon the algorithm 100*HSGPA + SATM in which a score of 950 or above placed a student into Calculus, a score of 850 to 949 placed a student in PreCalculus, and a score below 850 placed a student in Intermediate Algebra. While the algorithm is a simple combination of two measures with subjective cut points, the placement model is algorithmic in nature.

Mercer University implemented the simple algorithmic placement model beginning fall 2003 and compared the placement and success data from fall 2003 – fall 2009 using the algorithmic model to data from fall 1997 – fall 2002 using only SATM scores, the prior method of placement. Denny et al. (2012) found that course success increased from 57% to 68% for Intermediate Algebra, from 72% to 77% for PreCalculus, and from 79% to 83% for Calculus, all statistically significant increases. More specifically, males increased success in all three courses while females increased success in Intermediate Algebra only using the algorithmic placement model. African American success rates increased in all three courses while Asian/Pacific Islanders and Caucasian success rates did not increase in any of the three courses. Note that Calculus professors reported an anecdotal decrease in preparedness of trigonometric concepts for students placed using the algorithmic model. Denny et al. (2012) report that advisors found the model is “easy to use and appears to be effective” (p. 181).

The full level of success of any placement policy may not be solely attributable to the placement model. Placement policy changes may be one of many strategies that colleges enact in concert to increase success and completion. Both LBCC and NCCCS implemented multiple strategies for increasing the enrollment and success in Gateway mathematics and English courses in addition to placement policy. LBCC coordinates curricula alignment with local elementary and secondary schools (Willett, 2013) and invites fourth and fifth grade students from the feeder school districts to tour the college (Barnett et al., 2016). NCCCS individualized the ACCUPLACER exam, coordinates curricula alignment between the NCCCS and North Carolina high schools, is redesigning developmental education curricula, and is creating accelerated modules within the developmental education courses (Morrissey & Liston, 2012; Reeves Bracco et al., 2014).

The CHAID Method of Decision Tree Modeling in the Hierarchical Placement Model

The intent of including a discussion of the CHAID method of decision tree modeling in the literature review is to aid those without proficient knowledge of decision trees and how they may be used to determine cut scores in a hierarchical placement model. A description of the functionality of decision trees in determining cut scores might be beneficial to those charged with designing and implementing placement policy if they are considering a hierarchical multiple measure placement model.

Decision trees are a statistical tool that attempts to group members of a sample in increasingly homogenous subsets along the dependent variable (Hastie et al., 2009). Ramaswami and Bhaskaran (2010, p. 14) describe it as splitting the sample into segments such that "the variation of the response variable (categorical) is minimized within the segments and maximized among the segments." In this study, the response variable is success in an initial mathematics course. The CHAID method of decision tree modeling, as proposed by Kass (1980), utilizes the Pearson Chi squared statistic to merge and split the segments to form homogenous segments that are significantly different from one another. The Pearson Chi squared statistic compares the fit of observed categorical or ordinal data with what would be expected in each cell based on the total frequency of each category. Since data must be categorical or ordinal to use the Chi squared statistic, continuous data, such as HSGPA, must be separated into ordinal segments of similar size. Thus, the segments for HSGPA in this study will be defined by a connected range of HSGPA values with a similar number of elements in each segment. The ACTM values are defined as ordinal (instead of continuous) for the CHAID method so that each ACTM value is its own segment.

The formula for the Chi squared statistic is $X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$, where O_i is the observed frequency and E_i is the expected frequency. In the context of this study, the frequency of successes and failures for one set of placement measure values is compared to the frequency of successes and failures for another set of placement measure values. For example, suppose one segment is created by grouping all HSGPA values greater than 2.4 and less than or equal to 2.6 and another segment is created by grouping all HSGPA's greater than 2.6 and less than or equal to 2.7. Furthermore, suppose that the frequency of successes and failures is distributed as shown in Table 4.

Table 4

An Example of Calculating Chi Squared

	2.4 < HSGPA ≤ 2.6	2.6 < HSGPA ≤ 2.7	Total
Success	15 (18.8)	23(19.2)	38
Failure	27(23.2)	20(23.8)	47
Total	42	43	85

The expected frequencies are in parenthesis. The expected frequency for the first cell can be calculated as $\frac{(38)(42)}{85} = 18.8$.

The value used to calculate the Chi squared statistic for the first cell is $\frac{(15-18.8)^2}{18.8} = .77$. The Chi squared statistic is the sum of these values for all four cells and is 2.75 in this example. A p-value is calculated for each Chi squared value using the degrees of freedom value to standardize the calculations. The p-value is interpreted as the percent likelihood that the variables are independent; that is, the distribution of the data is due to chance. For example, the interpretation of a p-value of .10 is that there is a 10% probability that the variation is due entirely to chance. Traditionally, p-values less than .05 (5%) are deemed significant but this critical value can be adjusted as appropriate.

The CHAID method of decision tree modeling calculates a Chi squared statistic and p-value for each pair of adjacent, ordinally-ordered segments. The pair of segments whose p-value is least significant on a Chi squared test are merged. Then, the Chi squared and p-value for each pair of segments are recalculated (only the values for the pairs of segments adjacent to the newly merged segments need recalculated) and the pair of segments whose p-value is least significant on a Chi squared test are merged. This process continues until no non-significant Chi squared values exist based on a predetermined critical alpha-value of significance. Any merged segments comprised of three or more original segments are analyzed for any natural breaks by performing a Chi squared test within the merged section. If the Chi squared value of any two subsegments is significant, the section is split along that break. If this occurs, the merging and splitting process is performed again until "optimally" merged segments are formed that are significantly different from one another using a Chi squared test. This process of merging and splitting of segments to form "optimally" merged segments is performed for each predictor variable (Kass, 1980; Wilkinson, 1992).

Once "optimally" merged segments are formed for each predictor variable, a Chi squared value is calculated using the optimally merged segments. The predictor variable that yields the greatest Chi squared value splits the original sample into the "optimally" merged segments for that predictor variable. These split segments are referred to as nodes. The entire process is repeated on the subset of the sample in each node until no significantly different segments are found or until the tree reaches a predetermined number of levels (Kass, 1980; Wilkinson, 1992). At this point, the decision tree is considered formed. The break points in the predictor variables that define the different segments can be interpreted as cut point values for predictor variables in determining placement and will be used as such in the Hierarchical model.

Binary Logistic Regression in the Algorithmic Placement Model

The intent of including a discussion of binary logistic regression in the literature review is to aid those without proficient knowledge of logistic regression in utilizing an algorithmic placement model. A description of the functionality of logistic regression in the context of placement and course success might be beneficial to those charged with designing and implementing placement policy if they are considering an algorithmic multiple measure placement model.

Logistic regression models are designed to form the maximum likelihood relationship with, or best predict, a categorical outcome. In this study, the categorical outcome is binary: initial college mathematics course success or course failure. Logistic regression does not require data or errors to be normally distributed with equal variances and covariances, as is assumed in linear regression (Peng & So, 2002b). This is helpful since the outcome variable and its error are binomially distributed. For example, each student within the sample will either be successful or not successful in their initial mathematics course. Under linear regression, it would be assumed that the distribution of the outcome variable, course success, would be normal and violations of assumptions in modeling often lead to inaccurate results. Since logistic regression does not have assumptions of normality, it can often more accurately predict a categorical outcome than other forms of regression (Peng & So, 2002b). Hosmer et al. (2013) state that:

> The logistic regression model is remarkably flexible. Unless we are dealing with a set of data where most of the probabilities are very small or very large, or where the fit is extremely poor in an identifiable systematic manner, it is unlikely that any alternative model will provide a better fit (p. 200).

Binary logistic regression transforms the results of a linear regression equation into an outcome that only exists between 0 and 1 to align with the yes (1) or no (0) possibilities of a dichotomous outcome variable. A probability cut point is chosen (often at .50, or 50%) and values are entered into each predictor variable of the linear equation. Any transformed value of the linear equation that is greater than or equal to the probability cut point is assigned a predicted value of 1 (yes, or success) and any transformed value of the linear equation that is less than the probability cut point is assigned a predicted value of 0 (no, or not a success) (Hosmer et al., 2013). Suppose, for example, that the probability cut point is set at 0.50 and values of 3.00 and 21 are entered into a linear regression equation containing predictor variables of HSGPA and ACTM, respectively. If the transformed value of the linear equation is greater than or equal to 0.50, the model predicts that a student with a HSGPA of 3.00 and an ACT-Math score of 21 will successfully complete his/her initial mathematics course in college. Similarly, if the transformed value of the linear regression equation is less than 0.50, the model predicts that a student with a HSGPA of 3.00 and an ACT-Math score of 21 will not successfully complete his/her initial mathematics course. The probability cut point can be adjusted to maximize the accuracy, sensitivity, and/or specificity of the sample using comparisons of sensitivity to specificity and using a Receiver Operating Characteristic (ROC) curve (Peng & So, 2002b).

As mentioned above, the range of values of a linear regression equation is from negative infinity to infinity and must be transformed to a range between 0 and 1. The transformation of these values involves the taking the natural logarithm of the odds of an event occurring. Let the logistic regression equation (for one independent variable x and coefficients B_0 and B_1 for ease of notation) be denoted $g(x) = B_0 + B_1x$. This linear regression equation has a range of negative infinity to infinity. Taking e to the power of g(x) limits the range to 0 to infinity since e to the power of any number is greater than 0. Finally, divide $e^{g(x)}$ by $1 + e^{g(x)}$ and denote this as $\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}}$. Note that as $e^{g(x)}$ approaches 0, $\pi(x)$ approaches 0, and as $e^{g(x)}$ approaches infinity, $\pi(x)$ approaches 1. Thus, this is a transformation of values between negative infinity and infinity to values between 0 and 1. It can be shown that $\pi(x)$ is the estimated proportion, or probability, of outcomes that will be assigned a value of yes, or 1, given a value of the predictor variable x (Hosmer et al., 2013). In this study, a value of yes, or 1, indicates success in an initial mathematics course whereas a value of no, or 0, indicates a failure in an initial mathematics course.

Consider the following three examples of the probability of Y given a value of x: 0.5, 0.75, and 0.1. If the probability of a student's success is $\pi(x) = 0.5$ (or 50%) given a value of x, then $0.5 = \frac{e^{g(x)}}{1+e^{g(x)}}$. Multiplying both sides by the denominator results in $0.5 + 0.5e^{g(x)} = e^{g(x)}$. Combining like terms gives $0.5 = 0.5e^{g(x)}$. Dividing by 0.5 on each side gives $1 = e^{g(x)}$ and taking the natural log of both sides results in $0 = g(x)$. Therefore, a linear regression value of $g(x) = 0$ is transformed into a probability of 50%.

Similarly, if the probability of a student's success is $\pi(x) = 0.75$ (or 75%) given a value of x, then $0.75 = \frac{e^{g(x)}}{1+e^{g(x)}}$. Multiplying both sides by the denominator gives $0.75 +$

$0.75e^{g(x)} = e^{g(x)}$. Combining like terms gives $0.75 = 0.25e^{g(x)}$. Dividing by 0.25 on each side gives $3 = e^{g(x)}$ and taking the natural log of both sides results in 1.10 ≈ g(x). Therefore, a linear regression value of g(x) = 1.10 is transformed into a probability of approximately 75%.

Finally, if the probability of a student's success is $\pi(x)$ = 0.10 (or 10%) given a value of x, then 0.10 = $\frac{e^{g(x)}}{1+e^{g(x)}}$. Multiplying both sides by the denominator gives $0.10 + 0.10e^{g(x)} = e^{g(x)}$. Combining like terms gives $0.10 = 0.90e^{g(x)}$. Dividing by 0.90 on each side gives $0.111 \approx e^{g(x)}$ and taking the natural log of both sides results in -2.20 ≈ g(x). Therefore, a linear regression value of g(x) = -2.20 is transformed into a probability of approximately 10%. Using these formula, any value of g(x) from negative infinity to infinity can be transformed into a probability between 0 and 1 with g(x) = 0 corresponding to a 50% probability, g(x) < 0 corresponding to probabilities less than 50%, and g(x) > 0 corresponding to probabilities greater than 50%.

The odds of an event is defined as the probability of the event occurring, $\pi(x)$, divided by the probability of the event not occurring, $1 - \pi(x)$. Thus, the odds of an event is denoted $\frac{\pi(x)}{1-\pi(x)}$. It can be shown that the natural log of the odds is equal to the linear regression equation g(x) = B_0 + B_1x. Notice that since $(x) = \frac{e^{g(x)}}{1+e^{g(x)}}$, we have that the probability of an event not occurring, $1 - \pi(x)$, is $1 - \pi(x) = 1 - \frac{e^{g(x)}}{1+e^{g(x)}} = \frac{1+e^{g(x)}-e^{g(x)}}{1+e^{g(x)}} = \frac{1}{1+e^{g(x)}}$. Therefore, the natural log of the odds, $\ln\left(\frac{\pi(x)}{1-\pi(x)}\right)$, is $\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \ln\left(\frac{\frac{e^{g(x)}}{1+e^{g(x)}}}{\frac{1}{1+e^{g(x)}}}\right) = \ln\left(e^{g(x)}\right) = g(x)$. Since $\frac{\pi(x)}{1-\pi(x)}$ is the odds of an event given a value of x and we see that the logistic regression equation, g(x), is equal to the natural logarithm of the odds, the logistic regression equation is often referred to as the log odds equation (Hosmer et al., 2013). The term log odds will be used to mean the natural logarithm of the odds of an event occurring from this point forward.

Determining the values of the coefficients of each predictor variable in the logistic equation to maximize the likelihood that a dichotomous outcome occurs involves solving a system of differential equations and is often done via computer programs. However, once found, the coefficients form a linear relationship with the outcome variable (Hosmer et al., 2013). The slope coefficient, B, for a logistic regression is the amount of change in the log odds for a change of 1 in the predictor variable. While reporting the log odds is standard practice in a journal article, practical applications may be easier understood after transforming log odds to probability. For example, consider a slope coefficient B-value of 0.3 for the covariate ACT-Math score. This means that the log odds of success in the course change by 0.3 for each change of 1 in the student's ACT-Math score. If the other variables in the model are held constant, an increase of 1 in the ACT-Math score also increases the log odds of the linear regression equation by 0.3. In other words, g(x) increases by 0.3.

Converting this to probability may make the application of this change easier to understand. For example, consider the hypothetical univariate logistic equation g(x) = -5.5 + 0.3(x). In this example an ACTM score of 19 would yield g(19) = 0.2 and an ACTM score of 20 would yield g(20) = 0.5. We see from the earlier equations that $g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right)$ so then $e^{g(x)} = \frac{\pi(x)}{1-\pi(x)}$. We can calculate that $e^{0.2} = 1.22$ and $e^{0.5} = 1.65$ which are the odds ratios for course success at each value of ACT-Math score 19 and 20, respectively. Converting odds ratios to probability we have that $1.22 = \frac{\pi(19)}{1-\pi(19)}$ becomes $1.22 - 1.22\pi(19) = \pi(19)$ after multiplying both sides of the equation by the denominator. Combining like terms in the equation results in $1.22 = 2.22\pi(19)$. Solving for $\pi(19)$ we have $\pi(19) = \frac{1.22}{2.22} = 0.55$. Thus, it is estimated that 55% of students with an ACTM score of 19 will successfully complete the initial mathematics course. Similarly, $\pi(20) = \frac{1.65}{2.65} = 0.62$. So an increase in the ACT-Math score from 19 to 20 would increase the estimated proportion (probability) of those successfully completing the mathematics course from 55% to 62% for the sample.

If the probability cut point is set at 60% and ACT-Math score is the only independent variable, as in this example, an ACT-Math score of 19 would predict that the student would be unsuccessful since a 55% estimated proportion of success falls below the 60% threshold while an ACT-Math score of 20 would predict that the student would be successful since a 62% estimated proportion of success is above the 60% threshold. Note that the interpretation of probability is the estimated proportion of students successfully completing the mathematics course for the sample given the value of each predictor variable and not the probability for an individual's success in the course. However, the predicted outcome, success or failure, for each individual student is based on a comparison of the estimated proportion of success to the probability cut point for those covariate values (Hosmer et al., 2013 p. 18).

In the above hypothetical example, the probability cut point was set to 60%. If we set the probability cut point at an arbitrarily low value, such as 10%, then the sensitivity of our model will be high because every student who is actually successful would have likely been predicted to be successful. However, the specificity of the model will be low because many students who were not actually successful would have also been predicted to be successful.

This would result in a large false positive rate or Type I error. Conversely, if the probability cut point is set arbitrarily high, such as 90%, then the sensitivity of the model will be low and the specificity will be high. Many students who actually succeeded would have been predicted to not succeed leading to a large false negative rate or Type II error. A comparison of sensitivity to specificity as functions of probability cut point values allows the researcher to determine the optimum probability cut point at the intersection of the two functions. The area under the ROC curve, a comparison of sensitivity to 1 – specificity, is a statistical measure of the fit of the model to the data.

The value of the coefficient, B, in the logarithmic equation is also related to the ability of the equation to differentiate between an estimation of 1 or 0. The shape of a log odds equation is an S-curve with a steep increase somewhere in the middle of the curve that approaches 0 and 1 at each end of the S-curve, respectively. The steeper the slope during the increase the better the equation is at differentiating between 1 or 0 because a steeper slope means that the curve will approach the extremes of the graph more quickly. The greater the coefficient, B, of the variable, the greater the effect a change of 1 unit of the independent variable will have on the value of $g(x)$, and the greater the slope will be. Note that beta values must be standardized to compare the strength of one beta-value to another. The larger the absolute value of the standardized beta, the more predictive ability the corresponding variable has in determining the outcome variable.

CHAPTER THREE

METHODOLOGY

Overview of the Study

The difference between success and failure in an initial mathematics course may be affected by a student's placement along the mathematics sequence (Bailey et al., 2010; Edgecombe, 2011; Scott Clayton & Rodriguez, 2012). Ideally, colleges place students in the highest course needed for their degree path in which they can succeed given a dedicated approach to coursework. Dedication to learning does not always occur but giving students the best opportunity for success is the primary goal of placement. Students who are severely misplaced succeed at lower rates than correctly placed students (Scott-Clayton et al., 2014). Even underplacing by one course increases the amount of time a student needs to achieve academic goals and may unnecessarily use financial aid or out-of-pocket money.

Accurate placement has therefore been put at the fore of the college completion discussion (Bailey et al., 2010; Belfield & Crosta, 2012; Burdman, 2012; Reeves Bracco et al., 2014; Hughes & Scott-Clayton, 2011; Maruyama, 2012; Missouri Department of Higher Education, n.d.; Scott-Clayton et al., 2014). Research has shown that multiple measures, including high school transcript data, are more correlated to student success than any single measure (Bahr et al., 2014; Geiser & Santelices, 2007; Medhanie et al., 2012; Ngo & Kwon, 2015; Rothstein, 2004; Scott-Clayton, 2012). A lack of correlation to success is especially true when the single measure is a standardized placement test – a common practice for non-exempt students in many community colleges in the United States (Armstrong, 2000; Fields & Parsad, 2012; Hughes & Scott-Clayton, 2011; Mattern & Packman, 2009; Roska et al., 2009). The need for accurate placement using multiple measures has guided the purpose of this study, which was to analyze placement and success for one large community college and to develop and compare multiple measure placement models based on the data. A hierarchical model and an algorithmic model were created to reflect the multiple measure placement models currently used at colleges in the United States (Barnett et al., 2016; Reeves Bracco et al., 2014).

A non-experimental design was used to study placement at the community college so that no ethical issues arose in the placement of students as might occur in an experimental study. Data was collected from two consecutive historical samples, an initial sample and a validation sample. Two multiple measure placement models, an algorithmic model and a hierarchical model, were created based on the initial sample. The validity of the placement models were analyzed and compared on their ability to accurately place students and to differentiate between successful and unsuccessful students across the five mathematics course levels.

The dependent variable in the study was success or failure in a student's initial mathematics course at Metropolitan Community College. The variables were coded 1 = success and 0 = failure in which success includes a grade of A, B, C, or S (Successful) and failure includes a grade of D, F, U (unsuccessful), or W (withdrawal). A binary distribution of success and failure were given for each course level.

The independent variables, or placement measures, in the study were cumulative high school grade point average (HSGPA) and the mathematics score on the ACT exam (ACTM). Both variables were continuous (scaled) with HSGPA ranging from 0.0 to approximately 4.5 and ACTM ranging from 0 to 36. A histogram showing the distribution of the covariates was given even though it is not necessary in logistic regression to have normally distributed covariates (Peng, Lee, et al., 2002). The ACTM variable was recoded as ordinal in the Hierarchical model to ensure merging and splitting along integer values. Other variables such as socioeconomic status (SES), ethnicity, and gender were considered for inclusion in the models and, while these background characteristics might be effective in predicting success (Kuh et al., 2008; Rothstein, 2004), using these variables to place students in any practical sense creates issues of equality. Thus, SES, ethnicity, gender, and other background variables were not made available as possible covariates in either model.

Sampling, Data Collection, and Data Organization

Student data was collected using a data collection query on the operating system of Metropolitan Community College (MCC). Metropolitan Community College is a five-campus system with campuses in urban and suburban regions of the Kansas City, Missouri, metropolitan area with an overall enrollment of about 18,000 students (Metropolitan Community College, 2017). The query collected data on cumulative high school grade point average (HSGPA), ACT-Math subtest scores (ACTM), college mathematics courses enrolled and the corresponding term in which the student enrolled, the grade in the mathematics courses, and college student identification number. Data was collected using cohort sampling from a sixteen-semester sample of students and a six-semester sample of students placing and enrolling at Metropolitan Community College. The initial sample was collected from fall 2012 to fall 2017 while the validation sample was collected from spring 2018 to fall 2019 and will be non-overlapping. Only the initial sample was used to construct the placement models. The validation sample served as a tool for assessing model validity in predicting success on a new sample and the level of agreement of the models in placing students. Originally, the initial sample was to be collected from fall 2012 to fall 2015 while the validation sample was to be collected from spring 2016 to fall 2017. However, when data became available through fall 2019, it was included in the data collected with corresponding adjustments to the initial and validation sample.

Only student records with HSGPA, ACTM score, and course grade were included in the samples and were separated into files according to the initially placed course: Calculus, Trigonometry, Gateway, Developmental Math 1 (DM1) and Developmental Math 2 (DM2) (Refer to Table 5). Course grade data were recoded into a binary scale to determine success or failure. A course grade record of A, B, C, or S was coded into 1 for success and a course grade record of D, F, W, or U was coded into 0 for failure. If multiple entries existed for one data point for a student, the data was delimited according to the first enrollment term. For example, if a record showed that a student enrolled in DM2 in fall 2014 with an unsuccessful grade and enrolled in DM2 again in spring 2015, the fall 2014 unsuccessful data was used as this is the course in which the student placed. If multiple data exist for placement then only the highest placement score for each measure and the initially placed course were used for analysis. This aligns with the placement policies of MCC. For example, if a student had an ACTM score of 17 and an ACTM score of 19, the score of 19 was used in the data analysis. Any student record that does not include HSGPA, ACTM, or grade in the initially placed course was removed from the data set. Any placement score that was not the highest score and any course that was not the initially-placed course was also removed from the data set. Additionally, any HSGPA outside the range of 0.0 to 4.5 and any ACTM score outside the range of 0 to 36 was removed from the data set. Data was analyzed using SPSS software.

Table 5
MCC Mathematics Course Sequence

Calculus
Trigonometry
Gateway
Developmental Math 1 (DM1)
Developmental Math 2 (DM2)

Creating the Algorithmic Placement Model

The algorithmic placement model used binary logistic regression equations from student data in the initial sample to create an algorithm that best predicts success or failure in each mathematics course level beginning with the highest level and excluding the student data from the successive calculations that fit into the highest level. A unique logistic regression equation (or logit) was created for each mathematics course level except DM2 based upon the success data of that course level. The collection of four logistic regression equations are termed the algorithmic model. A prediction of success in the logistic regression equation corresponding to the highest mathematics course level was the predicted placement of the student. Predicted success or failure was determined by the value of the logit compared to the probability cut score with a logit value greater than the probability cut score denoting a prediction of success in the course level. A Receiver Operator Characteristic (ROC) curve comparing sensitivity to 1 – specificity was used to determine the maximum probability that the model correctly identified the students who would succeed in the course. This occurred at the value of the probability cut score that maximized the area under the ROC curve and hence, maximized the discrimination of the model using the Pythagorean Method. The probability cut score was set at this maximizing value and varied across the algorithm for each course level to not bias one course level equation over the other. A further description of logistic regression and probability cut scores can be found in the Logistic Regression section of the literature review.

Logistic regression requires that continuous variables be linearly related to the logit. This was analyzed using fractional polynomial analysis and adjusted accordingly to maximize model fit and, hence, placement accuracy. Each nonlinear covariate was adjusted accordingly using the simplest model that was not significantly different from the two-term fractional polynomial model. That is, if the linear model was not significantly different from the two-term fractional polynomial model, the linear model wasused (i.e., the covariate did not change). If the linear model was significantly different (not as good) from the two-term fractional polynomial model and the one-term fractional polynomial model was not significantly different, the covariate was adjusted according to the best one-term fractional polynomial model given by the test. Once the main-effect covariates were established, an interaction variable was built from the covariates and tested for significance using the Wald statistic and $p < .05$ for the interaction. The interaction term remained in the model if it was significant and removed if it was not significant. The resulting model was the full model.

Only two independent variables, HSGPA and ACTM, were considered for the study and were included in each logistic regression equation regardless of their significance. However, analyzing the significance of individual predictors and the fit of the regression equations added to the research on student placement and created a replicable model for future research. Individual predictor variables, or covariates, were evaluated using the Wald statistic to determine whether a significant relationship existed with the outcome variable, success in the initial mathematics course, controlling for other variables. An odds ratio for each predictor variable was given. The predictor variables were compared using partially standardized beta values and the change in the Chi Squared statistic of the Likelihood Ratio Test by adding the variables one at a time to the logistic regression analysis. The fit of each full model was evaluated against the null model using the Likelihood Ratio Test and Pseudo R^2 values. The Likelihood Ratio Test assessed whether the full model fits the model significantly better than the null model (the constant only model). Pseudo R^2 values are similar to explained proportion of variance for linear regression models with values closer to zero indicating no explained variance by the model and values near one indicating a large portion of the variance explained by the model.

Creating the Hierarchical Placement Model

The hierarchical placement model used student data from the initial sample to determine cut score values for HSGPA and ACTM scores – values that, if met or exceeded, placed a student in a certain course level. The highest course level cut points were determined first and the student data included in that course level were excluded from successive calculations. Cut scores for HSGPA, ACTM, or a combination of the two measures were determined using the Chi-squared Automatic Interaction Detection (CHAID) method of decision tree modeling with a maximum of two levels. Branches of the decision tree were included in the model if the branch differentiated success and failure at a rate greater than the rate of success for the course (or rate of failure if it was greater than the rate of success) and showed a significant differential effect as evidenced by a corresponding p-value less than .01. Decision tree modeling allowed for a statistical method of choosing cut points specific to the sample by processing the data through the CHAID algorithm that successively grouped students into increasingly homogenous subsets (Hastie et al., 2009). The goal in this study wass to divide the sample along the dependent variable, success in an initial mathematics course, using values of ACTM and HSGPA. The CHAID method used the Pearson Chi Squared algorithm to split the sample into the subsets and wasvutilized in this study because it allowed for multiple splits (i.e., multiple subsets) at each level and because the C4.5 and C5 methods of decision tree modeling were not available on SPSS without the additional SPSS Modeler feature. A further description of the CHAID method can be found in the CHAID Method of Decision Tree Modeling section of the literature review. The collection of cut scores for each of the five course levels constitutes the hierarchical placement model. The hierarchical model does not have statistical measures of fit or methods to measure the predictive ability of individual covariates, as in the algorithmic model. However, accuracy and discrimination could be and was calculated for both the hierarchical model and the algorithmic model to determine model validity.

Comparing the Validity of the Algorithmic and Hierarchical Placement Models on the Initial Sample

For each course level, the algorithmic model was compared to the hierarchical model in terms of accuracy and discrimination (c-statistic defined as the area under the ROC curve). While sensitivity and specificity was measured, these measures are best judged in tandem and should not be used to determine model fit since "they depend heavily on the distribution of the estimated probabilities in the sample" (Hosmer et al., 2013, p. 171). The accuracy of a model was defined as the sum of the number of correctly predicted successes and the number of correctly predicted failures divided by the total outcomes. The accuracy percentage for each model was compared directly without a statistical test to evaluate the significance of the difference in accuracy. One model was determined to be more accurate than another model if it tended to have greater accuracy rates across the course levels for both the initial and validation samples.

Discrimination was given by the area under the ROC curve and interpreted as the ability of a model to differentiate between a pair of opposite performing students. For example, consider a model with a discrimination value of .75. If a pair of students is selected with different outcomes, one success and one failure, there is a 75% probability that the model will predict the successful student to be successful. The significance of the difference of the areas can be determined by evaluating the significance of the value of the critical ratio test, Z, against areas under the normal probability density function according to the method set by Hanley and McNeil (1983). The critical ratio test, Z, is defined as $Z = \frac{A_1 - A_2}{\sqrt{SE_{A_1}^2 - SE_{A_2}^2 - 2rSE_{A_1}SE_{A_2}}}$, where A_1 and A_2 are the respective areas under the ROC curves (Pearce & Ferrier, 2000, p. 242). The difference in area under the ROC curves were calculated using an online calculator by Lowry (http://vassarstats.net/roc_comp.html, accessed 2017) since SPSS does not offer a way to compare the differences in area under two ROC curves at this time. If the areas under the ROC curves were significantly different from each other, one model would consistently differentiate success from failure better than the other model. Note that if the ROC curves intersected, the better discriminating model would only be better for a specific portion of probability cut points (Pearce & Ferrier, 2000).

Comparing the Agreement of the Algorithmic and Hierarchical Models on the Initial Sample and the Validation Sample

The final comparison of the placement models were comparisons of the agreement of placement patterns of the algorithmic and hierarchical models on both the initial sample and the validation sample. Every student was hypothetically placed in one of the five course levels by the algorithmic model and by the hierarchical model. The placement patterns were analyzed to determine the level to which the models agree on placement. Agreement of placement was defined as both models placing a student in the same course level whereas a discrepancy in placement meant that the models placed a student in different course levels. Varying degrees of discrepancy can exist. For example, placement in Trigonometry by the algorithmic model and placement in Gateway by the hierarchical model would be a discrepancy of one degree while placement in Trigonometry by the algorithmic model and placement in Developmental Mathematics 1 (DM1) by the hierarchical model would be a discrepancy of two degrees. The agreement in placement between the two models was determined using the weighted Kappa statistic, which can account for degrees of discrepancy. A weighted Kappa of 0 indicated that the models do not agree on placement on a level greater than what could be expected by chance whereas a weighted Kappa close to 1 indicated substantial agreement between the models (Landis and Koch, 1977).

The goals of these comparison tests were to compare the placement models on how accurately they predicted success in each level of mathematics course and the level of agreement between the models on the placement of students using a validation sample. A diagram of the tests of validity and agreement were given in Table 6. Using this battery of comparative tests gave a more complete comparison of the models than any one test alone. If one model outperformed the other model, then perhaps the better performing model should be considered by colleges to place students or considered for further research. If both models predicted success similarly then perhaps the model that is easier to implement should be considered by colleges to place students or for further research.

Table 6

The Validity, Reliability, and Agreement Tests to be Used in the Study

Internal Validity for the Algorithmic and Hierarchical Models		
	What is Measured	**How it is Measured**
Algorithmic Model	Accuracy	Direct Comparison to Course Success
	Discrimination	Area Under the ROC Curve
	Model Fit	Likelihood Ratio Test and Pseudo R^2
	Predictive Ability of Covariates	Wald Statistic, Partially Standardized Beta Values, and Change in Chi Squared for the Likelihood Ratio Test
Hierarchical Model	Accuracy	Direct Comparison to Course Success
	Discrimination	Area Under the ROC Curve
Comparisons of Internal Validity Between the Algorithmic and Hierarchical Models		
	What is Measured	**How it is Measured**
	Accuracy	Direct Comparison Between Placement Models
	Difference in Discrimination	Critical Ratio Z-Test
Agreement of Placement Between the Algorithmic and Hierarchical Models		
	What is Measured	**How it is Measured**
	Agreement on the Validation Sample	Weighted Kappa

Null Hypotheses

There were five null hypotheses for this study that aligned with the research questions. Research question one asked about the predictive power of the hierarchical model. The first set of null hypotheses were that the discrimination of the placement rules for each of the five course levels were equivalent to .50, or chance, and the accuracy was equivalent to the maximum of the percentage of students that succeed or the percentage of students that fail. Hosmer et al. (2013) give a generic guideline for level of discrimination. Levels of discrimination between 0.5 and 0.7 indicate that the model has poor discrimination, levels between 0.7 and 0.8 indicate acceptable discrimination, levels between 0.8 and 0.9 indicate excellent discrimination, and levels above 0.9 indicate outstanding discrimination (p. 177). For accuracy, chance is equal to the highest percentage of success or failure. In other words, if one was forced to guess whether a student would succeed without being given any information, one would choose the highest percentage out of success or failure and would be correct at a rate equal to that percentage given sufficient trials. Deviation from chance indicated an ability of the model to accurately predict initial course success above that rate. Because accuracy is situationally dependent, the significance of accuracy was only determined from a practical perspective.

Research question two asked about the predictive power of the algorithmic model. The second null hypothesis evaluated the predictive power of the algorithmic model in terms of statistical measures of fit, accuracy, and discrimination. Statistical measures of fit to used were Pseudo R^2 values and the Likelihood Ratio Test. Pseudo R^2 values ranged from zero to one with values near one indicating a large portion of the variance explained by the model. The Likelihood Ratio test compared the fit of the full model to the null model (the constant only model) using $G = -2(LogLikelihood_{newmodel} - LogLikelihood_{nullmodel})$ with p-values less than .05 indicating that the full model significantly fit the data better than the null model. Discrimination and accuracy were measured in the same manner as in the hierarchical model. Thus, the second set of null hypotheses were that Pseudo $R^2 = 0$, G = 0 where G is the value of the Likelihood Ratio statistic, discrimination equaled 0.50 (50%), and accuracy equaled the maximum of the percentage of success or the percentage of failure for each course level of the algorithmic model.

The third research question asked about the differences in predictive ability between the algorithmic and hierarchical models. The measures of predictive ability that could be compared were accuracy and discrimination (the area under the ROC curves). Thus, the third set of null hypotheses were that the difference in accuracy of the two models equaled 0 and the difference in discrimination of the two models equaled 0 for each course level. The difference in accuracy of the two placement models was compared directly without a statistical test to evaluate the significance of the difference in accuracy. The significance of the difference between the discrimination of two models were determined by evaluating the significance of the value of the critical ratio test, Z, against areas under the normal probability density function.

The fourth research question asked about the differences in placement results of the algorithmic and hierarchical models using the validation sample. In other words, did the two models generally place the same students in the same course levels? Using a weighted Kappa statistic in which a value of zero indicated that the agreement of placement was not better than chance, the level of agreement between the algorithmic and hierarchical models were calculated and analyzed from a practical perspective. Thus, the fourth null hypothesis was that the weighted Kappa for comparing the agreement of the two models was equal to 0, or chance, for the validation sample.

The fifth research question asked about the predictive strengths of the covariates HSGPA and ACTM in the binary logistic regression equations for each course level. The predictive strength was measured by the value of the coefficient of each variable. Since four of the five course levels (all except DM2) have binary logistic regression equations, the fifth set of null hypotheses was that $b_{ij} = 0$ with i = 1, 2, 3, and 4 representing each of the four course levels with logistic regression equations and j representing the various main effect and interaction variables present in each equation. The difference from zero for each beta coefficient was determined by the Wald statistic having a p-value less than .05. Additionally, the strength of each variable in each logistic regression equation was determined and compared by directly comparing partially standardized beta values and by the significance of the change in the Chi Squared statistic of the Likelihood Ratio Test when each variable was added one at a time to the logistic regression analysis. These tests attempted to give insight to which variable added more predictive strength to the model for particular course levels.

These five null-hypotheses were designed to answer the five research questions in this study. Taken together, the research questions sought to determine which multiple measure placement model more accurately placed students and determine the differences in the course placement of students. Using a non-experimental design and data from an initial sample and a validation sample, the study attemped to statistically analyze and compare the predictive ability of a hierarchical and algorithmic multiple measure placement model.

CHAPTER FOUR

RESULTS AND ANALYSIS

Overview of the Study

The purpose of this study was to create a hierarchical placement model and an algorithmic placement model using historical student data and compare the placement models based on accuracy, discrimination, and placement of students. The placement measures in the models included cumulative high school grade point average (HSGPA) and ACT-Math score (ACTM). Data were collected from Metropolitan Community College (MCC), a large urban and suburban community college in the Midwest United States from 2012 to 2019 and separated into an initial sample (fall 2012 to fall 2017) and a validation sample (spring 2018 to fall 2019). The models were built based on the initial sample data using CHAID decision tree modeling for the hierarchical model and binary logistic regression for the algorithmic model. Receiver Operator Curve (ROC) analyses were conducted on the algorithmic model to determine cut points to maximize sensitivity and specificity for each mathematics course level equation using the Pythagorean method. The course levels were Calculus, Trigonometry, Gateway Mathematics, Developmental Mathematics 1 (DM1), and Developmental Mathematics 2 (DM2). The accuracy of each placement model for each course level was compared on the initial sample and the validation sample. The discrimination of each placement model was compared on the initial sample. Additionally, simulated placement for each student in the validation sample was conducted for each model and the results were analyzed and compared for agreement in placement using weighted kappa. Finally, using the binary logistic regression equations of the algorithmic placement model, the significance of HSGPA and ACTM in the model was determined.

Data Collected

The data collected from the institutional research office at MCC included student ID, ethnic group, gender, ACTM score, HSGPA, math course enrolled, term of math course enrolled, and grade in math course. Student ID was needed to identify data specific to each individual as required to perform the statistical analyses. The data were organized to utilize only the initial mathematics course and grade and greatest ACTM score and HSGPA. Any student data that did not include both measures of placement were removed from the sample. Student ID's were then permanently deleted from the data to protect the identification of students. There were 10,525 distinct records of data for students in the initial sample and 3088 distinct records data for students in the validation sample. Each sample was further separated based on mathematics course level into the following: Calculus, Trigonometry, Gateway Mathematics, Developmental Mathematics 1 (DM1), and Developmental Mathematics 2 (DM2). In the initial sample, there were data for 243 students in Calculus, 247 students in Trigonometry, 4003 students in Gateway, 4268 students in DM1, and 1764 students in DM2. In the validation sample, there were data for 85 students in Calculus, 113 students in Trigonometry, 1086 students in Gateway, 1177 students in DM1, and 627 students in DM2. Table 7 shows the descriptive statistics for HSGPA, ACTM, and course success (defined as a grade of C or higher) for each course level of each sample. Ethnicity data and gender data were not analyzed in this study but are available for the initial sample in

Table 8 and for the validation sample in Table 9. An analysis of placement along the variables ethnicity and gender in each model would be valuable research for further study.

Table 7

Descriptive Statistics for HSGPA, ACTM, and Course Success for the Initial Sample and Validation Sample

		N	Mean* HSGPA	Mean ACTM	Mean Course Success Rate
Combined Sample	Total	13613	3.048	19.90	58.3%
Initial Sample	Total	10525	3.061	20.08	60.9%
	Calculus	243	3.581	27.00	67.1%
	Trigonometry	247	3.471	24.18	64.4%
	Gateway Math	4003	3.280	23.16	63.0%
	DM1	4268	2.975	18.37	59.1%
	DM2	1764	2.641	15.72	58.7%
Validation Sample	Total	3088	3.004	19.30	49.6%
	Calculus	85	3.464	26.61	44.7%
	Trigonometry	113	3.217	24.54	38.9%
	Gateway Math	1086	3.265	22.34	51.8%
	DM1	1177	2.917	17.57	50.2%
	DM2	627	2.612	15.35	47.4%

*represented on a 4-point scale

Table 8

Descriptive Statistics for HSGPA, ACTM, and Course Success across Ethnicity and Gender for the Initial Sample

		N	Mean* HSGPA	Mean ACTM	Mean Course Success Rate
Total		10525	3.061	20.08	60.9%
Gender	Female	5641	3.128	19.34	65.1%
	Male	4878	2.983	20.94	56.0%
	Not Specified	6	3.128	23.33	66.7%
Ethnicity	Asian	211	3.241	20.12	70.1%
	Black	1154	2.685	17.25	48.8%
	Hispanic	1040	2.972	19.07	58.7%
	Native American	31	2.839	19.13	51.6%
	Not Specified	126	3.058	20.61	61.9%
	Pacific Islander	27	3.009	20.19	51.9%
	Two or More	767	2.985	20.32	56.7%
	White	7169	3.140	20.72	63.3%

*represented on a 4-point scale

Table 9

Descriptive Statistics for HSGPA, ACTM, and Course Success across Ethnicity and Gender for the Validation Sample

		N	Mean* HSGPA	Mean ACTM	Mean Course Success Rate
Total		3088	3.003	19.30	49.6%
Gender	Female	1701	3.082	18.64	54.9%
	Male	1387	2.908	20.11	43.2%
Ethnicity	Asian	72	3.201	19.25	68.1%
	Black	427	2.682	16.48	35.6%
	Hispanic	431	2.897	18.13	45.5%
	Native American	6	3.005	19.50	66.7%
	Not Specified	45	3.249	20.49	62.2%
	Pacific Islander	5	3.094	22.40	100.0%
	Two or More	272	2.852	18.99	41.5%
	White	1830	3.004	20.24	53.8%

Preliminary Data Analysis

Before building the placement models, preliminary data analysis was performed on the data for each course level in both the initial and validation samples. Frequency tables were created to show the means of HSGPA, ACTM, and course success (coded success = 1 and not success = 0). As shown in Table 6, the mean HSGPA's was 3.061 for the initial sample and 3.004 for the validation sample. The mean ACTM score was 20.08 for the initial sample and 19.30 for the validation sample. The mean course success, defined as achieving a C or higher in graded courses or Satisfactory (S) in ungraded courses, was 0.609 (60.9%) for the initial sample and 0.496 (49.6%) for the validation sample. Independent samples t-tests were used to compare the two samples in terms of HSGPA and ACTM to determine if significant differences existed between the means. Both HSGPA and ACTM were reasonably normally distributed for both the initial and validation sample thus allowing the use of the independent samples t-test (Refer to Appendix A for histograms and Normal Q-Q plots). Course success data are binary (coded success = 1 and not success = 0) and therefore not normally distributed. The nonparametric Mann-Whitney U test was used to determine if significant differences existed in the means of course success from each sample.

Levene's test was performed prior to the t-test to assess the equality of variances between the data sets for HSGPA and ACTM. Levene's test showed that the variances of the data sets for HSGPA ($F = 7.313$, $p = .007$) differed significantly while the variances of the data sets for ACTM ($F = 3.021$, $p = .082$) differed but not significantly as demonstrated in Table 10. Even though the difference of variance for ACTM was not significant, the p-value was near $\alpha = 0.05$ and thus unequal variances were assumed for both variables in the t-test.

Table 10

Levene's Test for Equality of Variances between the Initial Sample and Validation Sample on the Means of HSGPA and ACTM

Variable	F	Equal Variance Assumed df	Equal Variance Not Assumed df
HSGPA	7.313**	13611	4903
ACTM	3.021	13611	5030

**p < .01. ***p < .001

The results of the t-test comparing the means of HSGPA and ACTM for the initial and validation samples showed that the differences between the means of the data sets for both HSGPA ($t = 4.866$, $p < .001$) and ACTM ($t = 9.272$, $p < .001$) were significantly different. Thus, the mean HSGPA of 3.061 in the initial sample was significantly greater than the mean HSGPA of 3.004 in the validation sample. In addition, the mean ACTM score of 20.08 in the initial sample was significantly greater than the mean ACTM score of 19.30 in the validation sample. These results are shown in Table 11.

Table 11

T-test Results Comparing the Difference of Means of HSGPA and ACTM on the Initial Sample and Validation Sample

					95% CI	
Variable	Mean Difference	SE Difference	t	df	Lower	Upper
HSGPA	0.057	0.012	4.866***	4903	0.034	0.080
ACTM	0.738	0.084	9.272***	5030	0.618	0.949

***p < .001

The results of the Mann-Whitney U test comparing the means of course success data from the initial and validation samples showed that the means were significantly different (Mann-Whitney U = 14419319, $p < .001$) as shown in Table 12. Therefore, the mean success rate of 60.9% in the initial sample is significantly greater than the mean success rate of 49.6% in the validation sample.

Table 12

Mann-Whitney U Test Results Comparing the Means of Course Success on the Initial Sample and Validation Sample

Variable	Mean Success Rate		
	Initial Sample n = 10,525	Validation Sample n = 3088	Mann-Whitney U
Course Success Success = 1, Not Success = 0	0.609 (61%)	0.496 (50%)	14419319***

***p < .001

Model Building: The Hierarchical Model

The hierarchical model was constructed of cut points or a combination of cut points of HSGPA and ACTM for the highest four course levels (Calculus, Trigonometry, Gateway, and DM1) using CHAID decision tree modeling. Placement into DM2 occurred if student data did not qualify for placement in any of the other course levels. For the CHAID decision tree modeling process, all course levels used a splitting significance level of $p < 0.1$, a grouping significance level of $p < .05$, and used 10 groups. The Calculus and Trigonometry course levels had relatively small sample sizes (N = 243 and N = 247, respectively). Therefore, for Calculus and Trigonometry, Bonferroni adjustment to the p-value was not included in an effort to increase the power of the model. Minimum values used to create parent and child nodes were set at 10 for the parent node and 5 for the child node to increase the splitting power of the model. Gateway and DM1 course levels had very large sample sizes (N = 4003 and N = 4268, respectively) so the Bonferroni adjustment was included in these models and minimum values for creating parent and child nodes were set at 100 and 50, respectively. Each node of the decision tree was compared to the mean success rate for the course level. The placement measurement was included in the placement model if the success rate for that node was greater than or equal to the mean course success rate. For example, in the Gateway decision tree, a node for HSGPA between 3.32 and 3.58 had a predicted success rate of 71.2%, which is greater than the mean success rate of 63% for the Gateway courses level. Thus, HSGPA between 3.32 and 3.58 was included in the model to place students into Gateway mathematics. The hierarchical placement model created by decision trees in the CHAID process were adjusted for practicality of implementation. In this way, some nodes that would have been included in the model due to having a higher predicted success rate than the mean success rate for the course were removed from the model. For example, the Gateway tree included the following parent/child node combinations and corresponding success rates: 3.04 < HSGPA < 3.32 (59.2%) → ACTM <= 26 (57.9%) → ACTM >= 19 (57.0%) and ACTM <= 18 (67.7%) (Refer to Figure C3 in Appendix C for the full Gateway decision tree). From a purely statistical standpoint, the model would include a HSGPA between 3.04 and 3.32 with an ACTM less than or equal to 18 because the success rate was 67.7% in this node but it would exclude from the model a HSGPA between 3.04 and 3.32 with an ACTM between 19 and 26. This is not feasible from a practical implementation perspective so a HSGPA between 3.04 and 3.32 and an ACTM less than or equal to 18 was not included in the placement model. The full decision trees are in Figures C1 through C4 in Appendix C. The resulting Hierarchical model is as follows:

Calculus: ACTM >= 30 or (HSGPA > 3.23 and ACTM >= 21)

Trig: HSGPA > 3.83 or (HSGPA > 3.54 and ACTM >=23) or ACTM >= 29

Gateway: HSGPA > 3.32 or (HSGPA > 2.86 and ACTM >= 23)

DM1: HSGPA > 3.01 or (HSGPA > 2.73 and ACTM >=17)

DM2: Any student data below that which would place the student into DM1

Model Building: The Algorithmic Model

The algorithmic placement model was constructed of four separate binary logistic regression equations and corresponding cut points from the highest four levels of mathematics courses: Calculus, Trigonometry, Gateway, and DM1. Placement in DM2 occurred if student data did not satisfy placement into any of the above course levels. Course grades in each course level were recoded into success (A, B, C, or Successful = 1) and not success (D, F, Withdrawal, or Unsuccessful = 0) to perform binary logistic regression.

While binary logistic regression does not require data to be normally distributed, it does require that each covariate (HSGPA and ACTM) is linearly related to the log odds (g-value). Therefore, the Box-Tidwell test was used to determine linearity by determining the significance of the interaction term of each covariate and natural log of the covariate in the model. If the interaction term was significant using the Wald statistic ($p < .05$), then the assumption that the covariate and the log odds (g-value) are linearly related was violated. If the assumption is violated, powers of the covariate were added to the model to determine if the relationship between covariate and log odds was polynomial in nature. Beginning with covariate squared, the term was added to the initial model of only HSGPA and ACTM covariates. If the squared covariate term was significant using the Wald statistic ($p < .05$), it remained in the model. If it was not significant, the covariate cubed term was tested in the same way. An interaction term of HSGPA by ACTM was also tested in the model in the same way by adding the interaction term to the initial model (HSGPA and ACTM only) and determining the significance of the interaction term using the Wald statistic ($p < .05$). If multiple interaction terms were significant in this way, all terms were included in the model and tested for significance. This is called the full model. If any interaction terms were not significant ($p > .05$) in this full model, then the insignificant interaction terms were removed resulting in the final model. If, however, multiple interactions terms were removed from the full model, the interaction term whose model, consisting of HSGPA, ACTM, and interaction term produced the least -2LogLikelihood value was added back to the model to create the final model. Each binary logistic regression equation for the highest four course levels (Calculus, Trigonometry, Gateway, and DM1) was created in this manner and, together, form the algorithmic placement model:

Calculus: $g = -10.944 + 1.963(\text{HSGPA}) + 0.176(\text{ACTM})$

Trigonometry: $g = -9.582 + 1.91(\text{HSGPA}) + 0.150(\text{ACTM})$

Gateway: $g = -3.685 + 1.965(\text{HSGPA}) – 0.292(\text{ACTM}) + 0.009(\text{ACTM})^2$

DM1: $g = -5.067 + 1.496(\text{HSGPA}) + 0.056(\text{ACTM})$

DM2: Any student data below that which would place the student into DM1

The g-value in these equations is the log odds. Table 13 presents a summary of the final model logistic regression analysis for each course level.

Table 13

Summary of Logistic Regression Analysis for Variables Predicting Course Success for Each Course Level

						95% CI for e^B	
Course Level	Variable	B	SE	Wald	e^B	Lower	Upper
Calculus	HSGPA	1.963	0.414	22.446***	7.117	3.160	16.029
	ACTM	0.176	0.044	16.387***	1.193	1.095	1.299
	Constant	-10.944					
Trigonometry	HSGPA	1.914	0.386	14.552***	6.780	3.180	14.455
	ACTM	0.150	0.047	10.193**	1.162	1.060	1.275
	Constant	-9.582					
Gateway	HSGPA	1.965	0.087	511.121***	7.135	6.017	8.460
	ACTM	-.292	0.138	4.453*	0.747	0.570	0.979
	$(ACTM)^2$	0.009	0.003	7.750**	1.009	1.003	1.015
	Constant	-3.685					
DM1	HSGPA	1.495	0.075	401.443***	4.459	3.853	5.162
	ACTM	0.056	0.013	17.947***	1.058	1.031	1.086
	Constant	-5.067					

*p < .05. **p < .01. ***p < .001
Probability Cut Points: Calculus 0.656, Trigonometry 0.6934, Gateway 0.6324, DM1 0.6078
e^B is exponentiated B and is the Odds Ratio

Calculating Probability Cut Points for Each Binary Logistic Equation

Once the binary logistic equations were created and log odds (g-value) determined for each course level, probability cut points for each binary logistic equation were determined. SPSS uses an initial probability cut point of 0.50 (50%). A probability cut point of 0.50 corresponds to a g-value of zero. Thus, student data that leads to a g-value greater than zero has a predicted probability of success greater than 50%. While a probability cut point of 0.50

may maximize the overall accuracy of the model in predicting success, it may lead to false positive rate (1 – sensitivity) or false negative rate (1 – specificity) that is higher than desired. A Receiver Operator Curve (ROC), which plots (1 – specificity) with sensitivity, can be created using the predicted probability of each student data and the corresponding course grade to determine a probability cut point other than 0.50 that maximizes the combination of sensitivity (true positive rate) and specificity (true negative rate). The Pythagorean method of determining probability cut point was utilized. In this method, the probability cut point is the probability for which the total of $(1 - \text{sensitivity})^2 + (1 - \text{specificity})^2$ is minimized. Hence, by minimizing the combination of false positive and false negative, sensitivity and specificity are maximized. Once a probability cut point has been determined for each log odds equation, the value of g that corresponds to that probability can be calculated using the formula: $g = \ln\left(\frac{prob\ cut\ point}{1-prob\ cut\ point}\right)$. Note that for a probability cut point of 0.50 the value of g becomes $g = \ln\left(\frac{0.50}{1-0.50}\right) = \ln\left(\frac{0.50}{0.50}\right) = \ln 1 = 0$. Thus, for each course level, a g-value greater than the g-value corresponding to the probability cut point will predict a successful student. The probability cut points determined by the Pythagorean method and the corresponding g-values for each course level were determined to be:

Calculus: Probability cut point = 0.6560 with g-value = 0.6455

Trigonometry: Probability cut point = 0.6934 with g-value = 0.8161

Gateway: Probability cut point = 0.6324 with g-value = 0.5425

DM1: Probability cut point = 0.6078 with g-value = 0.4381

The data analyses of the probability cut points are too large (more than 1500 lines in SPSS for the Gateway course level) to be shown in this research.

Analyzing the Fit of the Hierarchical Model

Research question one asks about the fit of the hierarchical model as measured by predicted accuracy on the initial and validation samples and discrimination on the initial sample. Predicted accuracy of a course level is determined by the sum of the students correctly predicted to succeed and correctly predicted to not succeed according to the rules of the placement model divided by the total number of students in the sample. The predicted accuracy of the hierarchical model was greater than the maximum of course success rate or unsuccessful rate in the initial sample for each course level thus showing an improvement in the ability of the model to predict success above chance in the initial sample. A direct comparison of the predicted accuracy rates to chance shows that, in Calculus, the predicted accuracy was 74.9%, which was 7.8 percentage points greater than the chance success rate of 67.1%. In Trigonometry, the predicted accuracy was 70.9%, which was 6.5 percentage points greater than the chance success rate of 64.4%. In Gateway, the predicted accuracy was 69.6%, which was 6.6 percentage points greater than the chance success rate of 63.0%. In DM1, the predicted accuracy was 66.7%, which was 7.6 percentage points greater than the chance success rate of 59.1%.

Tables 14 shows the predicted accuracy for each course level in cross-tabular form using the hierarchical placement model. In cross-tabular form, the rows show how many of the total students actually succeeded or did not succeed in the class. In Calculus, for example, of the 243 total students, 80 did not succeed while 163 did succeed (success defined as a C or higher). Thus, the chance success rate for Calculus in the initial sample is 163/243 = 67.1%. Cross-tabular form also separates each grouping of actually not successful and successful students into the number of students predicted by the hierarchical model to be not successful and successful. In Calculus, for example, of the 80 students who were actually not successful, 34 were predicted by the hierarchical placement model to be not successful while 46 were predicted to be successful. Specificity, the percent of correctly predicted negative outcomes, is calculated as 34/80 = 42.5% for Calculus. The false positive rate, or Type I Error, is calculated as 1 – specificity and, for Calculus, is 1 – 0.425 = 0.575 (57.5%). The grouping of actually successful students is also separated into the number of students predicted by the hierarchical model to be not successful and successful. In Calculus, for example, of the 163 students who were actually successful, 15 were predicted to be not successful while 148 were predicted to be successful. Sensitivity, the percent of correctly predicted positive outcomes, is calculated as 148/163 = 90.8% for Calculus. The false negative rate, or Type II Error, is found from 1 – sensitivity and, for Calculus, is 1 – 0.908 = 0.092 (9.2%). Accuracy (74.9% for Calculus) is defined as the sum of the correctly predicted negative outcomes (34 for Calculus) and the correctly predicted positive outcomes (148 for Calculus) divided by the total outcomes (243 for Calculus) and is given in the lower right-hand corner of the cross-tabular results found in Table 14.

Table 14

Hierarchical Model Crosstab Results for Predicted Placement for the Initial Sample

Course Level			Cross Tab Results				
			Predicted				
			Not Success	Success	Total		
Calculus		Not Success	34	46	80	42.5%	Specificity
	Actual	Success	15	148	163	90.8%	Sensitivity
		Total	49	194	243	74.9%	Accuracy
			Predicted				
			Not Success	Success	Total		
Trigonometry		Not Success	73	15	88	83.0%	Specificity
	Actual	Success	57	102	159	64.2%	Sensitivity
		Total	130	117	247	70.9%	Accuracy
			Predicted				
			Not Success	Success	Total		
Gateway		Not Success	794	686	1480	53.6%	Specificity
	Actual	Success	530	1993	2523	79.0%	Sensitivity
		Total	1324	2679	4003	69.6%	Accuracy
			Predicted				
			Not Success	Success	Total		
DM1		Not Success	923	821	1744	52.9%	Specificity
	Actual	Success	600	1924	2524	76.2%	Sensitivity
		Total	1523	2745	4268	66.7%	Accuracy

As was the case for the initial sample, the predicted accuracy of the hierarchical model was greater than the maximum of course success rate or unsuccessful rate in the validation sample for each course level thus showing an improvement in the ability of the model to predict success above chance in the validation sample. A direct comparison of the predicted accuracy rates to chance shows that, in Calculus, the predicted accuracy was 58.8%, which was 3.5 percentage points greater than the chance failure rate of 55.3%. In Trigonometry, the predicted accuracy was 68.1%, which was 7.1 percentage points greater than the chance failure rate of 61.1%. In Gateway, the predicted accuracy was 67.7%, which was 15.8 percentage points greater than the chance success rate of 51.8%. In DM1, the predicted accuracy was 63.8%, which was 13.6 percentage points greater than the chance success rate of 50.2%. Table 15 shows the cross-tabular results for accuracy as well as specificity and sensitivity of the hierarchical model for the validation sample.

Table 15

Hierarchical Model Crosstab Results for Predicted Placement for the Validation Sample

Course Level			Cross Tab Results				
			Predicted				
			Not Success	Success	Total		
Calculus		Not Success	17	30	47	36.2%	Specificity
	Actual	Success	5	33	38	86.8%	Sensitivity
		Total	22	63	85	58.8%	Accuracy
			Predicted				
			Not Success	Success	Total		
Trigonometry		Not Success	59	10	69	85.5%	Specificity
	Actual	Success	26	18	44	40.9%	Sensitivity
		Total	85	28	113	68.1%	Accuracy
			Predicted				
			Not Success	Success	Total		
Gateway		Not Success	284	239	523	54.3%	Specificity
	Actual	Success	112	451	563	80.1%	Sensitivity
		Total	396	690	1086	67.7%	Accuracy
			Predicted				
			Not Success	Success	Total		
DM1		Not Success	319	267	586	54.4%	Specificity
	Actual	Success	159	432	591	73.1%	Sensitivity
		Total	478	699	1177	63.8%	Accuracy

Discrimination is the area under the ROC curve and interpreted as the ability of a model to differentiate between a of a pair of opposite outcomes, or in terms of this study, opposite performing students. For example, given two students, one who was actually successful in Calculus and one who was actually not successful in Calculus, correctly discriminating between the two students means that the model correctly predicts the successful student to be successful and the not successful student to be not successful. A discrimination of 70% means that, for any given two students in the sample, the model will correctly select 70% of opposite performing pairs of students in the sample. The discrimination of the hierarchical model was shown to be significant at the $p < .001$ level for all course levels of the initial sample and were all above 70%, which is considered acceptable discrimination (Hosmer, 2013). The discrimination for each course level was: Calculus discrimination = 73.5% ($p < .001$), Trigonometry discrimination = 77.6% ($p < .001$), Gateway discrimination = 75.1% ($p < .001$), and DM1 discrimination = 70.5% ($p < .001$). Models with a p-value less than 0.05 differentiate the successful and unsuccessful students at a rate significantly greater than chance (50%). Table 16 shows the discrimination of both the hierarchical and algorithmic models at each course level with 95% confidence intervals, the difference of the discrimination between the placement models, and the significance of the difference of the discrimination between the placement models using the Critical Ratio Z-test. The difference of the discrimination between the placement models and its significance is discussed in the section "Comparative Analysis of Discrimination of the Placement Models" while comparing the discrimination of the models.

Table 16
Summary Results and Comparison of Model Discrimination on Initial Sample

Course Level	Placement Model	Discrimination	95% CI Lower	95% CI Upper	Difference in Discrimination	p-value^ (two-tailed)
Calculus	Algorithmic	0.740***	0.673	0.807	0.005	0.912
	Hierarchical	0.735***	0.667	0.802		
Trigonometry	Algorithmic	0.759***	0.699	0.820	0.017	0.684
	Hierarchical	0.776***	0.717	0.834		
Gateway	Algorithmic	0.752***	0.737	0.768	0.001	0.925
	Hierarchical	0.751***	0.735	0.766		
DM1	Algorithmic	0.708***	0.692	0.723	0.003	0.786
	Hierarchical	0.705***	0.689	0.720		

***p < .001
^Using the Critical Ratio Z-test to test the significance of the Difference in Discrimination

Analyzing the Fit of the Algorithmic Model

Research question two asks about the fit of the Algorithmic model as measured on the initial sample by the Chi Squared statistic of the Likelihood Ratio Test, Pseudo R^2 values, predicted accuracy, and discrimination. Predicted accuracy was also measured on the validation sample and used to determine model fit on a new set of students. Overall, the algorithmic model for each course level was shown to be a good fit. The Model Chi Squared value of the Likelihood Ratio Test was shown significant at $p < .001$ for each course level. The Model Chi Squared value for each course level was: Calculus Model Chi Squared = 44.255 ($p < .001$), Trigonometry Model Chi Squared = 46.108 ($p < .001$), Gateway Model Chi Squared = 790.520 ($p < .001$), and DM1 Model Chi Squared = 574.622 ($p < .001$). The

Pseudo R^2 values showed a moderate level of improvement over the constant model with the Nagelkerke R^2 values ranging between 0.170 and 0.245 for the DM1 model and Gateway model, respectively, and the Cox & Snell R^2 values ranging between 0.126 and 0.179 for the DM1 model and Gateway models, respectively. Table 17 shows the data for model fit of the algorithmic model.

Table 17

Model Summary for Logistic Regression Equations of the Algorithmic Model for Each Course Level

Course Level	-2 Log Likelihood	Chi Squared	df	Cox & Snell R Square	Nagelkerke R Square
Calculus	263.686	44.255***	2	0.166	0.232
Trigonometry	275.607	46.108***	2	0.170	0.234
Gateway	4483.896	790.520***	3	0.179	0.245
DM1	5198.729	574.622***	2	0.126	0.170

***$p < .001$

The third assessment of model fit for the algorithmic model was the predicted accuracy of simulated placement. The predicted accuracy of the algorithmic model was greater than the maximum of course success rate or unsuccessful rate in the initial sample for each course level thus showing an improvement in the ability of the model to predict success above chance. A direct comparison of the predicted accuracy rates to chance shows that, in Calculus, the predicted accuracy was 72.8%, which was 5.8 percentage points greater than the chance success rate of 67.1%. In Trigonometry, the predicted accuracy was 72.1%, which was 7.7 percentage points greater than the chance success rate of 64.4%. In Gateway, the predicted accuracy was 68.7%, which was 5.7 percentage points greater than the chance success rate of 63.0%. In DM1, the predicted accuracy was 65.3%, which was 6.2 percentage points greater than the chance success rate of 59.1%. Table 18 shows the cross-tabular results for accuracy as well as specificity and sensitivity of the algorithmic model for the initial sample.

Table 18

Algorithmic Model Crosstab Results for Simulated Placement for the Initial Sample

Course Level			Cross Tab Results				
			Predicted				
			Not Success	Success	Total		
Calculus		Not Success	53	27	80	66.3%	Specificity
	Actual	Success	39	124	163	76.1%	Sensitivity
		Total	92	151	243	72.8%	Accuracy
			Predicted				
			Not Success	Success	Total		
Trigonometry		Not Success	67	21	88	76.1%	Specificity
	Actual	Success	48	111	159	69.8%	Sensitivity
		Total	115	132	247	72.1%	Accuracy
			Predicted				
			Not Success	Success	Total		
Gateway		Not Success	1016	464	1480	68.6%	Specificity
	Actual	Success	787	1736	2523	68.8%	Sensitivity
		Total	1803	2200	4003	68.7%	Accuracy
			Predicted				
			Not Success	Success	Total		
DM1		Not Success	1173	571	1744	67.3%	Specificity
	Actual	Success	912	1612	2524	63.9%	Sensitivity
		Total	2085	2183	4268	65.3%	Accuracy

As was the case for the initial sample, the predicted accuracy of the algorithmic model was greater than the maximum of course success rate or unsuccessful rate in the validation sample for each course level thus showing the reliability of the model in a new sample to predict success above chance. A direct comparison of the predicted accuracy rates to chance shows that, in Calculus, the predicted accuracy was 60.0%, which was 4.7 percentage points greater than the chance failure rate of 55.3%. In Trigonometry, the predicted accuracy was 66.4%, which was 5.3 percentage points greater than the chance failure rate of 61.1%. In Gateway, the predicted accuracy was 67.8%, which was 16.0 percentage points greater than the chance success rate of 51.8%. In DM1, the predicted accuracy was 66.0%, which was 15.8 percentage points greater than the chance success rate of 50.2%. Table 19 shows the cross-tabular results for accuracy as well as specificity and sensitivity of the algorithmic model for the validation sample.

Table 19

Algorithmic Model Crosstab Results for Predicted Placement for the Validation Sample

Course Level			Cross Tab Results				
			Predicted				
			Not Success	Success	Total		
Calculus		Not Success	28	19	47	59.6%	Specificity
	Actual	Success	15	23	38	60.5%	Sensitivity
		Total	43	42	85	58.8%	Accuracy
			Predicted				
			Not Success	Success	Total		
Trigonometry		Not Success	56	13	69	81.2%	Specificity
	Actual	Success	25	19	44	43.2%	Sensitivity
		Total	81	32	113	66.4%	Accuracy
			Predicted				
			Not Success	Success	Total		
Gateway		Not Success	297	223	523	56.8%	Specificity
	Actual	Success	124	439	563	78.0%	Sensitivity
		Total	421	665	1086	67.8%	Accuracy
			Predicted				
			Not Success	Success	Total		
DM1		Not Success	420	166	586	71.7%	Specificity
	Actual	Success	234	357	591	60.4%	Sensitivity
		Total	654	523	1177	66.0%	Accuracy

The fourth assessment of model fit for the algorithmic model was discrimination, the ability of a model to differentiate between the success and failure of a pair of opposite performing students. The discrimination of the algorithmic model was shown to be significant at the $p < .001$ level for all course levels in the initial sample and were all above 70%, which is considered acceptable discrimination (Hosmer, 2013). The discrimination for each course level was: Calculus discrimination = 74.0% ($p < .001$), Trigonometry discrimination = 75.9% ($p < .001$), Gateway discrimination = 75.2% ($p < .001$), and DM1 discrimination = 70.8% ($p < .001$). Table 16 shows the discrimination of both the hierarchical and algorithmic models at each course level with 95% confidence intervals, the difference of the discrimination between the placement models, and the significance of the difference of the discrimination between the placement models using the Critical Ratio Z-test.

Comparative Analysis of Predicted Accuracy and Discrimination of the Placement Models

Research question three asks about the differences in predictive ability between the Algorithmic and Hierarchical models as measured by the differences in predicted model accuracy on the initial sample and the validation sample and the differences in discrimination on the initial sample. Comparing predicted accuracy and discrimination on the initial sample compares the fit of the models. Comparisons of predicted accuracy on the validation sample tests the ability of the placement models to retain predictive power on a new set of students.

Comparative Analysis of Predicted Accuracy of the Placement Models

Accuracy percentages for the algorithmic and hierarchical models for each course level of the initial sample were produced using crosstabs between course success (coded success = 1 and not success = 0) and the predicted success of each placement model (coded success = 1 and not success = 0). The algorithmic model predicted success if the log odds (g-value) of the logistic equation was greater than the corresponding probability cut point. The hierarchical model predicted success if the values of HSGPA or ACTM or a combination were greater than (or equal to) the cut point values created by the CHAID decision tree model. The results showed that predicted accuracy was similar for the placement models on each course level of the initial sample with a maximum difference in predicted accuracy of 2.1%. A direct comparison of the accuracy rates for the two models shows that the hierarchical model had greater predicted accuracy than the algorithmic model in Calculus (74.9% to 72.8%), Gateway (69.6% to 68.7%), and DM1 (66.7% to 65.3%) and a lesser predicted accuracy in Trigonometry (70.9% to 72.1%). No statistical analysis was completed to test the significance of these differences in accuracy. Table 20 shows the accuracy rates and the difference of the accuracy rates of the hierarchical and algorithmic placement models for each course level of the initial sample.

Table 20

Comparison of Accuracy Rates for the Hierarchical and Algorithmic Placement Models for the Initial Sample

Course Level	Placement Model	Predicted Accuracy	Difference in Accuracy
Calculus	Hierarchical	74.9%	2.1%
	Algorithmic	72.8%	
Trigonometry	Hierarchical	70.9%	-1.2%
	Algorithmic	72.1%	
Gateway	Hierarchical	69.6%	0.9%
	Algorithmic	68.7%	
DM1	Hierarchical	66.7%	1.4%
	Algorithmic	65.3%	

Similarly, for the validation sample, the predicted accuracy was similar for the placement models on each course level with a maximum difference in predicted accuracy of 2.2%. However, the comparisons of predicted accuracy on the validation sample yielded opposite results as the comparisons of predicted accuracy on the initial sample. A direct comparison of the accuracy rates for the two models showed that the hierarchical model had greater predicted accuracy than the algorithmic model in Trigonometry (68.1% to 66.4%) and a lesser predicted accuracy in Calculus (58.8% to 60.0%), Gateway (67.7% to 67.8%), and DM1 (63.8% to 66.0%). Once again, no statistical analysis was completed to test the significance of these differences in accuracy. Table 21 shows the accuracy rates and the difference of the accuracy rates of the hierarchical and algorithmic placement models for each course level of the validation sample.

Table 21
Comparison of Accuracy Rates for the Hierarchical and Algorithmic Placement Models for the Validation Sample

Course Level	Placement Model	Predicted Accuracy	Difference in Accuracy
Calculus	Hierarchical	58.8%	-1.2%
	Algorithmic	60.0%	
Trigonometry	Hierarchical	68.1%	1.7%
	Algorithmic	66.4%	
Gateway	Hierarchical	67.7%	-0.1%
	Algorithmic	67.8%	
DM1	Hierarchical	63.8%	-2.2%
	Algorithmic	66.0%	

Comparative Analysis of Discrimination of the Placement Models

The discrimination of the placement models can be interpreted as the ability of the model to differentiate between success and not success for a pair of opposite performing students. Models with a p-value less than 0.05 differentiate between the successful and unsuccessful students at a rate significantly greater than chance (50%). The results showed that discrimination rates were similar for the placement models on each course level of the initial sample with a maximum difference in the discrimination rate of 1.7%. The hierarchical model had greater discrimination than the algorithmic model in Trigonometry (77.6% to 75.9%) and lesser discrimination in Calculus (73.5% to 74.0%), Gateway (75.1% to 75.2%), and DM1 (70.5% to 70.8%). The Critical Ratio Z-test was used to determine whether the difference in discrimination of the models was significant. It was found that the hierarchical and algorithmic models had no significant difference in discrimination at any course level with the lowest p-value = .684 in the Trigonometric models. Thus, the placement models differentiate between successful and not successful students and statistically similar rates for each course level. Table 16 shows the discrimination of both the hierarchical and algorithmic models at each course level with 95% confidence intervals, the difference of the discrimination between the placement models, and the significance of the difference of the discrimination between the placement models using the Critical Ratio Z-test.

Analysis of Agreement in Placement on the Validation Sample

The fourth research question asks about the differences in placement results of the algorithmic and hierarchical models. The complete validation sample with student data from all course levels was used to create simulated placement of students using the Hierarchical model and the Algorithmic model. For each model, placement into the Gateway course level was determined. Any student not placing into Gateway was assessed for placement into DM1. Any student not placed into DM1 was placed into DM2. Placement into the Calculus and Trigonometry course levels was not used due to the practical requirement of additional prerequisite course success in addition to the placement criteria. In practice, a student who places into Calculus according to the placement model but has not successfully completed Trigonometry, the prerequisite course for Calculus, would be placed into Trigonometry. Similarly, in practice, a student who places into Trigonometry according to the placement model but has not successfully completed the appropriate Gateway course, the prerequisite for Trigonometry, would be placed into the Gateway course. Since individual success in prerequisite courses was unknown, any student placing into Calculus or Trigonometry was placed into Gateway for this analysis. The hierarchical model placed students as follows: DM2 = 1111 (36.0%), DM1 = 779 (25.2%), and Gateway = 1198 (38.8%) while the algorithmic model placed the students as follows: DM2 = 1676 (54.3%), DM1 = 252 (8.2%), and Gateway 1160 (37.6%).   Table 22 shows a summary of the simulated placement for the hierarchical and algorithmic models.

Table 22

Summary of Simulated Placement on the Validation Sample

	Course Level	Frequency	Percent	Cumulative Percent
Algorithmic Model	DM2	1676	54.3%	54.3%
	DM1	252	8.2%	62.4%
	Gateway	1160	37.6%	100.0%
Hierarchical Model	DM2	1111	36.0%	36.0%
	DM1	779	25.2%	61.2%
	Gateway	1198	38.8%	100.0%

The agreement of placement for the models is determined using weighted kappa, a statistic that measures the degree to which two mutually exclusive categories agree. A weighted kappa of 0 indicates that the models do not agree on placement on a level greater than what could be expected by chance whereas a weighted kappa close to 1 indicates substantial agreement between the models. Weighted kappa is used instead of kappa because, for example, the difference of a DM2 placement by one model and a Gateway placement by another model is greater than the difference of a DM1 placement by one model and a Gateway placement by another model. Weighted kappa takes into account the degree of differences in placement. Table 23 shows the weights of disagreement used to calculate weighted kappa.

Table 23

Weights for Calculating Weighted Kappa in the Comparison of Model Placement

		Course Level		
		DM2	DM1	Gateway
Course Level	DM2	0	1	2
	DM1	1	0	1
	Gateway	2	1	0

An analysis of the simulated placement shows that the hierarchical and algorithmic models agreed on the placement of 2472 out of 3088 students (80.1%) in the validation sample. The resulting weighted kappa equaled 0.777 indicating a substantial agreement of placement between the two placement models (Landis and Koch, 1977). Most of the differences in placement (507 students or 82% of the differing-placed students) occurred with the hierarchical model placing students in DM1 that the algorithmic model placed in DM2. Table 24 shows the cross-tabular results of the simulated placement with the rows showing the placement by the algorithmic model and the columns the placement by the hierarchical model. Cross-tabular form subdivides the number of students placing into each course level by one placement model into the three placement levels using the other placement model. For example, the hierarchical model placed 1198 students into Gateway. Of those 1198 students, the algorithmic model placed 60 in DM2, 13 in DM1, and 1125 in Gateway. Likewise, of the 1160 students that the algorithmic model placed into Gateway, the hierarchical model placed 1 in DM2, 34 in DM1, and 1125 in Gateway. Thus, the algorithmic and hierarchical models agreed on the placement of 1125 students into the Gateway course level. The 80.1% overall agreement of placement is the sum of the agreed upon placement from DM2, DM1, and Gateway divided by the total students ((1109 + 238 + 1125)/3088) = 80.1%).

Table 24

Crosstabs of Simulated Placement Results on the Validation Sample

		Hierarchical Model			
		DM2	DM1	Gateway	Total
Algorithmic Model	DM2	1109	507	60	1676
	DM1	1	238	13	252
	Gateway	1	34	1125	1160
	Total	1111	779	1198	3088
			N	Percent	
Agreement of Placement			2472	80.1%	
Weighted Kappa			0.777		

Course levels DM1 and DM2 can be combined as developmental mathematics to further analyze the placement into either Gateway or developmental mathematics. Developmental mathematics is any mathematics course that is not credit bearing. An analysis of the simulated placement into Gateway or developmental mathematics (DM1 combined with DM2) shows a very high level of agreement between the hierarchical and algorithmic models. The models agreed upon the placement of 2980 out of 3088 students

(96.5%) with an unweighted kappa of 0.926. Table 25 shows the cross-tabular results of the simulated placement into either Gateway or developmental mathematics with the rows showing the placement by the algorithmic model and the columns the placement by the hierarchical model.

Table 25

Crosstabs of Simulated Placement Results into only Gateway or Developmental Mathematics on the Validation Sample

		Hierarchical Model		
		Developmental Mathematics	Gateway	Total
Algorithmic Model	Developmental Mathematics	1855	73	1928
	Gateway	35	1125	1160
	Total	1890	1198	3088
		N	Percent	
Agreement of Placement		2980	96.5%	
Weighted Kappa		0.926		

Analysis of the Predictive Strengths of the Covariates HSPGA and ACTM in Course Success

The fifth research question asks about the predictive strengths of the covariates HSGPA and ACTM in the binary logistic regression equations of the algorithmic model for each course level. The predictive strength of the covariates in the logistic regression equations can be determined by the value of the Wald statistic and its corresponding p-value, directly comparing partially standardized beta values, and comparing the values of the Chi

Squared statistic of the model when covariates are added to the model in succession. To begin with, the Wald statistic showed that both HSGPA and ACTM were statistically significant predictors of success in the binary logistic equation for each course level with $p < .01$ for each covariate at each course level. Comparatively, a greater Wald statistic and lower p-value represents the covariate that has greater influence in the model's ability to predict course success. The covariate HSGPA had a greater Wald statistic than ACTM for each course level (DM1 through Calculus) indicating that HSGPA has greater predictive strength in the binary logistic equations. The Wald statistic for HSGPA and ACTM was 22.446 and 16.387, respectively, for Calculus, 14.552 and 10.193, respectively, for Trigonometry, 515.993 and 55.576, respectively, for Gateway, and 401.443 and 17.947, respectively, for DM1. Table 26 gives the beta values and Wald statistic of the logistic equations for course levels: Calculus, Trigonometry, Gateway, and DM1. The results in Table 26 are used to compare the predictive strength of the covariates on the models without other interaction variables and, therefore, are not necessarily the final models derived from Table 13 and given in the section "Model Building: The Algorithmic Model".

Table 26

Summary of Logistic Regression Analysis for HSGPA and ACTM Variables Only

Course Level	Variable	B	SE	Wald
	HSGPA	1.963	0.414	22.446***
Calculus	ACTM	0.176	0.044	16.387***
	Constant	-10.944		
	HSGPA	1.914	0.386	14.552***
Trigonometry	ACTM	0.150	0.047	10.193**
	Constant	-9.582		
	HSGPA	1.971	0.087	515.993***
Gateway	ACTM	0.093	0.012	55.576***
	Constant	-3.685		
	HSGPA	1.495	0.075	401.443***
DM1	ACTM	0.056	0.013	17.947***
	Constant	-5.067		

** $p < .01$, *** $p < .001$

Partially standardized beta values can be analyzed to compare the predictive strength of the covariates. The beta values must be standardized (made to be without scale) since the covariates HSGPA and ACTM have different scales (0-4 for HSGPA and 0-36 for ACTM) and can be done by converting the covariates to z-scores and then using the z-scores in the binary logistic regression equations. In logistic regression, the beta values are only partially standardized because only the covariates are standardized since the output is a logit. The greater the partially standardized beta values, the greater the influence the predictor has on the logistic regression equation. Thus, the covariate with the greater partially standardized beta value is a stronger predictor of success in the algorithmic placement model. The covariate HSGPA had greater partially standardized beta values than ACTM for each course level. The partial standardized beta values for HSGPA and ACTM are 0.767 and 0.646, respectively, for Calculus, 0.814 and 0.486, respectively, for Trigonometry, 0.945 and 0.281, respectively, for Gateway, and 0.764 and 0.154, respectively, for DM1. Table 27 shows the partially standardized beta values for each course level.

Table 27

Partially Standardized Beta Values in Initial Logistic Regression Equations Using Only HSGPA and ACTM for Each Course Level

Course Level	Variable	Partially Standardized B
Calculus	HSGPA	0.767
	ACTM	0.646
Trigonometry	HSGPA	0.814
	ACTM	0.486
Gateway	HSGPA	0.945
	ACTM	0.281
DM1	HSGPA	0.764
	ACTM	0.154

Finally, the predictive strength of the covariates can be measured by comparing the Chi Squared statistic when each variable is added in succession to the binary logistic regression model. In binary logistic regression, Chi Squared represents the difference of goodness of fit of the model with and without the variable. By adding variables in succession, the goodness of fit of the model is calculated with one variable and then again with both variables. The larger the Chi Squared value, the better the model fits the data. When the first variable is added, the goodness of fit and corresponding Chi Squared value are calculated for that variable only indicating how much that variable accounts for the fit of the model. When the second variable is added, the additional Chi Squared value indicates how much better the model fits the data with the new variable added. Entering the covariates successively in the initial binary logistic regression equation for Calculus with HSGPA first, HSGPA had a Chi Squared of 25.696 and ACTM added 18.559 for a total of 44.255 while entered with ACTM first, ACTM had a Chi Squared of 18.103 and HSGPA added 26.152. For Trigonometry with HSGPA first, HSGPA had a Chi Squared of 35.300 and ACTM added 10.808 for a total of 46.108 while entered with ACTM first, ACTM had a Chi Squared of 16.486 and HSGPA added 29.621. Entering the covariates successively in the initial binary logistic regression equation for Gateway with HSGPA first, HSGPA had a Chi Squared of 725.511 and ACTM added 56.910 for a total of 782.421 while entered with ACTM first, ACTM had a Chi Squared of 144.218 and HSGPA added 638.202. For DM1 with HSGPA first, HSGPA had a Chi Squared of 556.304 and ACTM added 18.318 for a total of 574.622 while entered with ACTM first, ACTM had a Chi Squared of 105.507 and HSGPA added 469.115. In each course level, HSGPA increased the fit of the model more than ACTM whether it was added first or second. The statistically significant increase of Chi Squared value when adding the second variable (whether it was HSGPA or ACTM) in each course level indicates that each variable contributes to the fit of the model in a way that the other variable does not, although more so for HSGPA than ACTM. Table 28 is a summary of the change in Chi Squared value for successively adding the covariates HSGPA and ACTM to the initial logistic regression equations.

Table 28

Summary of Change in Chi Squared for Logistic Regression Models

Course Level	Order	Variable	Chi Squared	df
Calculus	Step 1	HSGPA	25.696***	1
	Step 2	ACTM	18.559***	1
	Model		44.255***	2
	Step 1	ACTM	18.103***	1
	Step 2	HSGPA	26.152***	1
	Model		44.255***	2
Trigonometry	Step 1	HSGPA	35.300***	1
	Step 2	ACTM	10.808**	1
	Model		46.108***	2
	Step 1	ACTM	16.486***	1
	Step 2	HSGPA	29.621***	1
	Model		46.108***	2
Gateway	Step 1	HSGPA	725.511***	1
	Step 2	ACTM	56.910***	1
	Model		782.421***	2
	Step 1	ACTM	144.218***	1
	Step 2	HSGPA	638.203***	1
	Model		782.421***	2
DM1	Step 1	HSGPA	556.304***	1
	Step 2	ACTM	18.318***	1
	Model		574.622***	2
	Step 1	ACTM	105.507***	1
	Step 2	HSGPA	469.115***	1
	Model		574.622***	2

$^{**}p < .01$. $^{***}p < .001$

In summary, both HSGPA and ACTM were significant predictors of course success by both the Wald Statistic ($p < .001$) and the Chi Squared statistic of overall model improvement ($p < .001$) for each course level (DM1 through Calculus). However, HSGPA had a greater influence than ACTM on each model's ability to predict course success in every course level for all three measures: Wald Statistic, partially standardized beta value, and Chi Squared statistic. The predictive ability of HSGPA was substantially greater than ACTM in the Gateway and DM1 course levels and moderately greater than ACTM in the Calculus and Trigonometry course levels.

Additional Analysis

Comparative Analysis of Actual Placement and Simulated Placement using the Hierarchical and Algorithmic Models

The actual placement of students in the validation sample was also determined and compared to the simulated placement by the hierarchical and algorithmic models. Table 29 shows the actual placement of students along with the simulated placement by the hierarchical and algorithmic models that are in Table 22. Placement into the Calculus and Trigonometry course levels was not used due to the practical requirement of additional prerequisite course success in addition to the placement criteria. A slightly greater number of students actually placed into Gateway, 42.6%, as were simulated to be placed into Gateway by both the algorithmic and hierarchical placement models, 37.6% and 38.8%, respectively. Correspondingly, a slightly fewer number of students actually placed into developmental mathematics (DM1 or DM2), 58.4%, as were simulated to be placed into developmental mathematics by both the algorithmic and hierarchical models, 62.4% and 61.2%, respectively. The greatest difference in placement existed within developmental mathematics. Students were actually placed in DM1 at a much greater rate, 38.1%, than by either simulated placement model, 8.2% for the algorithmic model and 25.2% for the hierarchical model. Correspondingly, the algorithmic model placed the greatest number of students in the lowest level of mathematics (DM2), 54.3%, as compared with 36.0% for hierarchical model and 20.3% for actual placement.

Table 29

Summary of Actual Placement on the Validation Sample and Simulated Placement by the Hierarchical and Algorithmic Models on the Validation Sample

	Course Level	Frequency	Percent	Cumulative Percent
Actual Placement	DM2	627	20.3%	20.3%
	DM1	1177	38.1%	58.4%
	Gateway	1284	42.6%	100.0%
Algorithmic Model	DM2	1676	54.3%	54.3%
	DM1	252	8.2%	62.4%
	Gateway	1160	37.6%	100%
Hierarchical Model	DM2	1111	36.0%	36.0%
	DM1	779	25.2%	61.2%
	Gateway	1198	38.8%	100.0%

Cross-tabulations were calculated for placement into DM2, DM1, and Gateway for both actual placement and the algorithmic model and for actual placement and the hierarchical model. The cross-tabular results show, of the number of students who were actually placed in a particular course level, how many were place by the algorithmic model or hierarchical model in the three course levels. For example, of the 1284 students who actually placed into Gateway, the algorithmic model simulated that 371 would be placed in DM2, 99 would be placed in DM1, and 814 would be placed in Gateway. The algorithmic model agreed with the actual placement of 814 (63.6%) students into Gateway. Similarly, of the 1284 students who actually placed into Gateway, the hierarchical model simulated that 182 would be placed in DM2, 263 would be placed in DM1, and 839 would be placed in Gateway. The hierarchical model agreed with the actual placement of 839 (65.3%) students into Gateway. Overall, the algorithmic model and actual placement agreed on the placement of 1476 out of 3088 students (47.8%) for a resulting weighted kappa of 0.339. The hierarchical model and actual placement agreed on the placement of 1700 out of 3088 students (55.1%) for a resulting weighted kappa = 0.408. Thus, the hierarchical model aligned more closely to actual placement than the algorithmic model. The algorithmic and hierarchical placement models and actual placement generally agreed on the number of students placed into Gateway, 1160, 1198, and 1284, respectively, but not necessarily which students. Of the 1284 students actually placed in Gateway, the algorithmic model would have placed 470 (36.6%) in developmental mathematics with 371 (28.9%) being placed in DM2. Similarly, the hierarchical model would have placed 445 (34.7%) students in developmental mathematics with 182 (14.2%) students in DM2. Table 30 shows the cross-tabular results of actual placement and the simulated placement by the algorithmic model for the validation sample. Table 31 shows the cross-tabular results of actual placement and the simulated placement by the hierarchical model for the validation sample.

Table 30

Summary of Crosstabs for Actual Placement and Simulated Placement by the Algorithmic Model on the Validation Sample

		Algorithmic Model			
		DM2	DM1	Gateway	Total
Actual Placement	DM2	535	26	66	627
	DM1	770	127	280	1177
	Gateway	371	99	814	1284
	Total	1676	252	1160	3088

	N	Percent
Agreement of Placement	1476	47.8%
Weighted Kappa	0.339	

Table 31

Summary of Crosstabs for Actual Placement and Simulated Placement by the Hierarchical Model on the Validation Sample

		Hierarchical Model			
		DM2	DM1	Gateway	Total
Actual Placement	DM2	451	106	70	627
	DM1	478	410	289	1177
	Gateway	182	263	839	1284
	Total	1111	779	1198	3088

	N	Percent
Agreement of Placement	1700	55.1%
Weighted Kappa	0.408	

CHAPTER FIVE

CONCLUSION

Summary of the Study

The purpose of the study was to compare a hierarchical placement model and an algorithmic placement model for placement into college mathematics courses using high school grade point average (HSGPA) and ACT-Math scores (ACTM) as measures of placement. Prior research has shown that multiple measure placement models are more accurate than a single measure placement (ACT, 2008; DesJardins & Lindsay, 2008; Ngo & Kwon, 2015; Westrick & Allen, 2014) and the measures of HSGPA and ACTM are significant predictors of course success (Ngo and Kwon, 2015). Two current multiple measure placement models are the hierarchical model and the algorithmic placement model (Barnett, Bostian, Peterson, & Welbeck, 2016; Reeves Bracco et al., 2014) but minimal research exists comparing the two models which led to the development of this study.

This study compared the hierarchical and algorithmic placement models on the accuracy of placement in the initial college level math course using course success to determine accuracy, the ability to differentiate between a pair of successful and unsuccessful students, and the simulated placement of students into the Gateway, Developmental Mathematics 1 (DM1) and Developmental Mathematics 2 (DM2) course levels. The sample of students was separated into an initial sample and a validation sample. Both models were created using data from the initial sample. The validation sample was used to test the accuracy of the models on a different set of student data and to simulate the placement of a different set of students. The accuracy of the placement models was compared on the initial sample to test the internal validity of each model and the validation sample to test the reliability of each model. Discrimination, the ability of each model to differentiate between the success and failure of a pair of opposite performing students, was compared on the initial sample. The final comparison of the placement models was the extent of agreement of the simulated placement of students in the validation sample. A secondary purpose of the study was to measure and compare the influence of HSGPA and ACTM on course success in the algorithmic placement model to add to the current research on these placement measures.

This chapter includes a summary of the results, the implications of these findings, and concludes with a discussion of limitations, suggestions for future research, and a brief conclusion. The research questions for this study were:

1. What is the predictive ability of the hierarchical placement model in terms of accuracy and discrimination?
2. What is the predictive ability of the algorithmic placement model in terms of measures of model fit, accuracy, and discrimination?
3. What are the differences in predictive ability between the hierarchical and algorithmic placement models in terms of accuracy and discrimination?
4. What are the differences in course placement results between the hierarchical and algorithmic placement models?
5. What are the predictive strengths of HSGPA and ACTM scores on course success at each course level for the algorithmic model?

Summary of the Results

Research questions one and two asked about the predictive ability of each placement model. The research showed that both placement models were viable models to predict course success for each mathematics course level: Calculus, Trigonometry, Gateway, and Developmental Mathematics 1 (DM1). In terms of accuracy, the accuracy of placement on the initial sample was greater than chance on the four course levels by an average of 7.1 percentage points for the hierarchical model (Table 14) and by an average of 6.3 percentage points for the algorithmic model (Table 18). Furthermore, the accuracy rates on a different set of students in the validation sample were also greater than chance for each placement model in every course level. The accuracy of placement on the validation sample was greater than chance on the four course levels by an average of 10.0 percentage points for the hierarchical model (Table 15) and by an average of 10.4 percentage points for the algorithmic model (Table 19). Thus, the placement models exhibit both internal validity on the initial sample and reliability on the validation sample. The validity of each model for placement was also supported by their ability to differentiate between pairs of successful and not successful students. Discrimination rates above 70% are considered acceptable levels of discrimination (Hosmer, 2013). The discrimination rate for each model was greater than 70% for each course level with an average discrimination rate of 74.2% for the hierarchical model and 74.0% for the algorithmic model (Table 16). The validity of the algorithmic model, which is a collection of logistic regression equations, can also be measured by the fit of these logistic regression equations. For each course level, the fit of the logistic regression equation to the data was statistically significant at the $p < .001$ level for the Chi Squared statistic of model fit with pseudo R^2 values between 0.126 and 0.179 for Cox & Snell R^2

values and between 0.170 and 0.245 for Nagelkerke R^2 values for the different course levels (Table 17). In total, the hierarchical and algorithmic placement models were shown to be statistically better predictors of course success than chance and thus viable models for placement into mathematics course levels.

The crux of this study and the differentiating factor of this study to other studies exists in research questions three and four, which compare the hierarchical and algorithmic placement models. Research question three asks about the differences in predictive ability of the hierarchical and algorithmic placement models. The two models were quite similar in their ability to accurately predict course success across the mathematics course levels with one model slightly outperforming the other model in some metrics and vice versa. The hierarchical model more accurately predicted course success than the algorithmic model on the initial sample in every course level except Trigonometry with the greatest difference in accuracy of 2.1 percentage points in Calculus (74.9% to 72.8% for the hierarchical and algorithmic models, respectively) (Table 20). The algorithmic model more accurately predicted course success on the validation sample in every course level except Trigonometry with the greatest difference in accuracy of 2.2 percentage points in DM1 (63.8% to 66.0% for the hierarchical and algorithmic models, respectively) (Table 21). Thus, the accuracy percentages of the two models were similar and there was not a pattern of one model predicting success among the course levels better than the other model.

The discrimination of the hierarchical and algorithmic models between pairs of opposite performing students were also similar for each course level. The algorithmic model differentiated between opposite performing students slightly better than the hierarchical model in every course level except Trigonometry, but the difference of discrimination was not statistically significant in any course level using the Critical Ratio Z-test (Table 16). The smallest p-value for the Critical Ratio Z-test was .684 in Trigonometry in which the discrimination was .776 for the hierarchical model and .759 for the algorithmic model. Thus, while both models differentiated between a successful student and not successful student better than chance for each course level, neither model significantly differentiated better than the other model in any course level. Thus, the answer to research question three was that the hierarchical and algorithmic placement models were nearly equivalent in their ability to accurately predict course success and to differentiate between pairs of opposite performing students at each course level.

Research question four asks about the differences in simulated course placement for the hierarchical and algorithmic placement models of students in the validation sample. While the hierarchical and algorithmic models performed similarly, there was a substantial difference in placement of students within developmental mathematics. The two models agreed on 80.1% of the placement of students with a weighted kappa value of 0.777 (Table 24) indicating substantial agreement of the models (Landis and Koch, 1977). Both models placed about the same percentage of students in gateway mathematics at just under 40% of students with the hierarchical model placing 38 more students in Gateway than the algorithmic model, 1198 to 1160, respectively (Table 22). Not only was the number of students placed into Gateway similar, the two placement models placed nearly all the same students into Gateway. Of the 1198 students placed into Gateway by the hierarchical model, 1125 (93.9%) of those students were also placed into Gateway by the algorithmic model. Of the 1160 students placed into Gateway by the algorithmic model, 1125 (97.0%) of those students were also placed into Gateway by the hierarchical model (Table 25).

Correspondingly, the two placement models placed nearly all the same students into developmental mathematics.

The main difference between the hierarchical and algorithmic models existed in the placement of students within developmental mathematics. While both models placed about 60% of students into developmental mathematics, the algorithmic model placed substantially more students in the lowest level mathematics (DM2) than the hierarchical model. The algorithmic model placed 1676 students (54.3% of the total students or 86.9% of the students placed into developmental mathematics) in the lowest level of mathematics (DM2) while the hierarchical model placed 1111 students (36.0% of the total students or 58.8% of the students placed into developmental mathematics) in the lowest level of mathematics (Table 22). This equates to 565 more students placed into the lowest level of mathematics by the algorithmic model than the hierarchical model or, stated another way, 527 more students placed into DM1 and 38 more students placed into Gateway by the hierarchical model than the algorithmic model. Thus, the answer to research question four was that the simulated placement of the models was nearly equivalent between Gateway and developmental mathematics but, within developmental mathematics, the algorithmic model placed substantially more students in the lowest level of mathematics than the hierarchical model.

The fifth research question asked about the predictive ability of HSGPA and ACTM in the logistic regression equations of the algorithmic model. In terms of predictive ability, the first result was that both covariates are statistically significant predictors of course success in the logistic regression equations by the Wald statistics for beta values (Table 26) and the change in Chi Squared value for model fit (Table 28). Therefore, both covariates significantly contribute to the predictive ability of the model. Secondly, when added to the model consisting of only the other covariate, each covariate significantly improves the fit of the model by the change in Chi Squared value (Table 28). Therefore, both HSGPA and ACTM significantly add predictive power to the model above what is contained in the other covariate.

A comparison of the predictive ability of the covariates HSGPA and ACTM shows that HSGPA is a stronger predictor of course success than ACTM in each course level. Every statistical measure utilized in the study to compare the covariates, the Wald statistic for beta values in the logistic regression equations (Table 26), a comparison of partially standardized beta values (Table 27), and the Chi Squared statistic for adding successive covariates to the model (Table 28), showed that HSGPA was a better predictor of course success than ACTM at each mathematics course level. Additionally, HSGPA was a substantially better predictor of course success in developmental and gateway mathematics courses.

Implications of the Predictive Ability and Simulated Placement of Each Placement Model

In summary, the study showed that the hierarchical and algorithmic placement models are both viable models for placing students using HSGPA and ACTM as placement measures, that the models performed similarly in terms of accuracy and in discriminating between a pair of opposite performing students, and that the simulated placement of students by the two models between Gateway and developmental mathematics was nearly identical. The study also showed that, within developmental mathematics, the algorithmic model placed a far greater number of students in the lower developmental mathematics level (DM2) than the hierarchical model.

The implications of these findings are three-fold. First, a college looking to implement a multiple measure placement model should expect that either the hierarchical or the algorithmic model can be utilized successfully and that the models will separate students into Gateway and developmental mathematics (DM1 combined with DM2) similarly. Second, this study indicates that the algorithmic model may underplace students as evidenced by placing a substantial number of students into the lowest level of developmental mathematics without gaining significant improvements in placement accuracy. Third, the hierarchical model performed equally well as the algorithmic model in predicting course success despite the hard cut points for the hierarchical model.

The conclusion of this study, that both the hierarchical placement model and the algorithmic placement model are acceptable models for placement of students in an initial mathematics course, aligns with studies showing that multiple measure placement models may be the best predictor for college success (Bahr et al., 2014; Belfield & Crosta, 2012; Lewallen, 1994; Ngo & Kwon, 2015; Scott-Clayton, et al., 2014). Prior knowledge is extended in this study by comparing the hierarchical and algorithmic multiple measure placement models. A comparison of the placement models showed that they performed similarly, especially in terms of accuracy, discrimination, and placement of students into Gateway and developmental mathematics. Perhaps the most convincing result showing the similarity of the two models was that, when determining whether students should be placed in Gateway or developmental mathematics (Table 25), the models agreed on 2980 out of 3088 students (96.5%) for a kappa value of 0.926, a near perfect level of agreement (Landis and Koch, 1977). Since two distinct, viable placement models agree to near perfection on which students should be in developmental mathematics, it is realistic to expect that a multiple measure placement model can reasonably distinguish between those students who are prepared for Gateway and those students who would benefit from developmental mathematics.

Whereas the hierarchical and algorithmic placement models agreed to near perfection which students should be in developmental mathematics and which students should be in Gateway, the models differed substantially in the simulated placement of students within developmental mathematics. The algorithmic model in this study placed 565 more students from the validation sample in DM2 than the hierarchical model while achieving only a minimally better accuracy rate in DM1. This suggests that the algorithmic model may significantly underplace students by placing more students than necessary in DM2 instead of DM1. Students placed into the lowest level of developmental mathematics are at greater risk of not completing a Gateway mathematics course due to compounded completion rates and stop out rates. Baker et al. (2014) found that the probability of completing a Gateway math course starting from two levels below Gateway is 16% if the student does not stop out while the probability of completing a Gateway math course starting one level below Gateway is about 36%. Nationally, the stop out rate, the percentage of students who successfully complete a course but do not enroll in the next course, is about 35% (Bailey et al., 2010). If the results of this study are generalizable, colleges with multiple levels of developmental mathematics should consider whether an algorithmic placement model would underplace a significant number of students into the lowest level of developmental mathematics, which may potentially negatively affect their academic progress.

It would have been reasonable to assume that the algorithmic model would outperform the hierarchical model due the hard cut points of the hierarchical model. For example, consider the hierarchical placement model for Gateway: HSGPA > 3.32 or (HSGPA > 2.86 and ACTM >= 23). A student with a HSGPA of 2.85 would not qualify for Gateway in the hierarchical model. This is a hard cut point. However, in the algorithmic model, a student with a HSGPA of 2.85 and an ACTM of 27 would meet the requirement for placement into Gateway. Recall that the logistic regression equation for gateway is $g = -3.685 + 1.965(\text{HSGPA}) - 0.292(\text{ACTM}) + 0.009(\text{ACTM})^2$ and requires a g-value greater than 0.5425 for placement into Gateway. Substituting 2.85 for HSGPA and 27 for ACTM gives a g-value of 0.59225 which is greater than 0.5425. The algorithmic model allows a more sliding scale of HSGPA and ACTM for placement into the course level. For this reason, it would be reasonable to assume that the algorithmic model would more accurately place students into each course level. However, as stated above, the accuracy rates of predicting course success for the two models were very similar, as was the ability of each model to differentiate between pairs of opposite performing students. Additionally, the hierarchical model placed a fewer number of students into the lowest level of developmental mathematics in the simulated placement. Therefore, despite the hard cut points, the hierarchical model performed as well as the algorithmic model in terms of accuracy and discrimination and possibly better than the algorithmic model in terms of placement.

The hierarchical model has additional practical advantages over the algorithmic model. The placement rules and cut points of the hierarchical model are more easily defined and explained than the logistic regression equations and corresponding cut points for the g-value of the algorithmic model. It is likely that students will more easily understand why they are placed into a course level using the hierarchical model and for advisors to explain the placement results. It is also likely that decision makers and researchers can more easily analyze and adjust the cut points of the hierarchical model than the logistic regression equations of the algorithmic model. For example, if a college's Gateway placement is: HSGPA > 3.32 or (HSGPA > 2.86 and ACTM >= 23), students with scores just above and/or below these placement cut-values can be identified and their success rates analyzed to determine if adjustment in the placement model is needed. Alternatively, pilot placement models with cut scores just below current cut scores can be implemented and analyzed to determine whether current placement policy can be lowered to include additional students in the higher-level courses. These analyses may be more difficult to perform and communicate about using the logistic regression equations of the algorithmic model. Therefore, if the results of this study are generalizable, colleges considering implementing a multiple measure placement policy should consider using a hierarchical placement model due to its nearly identical ability to predict course success as the algorithmic model, its tendency to place fewer students in the lowest level of mathematics, and the greater ease in which the hierarchical model is understood, analyzed, and adjusted.

Implications of the Significance of Predictor Variables in Determining Course Success

The results from this study suggests that HSGPA and ACTM are significant predictors of course success in mathematics and should be considered as measures of placement in any mathematics placement model. This result aligns with the research by Ngo and Kwon (2015). Of the two, HSGPA was shown to be the greater predictor of course success, a result that has consistently appeared in placement research (ACT, 2014; Allen et al., 2008; Barnett et al., 2016; Camara & Echternacht, 2000; Geiser & Studley, 2002; Maruyama, 2012; Noble & Sawyer, 2004; Robbins et al., 2006; Rothstein, 2004; Sanchez, 2013; Sawyer, 2013; Schmitt, 2012; Westrick & Allen, 2014). Finally, similar to the results obtained by the ACT Corporation (2008), DesJardins and Lindsay (2008), and Ngo and Kwon (2015), even though HSGPA was shown to be the stronger predictor, ACTM added statistically significant predictive strength to the placement model. Adding ACTM to the algorithmic model after HSGPA showed a significant increase in model fit for each course level. A further example of the significance of ACTM is that, in the hierarchical model, the placement rules for each course level used both ACTM and HSGPA as a measure for differentiating placement, not just HSGPA. Thus, while HSGPA is the stronger measure, using both measures increased the predictive strength of each placement model. The study adds to the research on HSGPA and ACTM by showing that these two placement measures are not just significant predictors of success in a Gateway course but significant predictors of success at every course level, DM1, Trigonometry, and Calculus. Additionally, this study has shown that HSGPA is a substantially greater predictor of course success than ACTM in Gateway and DM1. Therefore, if the results of this study are generalizable, colleges considering a multiple measure placement model should consider using both HSGPA and ACTM and should strongly consider using HSGPA as a placement measure for placement in Gateway and developmental mathematics.

Limitations

Lack of Model Precision Due to Possible Variations in the Model Building Process

The hierarchical and algorithmic placement models are both susceptible to a lack of precision due to choices and variations in the model building process. Different decisions within the model building process can result in different values and cut points in the placement models leading to different placement results. In the algorithmic model, multiple methods exist to determine the probability cut point for g. The algorithmic model in this study was built using the Pythagorean method of determining probability cut points to maximize sensitivity and specificity in which $(1 – \text{sensitivity})^2 + (1 – \text{specificity})^2$ is minimized. The summation method, in which the sum of sensitivity and specificity is maximized, is another method for determining the probability cut point g that maximizes sensitivity and specificity but may produce different results. A third method, the 50% method, is the default method for SPSS and sets all probability cut points for g at 50%. Both the summation method and the 50% method are viable alternatives to the Pythagorean method of determining cut points and may yield different cut points for g and thus different placement results.

Additional research was performed to document the differences of accuracy and placement using the summation method and the 50% method for the Gateway course level of the algorithmic model. The accuracy rate for predicting course success using the summation method in the Gateway level course was 68.4% (instead of 68.7%) for the initial sample and 67.7% (instead of 67.8%) for the validation sample. The number of students in the validation sample placed into the Gateway course level by the algorithmic model using the summation method was 1128 (instead of 1160). Therefore, a difference in placement for the summation method, instead of the Pythagorean method, existed but was minimal. The accuracy rate for predicting course success using the 50% method for the Gateway level course was 70.8% (instead of 68.7%) for the initial sample and 65.0% (instead of 67.8%) for the validation sample. The number of students in the validation sample placed into the Gateway course level by the algorithmic model using the 50% method was 1674 (instead of 1160). Therefore, a difference in placement for the 50% method instead of the Pythagorean method existed and was substantial.

The result of this study would have likely been at least slightly different if a different method of probability cut point had been used in the algorithmic model, resulting in a possible lack of precision. From an implementation perspective, reporting only the logistic regression equations and each probability cut point g is not sufficient. The method for calculating g needs to be understood, reported, and maintained if placement is to be reliable and if placement data are to be continually analyzed and adjusted. For example, using a Pythagorean method of determining the cut point for g one year and the 50% method the following year will yield unreliable and incomparable results. Colleges considering using an algorithmic model should be equipped to understand the binary logistic regression equations, the corresponding cut point value for g and the method used to determine g, and be able to communicate that information to decision makers and those communicating with students such as advisors.

The hierarchical placement model was also susceptible to a lack of precision. Choices in the hierarchical model building process that affect the placement rules include choosing the significance levels for splitting or joining nodes, the minimum number of data points needed for child and parent nodes to exist, the number of intervals used to change scaled variables into ordinal variables, whether or not to use Bonferroni adjustment, and whether or not to use validation/bootstrapping. The results of this study would have likely been at least slightly different if changes to any of these factors had been different, resulting in a possible lack of precision in the study. Adjusting the number of intervals for scaled variables tended to have the greatest impact on differences in placement cut points determined by the CHAID method of decision tree modeling in the hierarchical placement model. For example, this study used ten intervals to transform the scaled variable HSGPA into an ordinal variable for each course level. Adjusting the number of intervals in the Gateway course level from ten to twenty, changed the placement rule from HSGPA > 3.32 or (HSGPA > 2.86 and ACTM >= 23) to HSGPA > 3.11 or (HSGPA > 2.86 and ACTM >= 26), a sizeable difference. The accuracy rate for predicting course success using this new placement rule was 70.6% (instead of 69.6%) for the initial sample and 63.3% (instead of 67.7%) for the validation sample. The number of students in the validation sample placed into the Gateway course level using this new placement rule was 1447 (instead of 1198). Therefore, a difference in placement and accuracy existed depending on the choice of the number of intervals used to change a scaled variable (HSGPA) into an ordinal variable. Similarly, other choices in the hierarchical model building process such as the significance levels for splitting and joining nodes, the minimum number of data points needed for parent and child nodes to exist, whether or not to use a Bonferroni adjustment, and whether or not to use validation/bootstrapping may affect the placement model when adjusted.

Colleges considering using a hierarchical model should be equipped to understand the CHAID decision tree model, the entire model building process including choices within the model, and be able to communicate that information to decision makers and those communicating with students such as advisors. There should be a recognition that the hierarchical and algorithmic placement models are delicately constructed and that small changes in the model building process can result in differences of placement results. The placement models should be constructed consistently from year to year using the same model building process to have consistent placement results that can be analyzed and compared.

Possible Sample Bias

The results of the study showed a statistically significant difference in the mean values of HSGPA, ACTM, and course success rates between the initial sample and validation sample (Tables 11 and 12). The mean HSGPA, ACTM, and course success rate were 3.061, 20.08, and 61%, respectively, for the initial sample and 3.004, 19.30, and 50% for the validation sample, all statistically significant differences. This difference in the means of these measures indicate a possible bias in the initial sample from which the hierarchical and algorithmic models were constructed. Metropolitan Community College (MCC) did not require every incoming student to submit HSGPA and ACTM scores during the time period of the initial sample but did require those transcript measures, if available, during a portion of the time period of the validation sample. Thus, the validation sample likely represents a broader spectrum of the student population than the initial sample, which may have led to sample bias in this study. The possible sample bias may not have affected the comparisons of accuracy and discrimination between the models since both models were tested on the same samples. However, the reliability of the models and the simulated placement of the students in the validation sample may have been affected by the possible sample bias. While the placement models accurately placed students in the validation sample when the course level was predetermined (Tables 15 and 19), both models placed many more students at a lower course level than their actual placement when course level was not predetermined (Table 29). The agreement between the hierarchical model placement and actual placement and the agreement between the algorithmic model placement and actual placement was poor, agreeing on only 55.1% and 47.8% of the placement of students, respectively, with weighted kappa values of 0.408 and 0.339, respectively (Tables 30 and 31). This could be the result of the significantly different means for HSGPA, ACTM, and success rates between the initial and validation samples. Further research is suggested to create and analyze the hierarchical and algorithmic placement models using a sample of students in which submission of high school transcript data was required.

Lack of Generalizability of the Research

This study is limited by a possible lack of generalizability to other two-year colleges and four-year colleges because the data were collected from only one college. While the sample sizes in both the initial and validation sample were large and the college from which the data are collected has a diverse population of students in both urban and suburban locations, the placement models were built for that particular set of student data. However, since the comparisons of the models was the focus of the research and not the models themselves, the possible lack of generalizability does not negatively affect the research in this study. At a minimum, this research can provide a basis for replication at other colleges.

Limited Number of Predictor Variables

This study used HSGPA and ACTM as the sole predictor variables for the multiple measure placement models. The predictor variables were chosen after extensive review of the literature on placement and success in college mathematics courses and, in general, follow the suggestions of the literature. However, the limited number of predictors and the fact that they were predetermined is a limitation of the study. Further research is suggested using additional predictive measures such as the highest mathematics course completed in high school, mathematics-specific grade point average in high school, affective measurements such as academic motivation and academic self-efficacy, and intended major or academic pathway.

Future Research

Further research using student data from additional two-year colleges and four-year colleges and universities should be conducted to broaden the scope of the study. Differences among two-year colleges could be compared as well as differences between two-year colleges and four-year colleges. Continued study of the hierarchical and algorithmic placement models using additional placement measures such as prior mathematics courses taken and affective measures should be conducted to test the predictive power of those measures in each model. Research should also be conducted on the effects of the placement models for various socioeconomic, gender, and ethnic groups to determine whether one placement model benefits certain people groups. Finally, as stated earlier, the hierarchical and algorithmic placement models should be created and analyzed using a sample of students in which submission of transcript data was required.

Conclusion

Literature has consistently shown that multiple measure placement models place students more accurately in an initial college mathematics course than any single measure (ACT, 2008; DesJardins & Lindsay, 2008; Ngo & Kwon, 2015; Westrick & Allen, 2014). This study compared two of the multiple measure placement models that have been used by colleges in the United States, the hierarchical placement model and the algorithmic placement model (Barnett et al., 2016; Reeves Bracco et al., 2014) using HSGPA and ACTM as placement measures. The results showed a similarity of the models in accurately predicting course success, discriminating between a pair of successful and not successful students, and placing students into either Gateway or developmental mathematics. The results also showed a difference between the models in placing students within developmental mathematics and in ease of use and interpretation. The models were astonishingly similar in separating students into Gateway or developmental mathematics with one model placing 96.5% of students into the same course level as the other model. Conversely, the models were remarkably different in their placement of students into DM1 and DM2, one and two levels below Gateway mathematics, respectively. The algorithmic model placed a far greater number of students in DM2 than the hierarchical model, 54.3% to 36.0%. Since the hierarchical and algorithmic models had similar accuracy and discrimination rates, there is concern that the Algorithmic model may underplace students by placing a substantial number of students in DM2 that may have been successful in DM1 or higher. Finally, the study illuminated that the hierarchical model may have some advantage over the algorithmic model in ease of use and understanding by students, advisors, and other stakeholders. The combination of the hierarchical and algorithmic models having equivalent rates of accuracy and discrimination, the underplacement of students into DM2 by the algorithmic model, and the ease of use of the hierarchical model leads to the conclusion that the hierarchical model may be a more useful multiple measure placement model for colleges.

APPENDIX A

The North Carolina Community College System's (NCCCS) Math Benchmark Courses Eligible for Multiple Measures Placement (as of 2016)

- Advanced Functions and Modeling
- AP Calculus
- AP Statistics
- Discrete Mathematics
- Essentials for College Math (a Southern Region Education Board (SREB) course – Math Ready)
- Integrated Mathematics IV
- International Baccalaureate Mathematics
- International Baccalaureate Computer Science
- Mindset
- Pre-Calculus

(North Carolina Community College System, n. d.)

APPENDIX B

Histograms and Normal Q-Q Plots for HSGPA and ACTM Data

Figure B1. Histogram of HSGPA Data in the Initial Sample

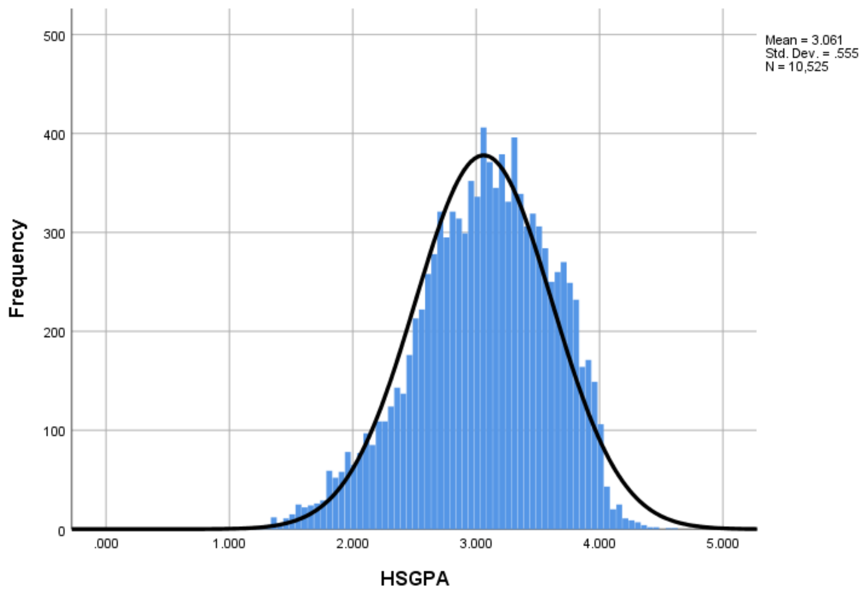


Figure B2. Normal Q-Q Plot of HSGPA Data in the Initial Sample

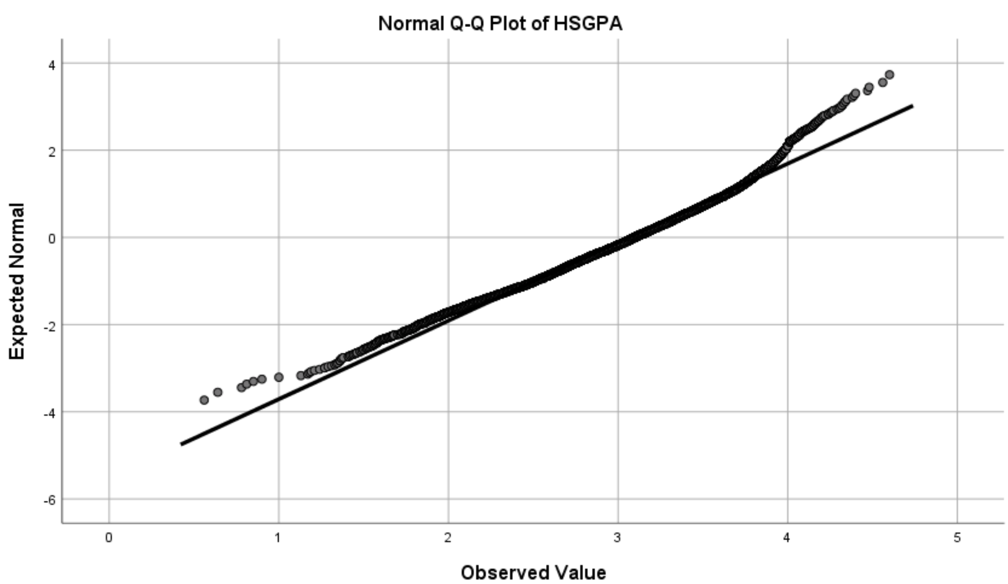


Figure B3. Histogram of ACTM Data in the Initial Sample

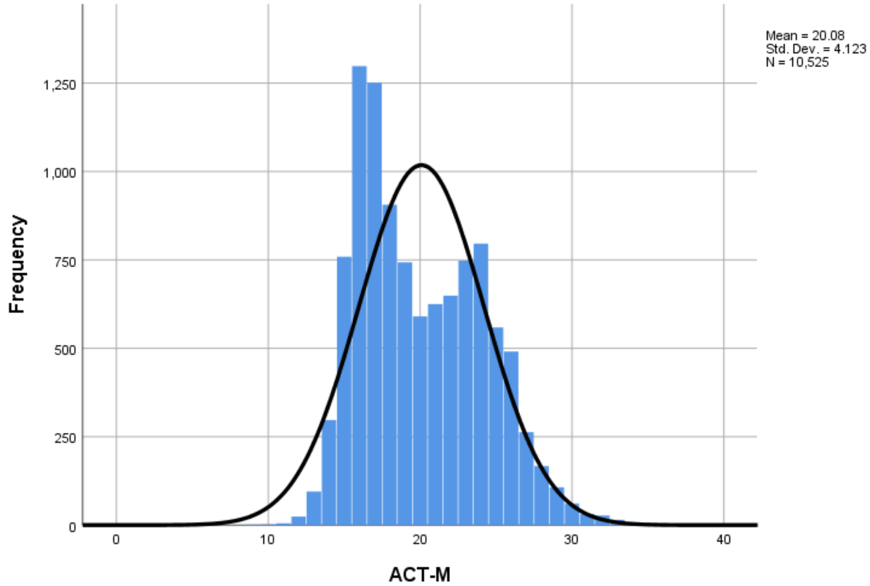


Figure B4. Normal Q-Q Plot of ACTM Data in the Initial Sample

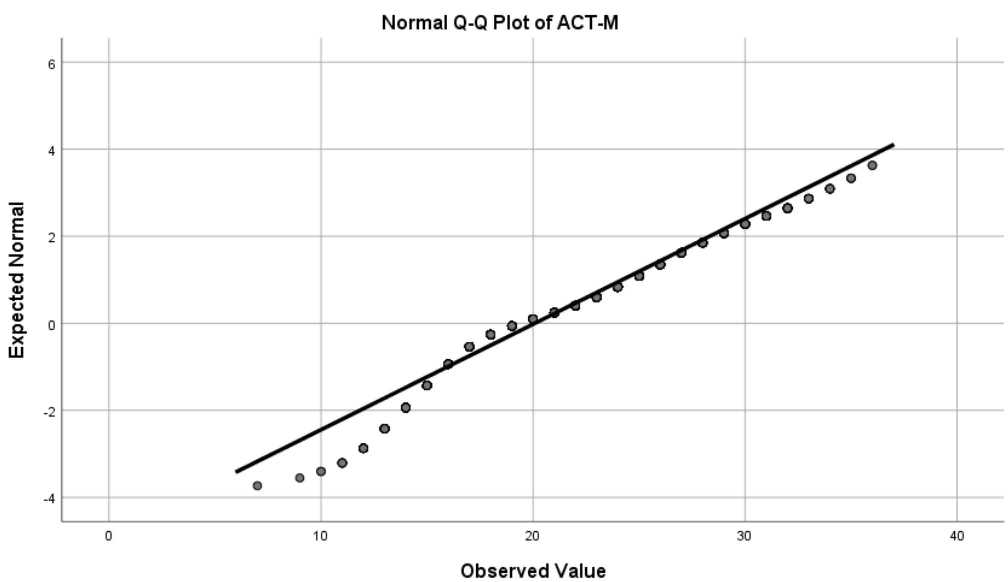


Figure B5. Histogram of HSGPA Data in the Validation Sample

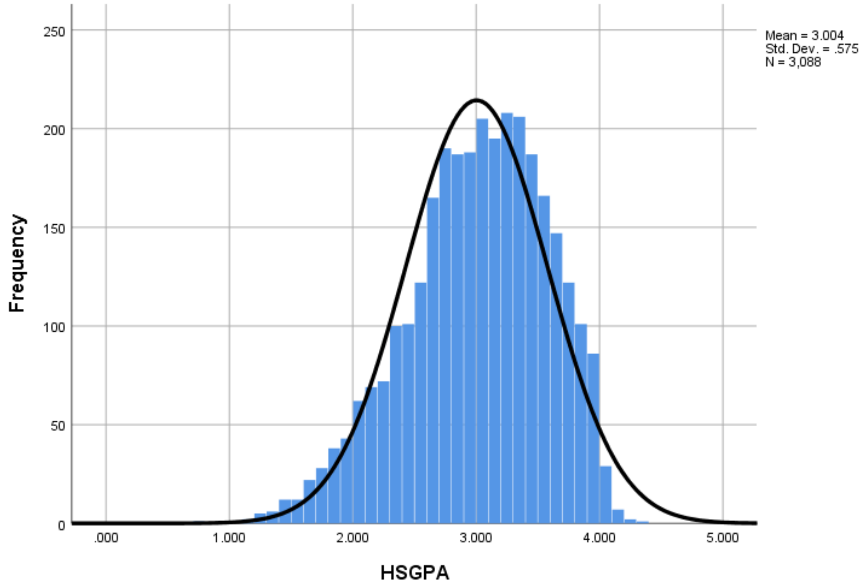


Figure B6. Normal Q-Q Plot of HSGPA Data in the Validation Sample

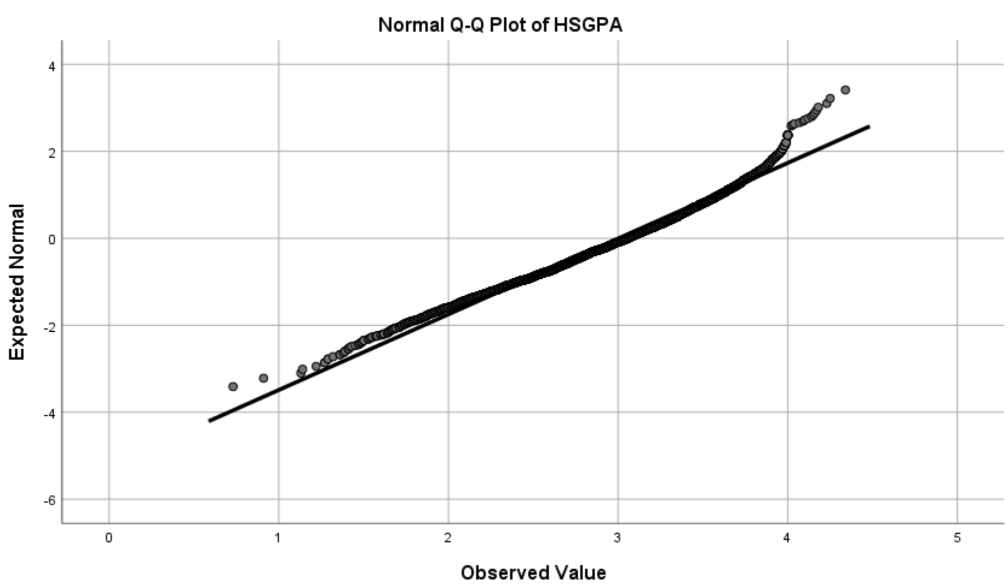


Figure B7. Histogram of ACTM Data in the Validation Sample

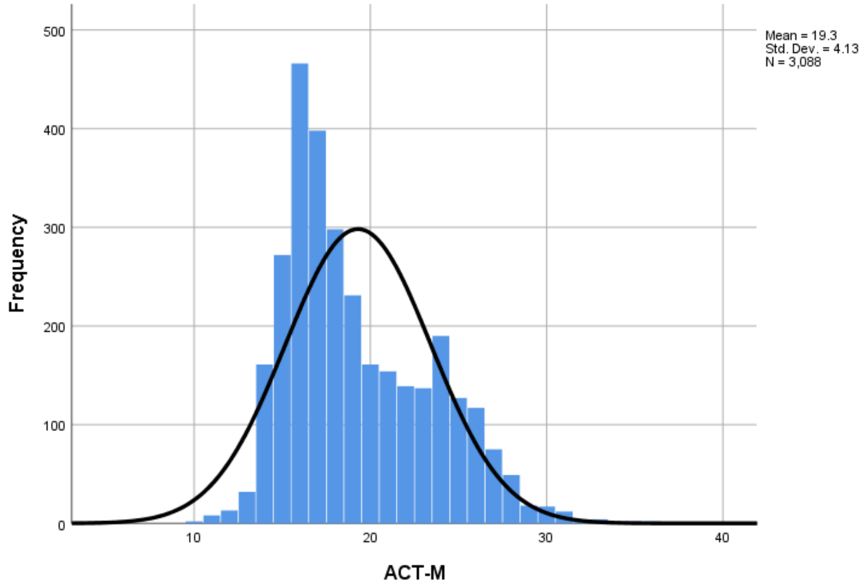
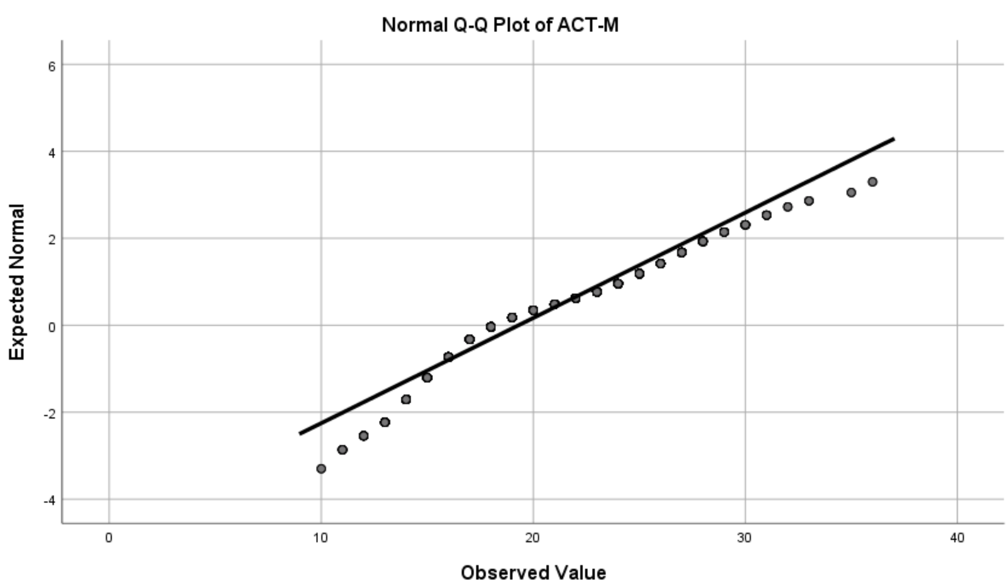


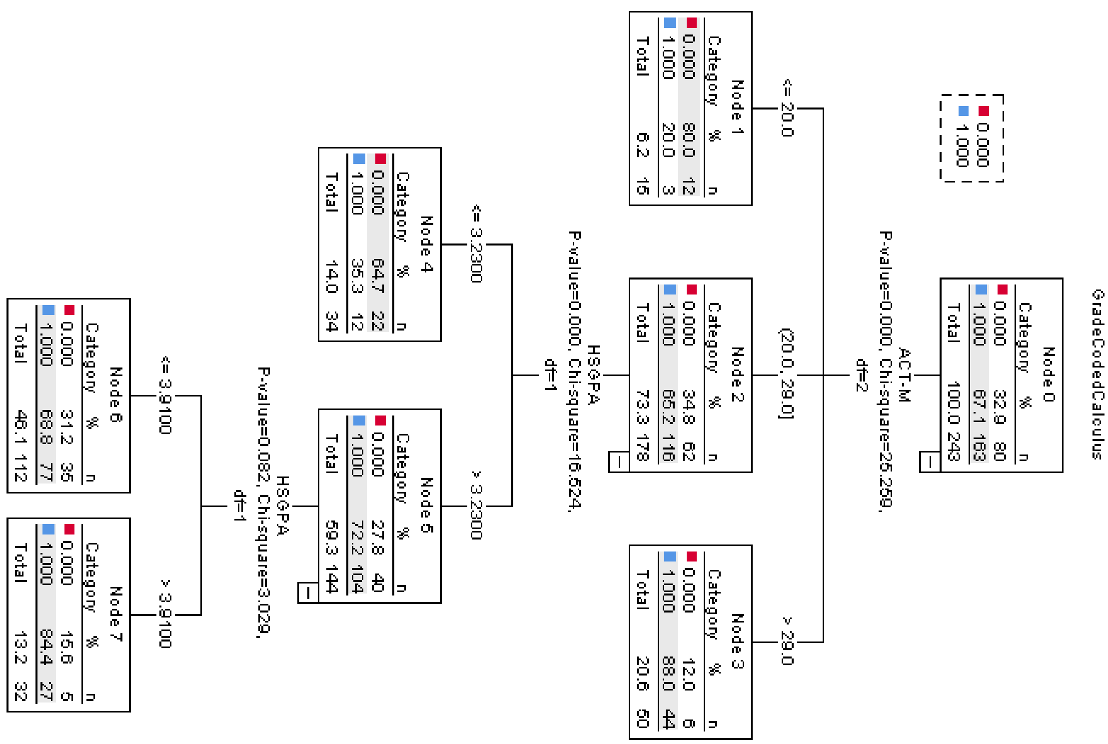
Figure B8. Normal Q-Q Plot of ACTM Data in the Validation Sample



APPENDIX C

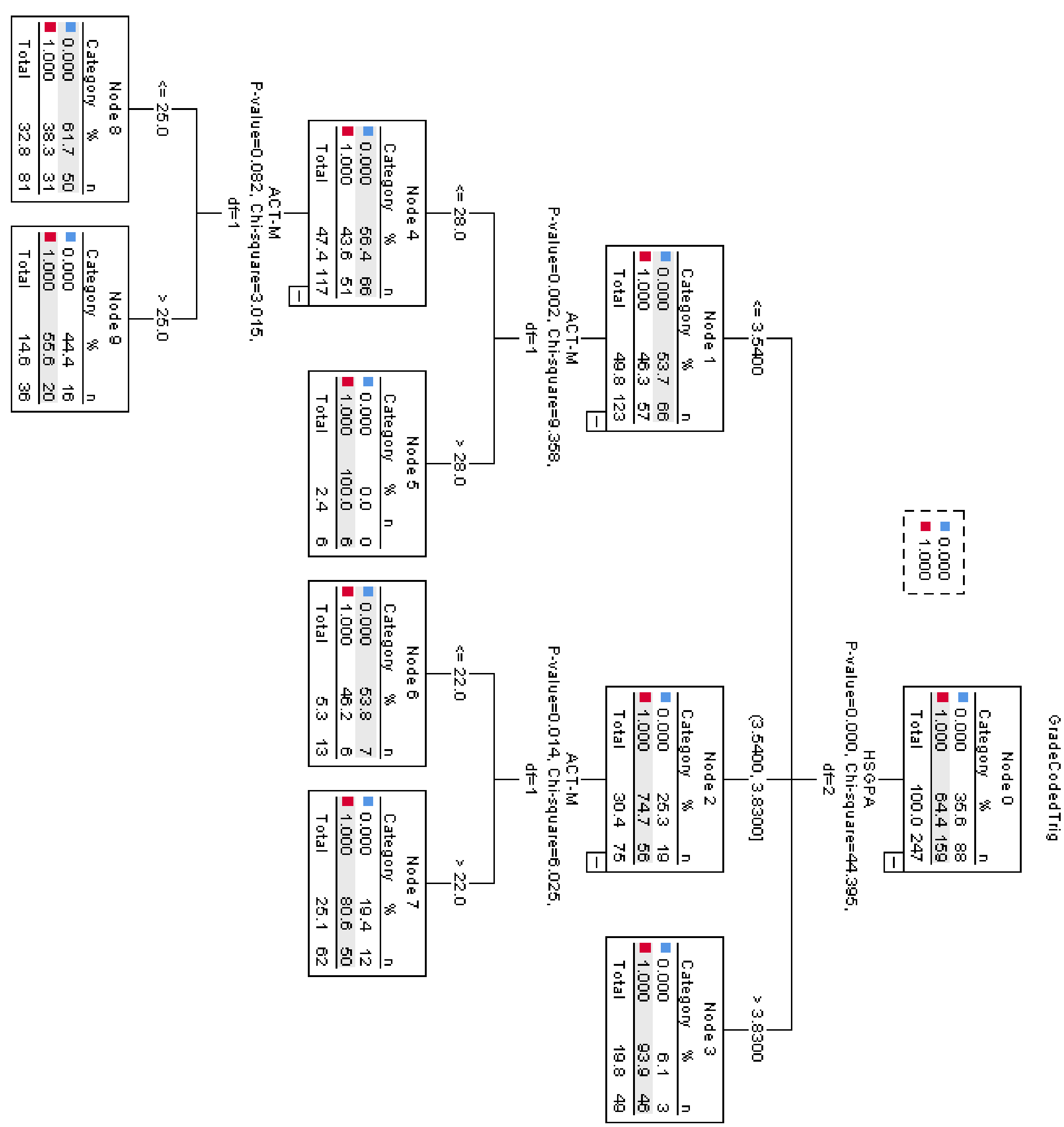
CHAID Decision Tree Model Outputs from SPSS

Figure C1. CHAID Decision Tree for Calculus



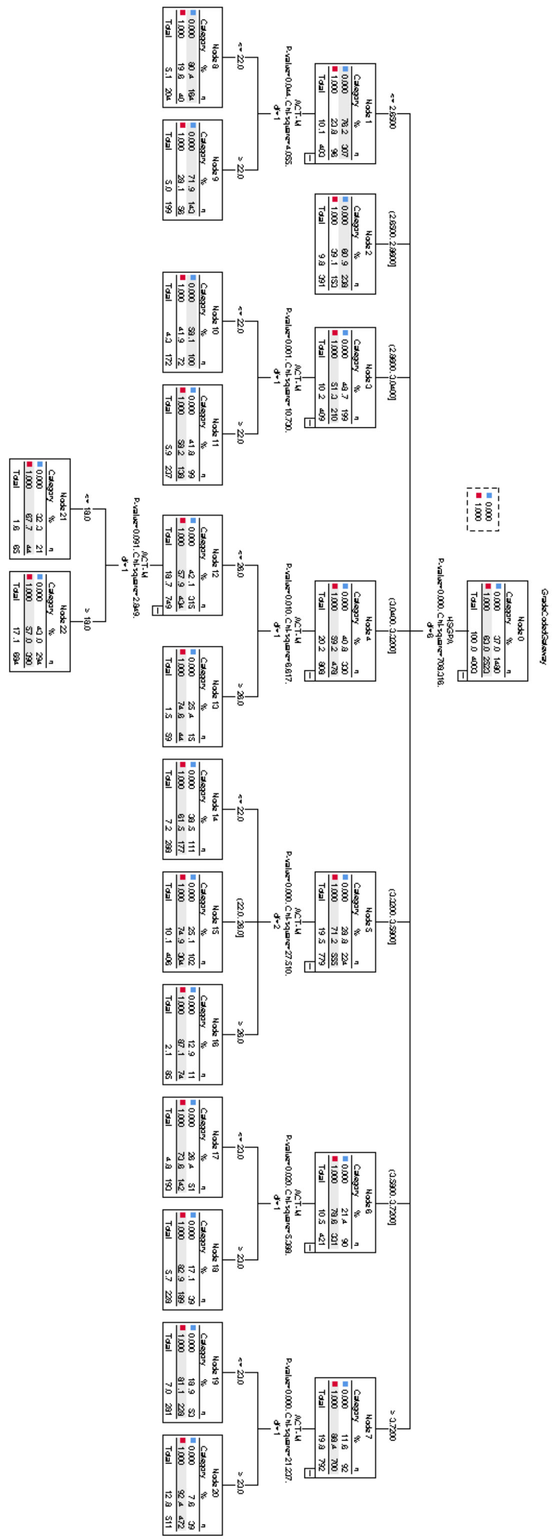
Coded: 0 = Not Success and 1 = Success

Figure C2. CHAID Decision Tree for Trigonometry



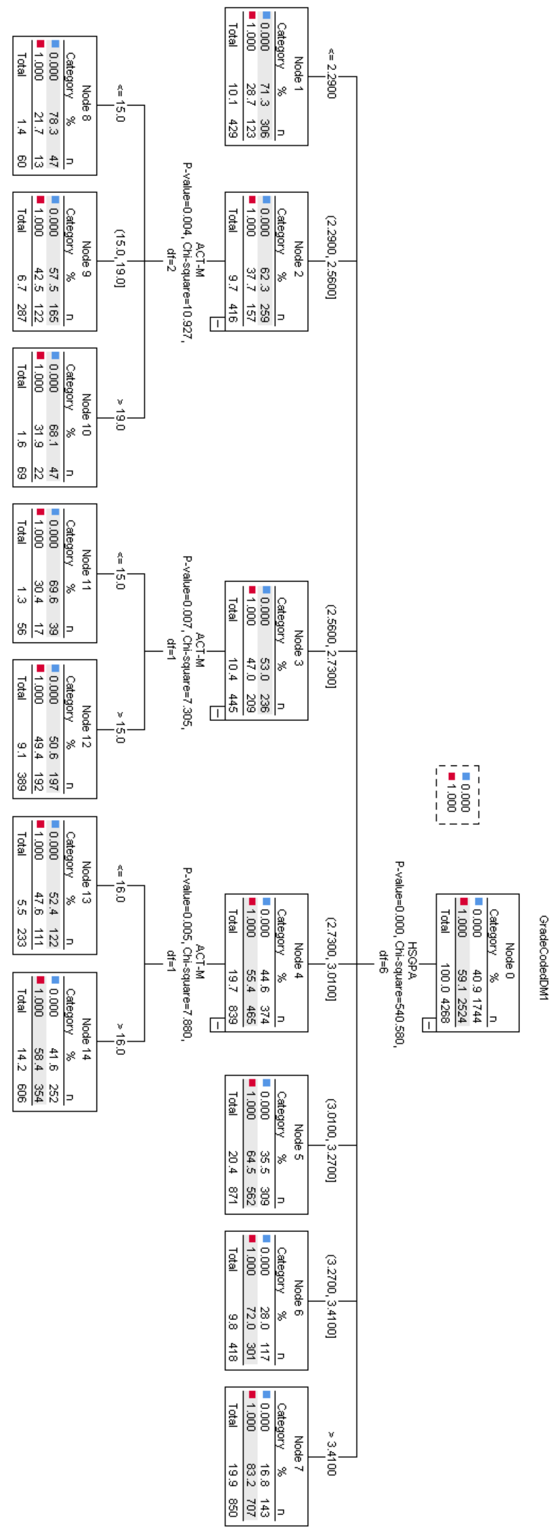
Coded: 0 = Not Success and 1 = Success

Figure C3. CHAID Decision Tree for Gateway



Coded: 0 = Not Success and 1 = Success

Figure C4. CHAID Decision Tree for DM1



Coded: 0 = Not Success and 1 = Success

REFERENCES

ACT (2008). *The relative predictive validity of ACT scores and high school grades in making college admission decisions. Issues in College Success*. Retrieved from https://www.act.org/research/policymakers/pdf/PredictiveValidity.pdf

ACT (n.d.). *ACT Compass placement test*. Retrieved from http://www.act.org/products/higher-education-act-compass/

ACT (2014). *ACT Technical Manual 2014.* ACT.

ACT (2016). *The condition of college and career readiness 2016: Missouri key findings*. ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/state26_Missouri_Web_Secured.pdf

ACT (2017). *National distributions of cumulative percents for ACT test scores: ACT-tested high school graduates from 2015, 2016, and 2017*. ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/state26_Missouri_Web_Secured.pdf

Allen, J., Robbins, S., Casillas, A., & Oh, I. (2008). Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness. *Research in Higher Education*, *49*(7), 647-664.

Armstrong, W. B. (2000). The association among student success in courses, placement test scores, student background data, and instructor grading practices. *Community College Journal of Research and Practice*, *24*(8), 681-695.

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *The Journal of Higher Education*, *77*(5), 886-924.

Bahr, P., Hayward, C., Hetts, J., Lamoree, D., Newell, M., Pellegrin, N, … Willett, T. (2014). *Multiple measures for assessment and placement* [White Paper]. Retrieved from http://rpgroup.org/system/files/MMAP_WhitePaper_Final_September2014.pdf

Bailey, T., Jeong, D. W., & Cho, S. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, *29*, 255-270.

Barnett, E., Bostian, B., Peterson, G., & Welbeck, R. (2016). *Moving beyond the placement test: Multiple measures assessment. Great Lakes assessment webinar*. Retrieved from https://ccrc.tc.columbia.edu/media/k2/attachments/moving-beyond-placement-test-multiple-measures.pdf

Belfield, C. R., & Crosta, P. M. (2012). *Predicting success in college: The importance of placement tests and high school transcripts*. CCRC Working Paper No. 42. Community College Research Center, Columbia University. Retrieved from http://hdl.handle.net/10022/AC:P:13086.

Bettinger, E., Evans, B., & Pope, D. (2011). *Improving college performance and retention the easy way: Unpacking the ACT exam*. National Bureau of Economic Research Working Paper No. 17119. Retrieved from https://www.nber.org/system/files/working_papers/w17119/w17119.pdf

Bettinger, E., & Long, B. T. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work? *Journal of Human Resources*, *44*(3), 736-771.

Bleeker, S., Moll, H., Steyerberg, E., Donders, A., Derksen-Lubsen, G., Grobbee, D., & Moons, K. (2003). External validation is necessary in prediction research: A clinical example. *Journal of Clinical Epidemiology*, *56*, 826-832.

Boatman, A., & Long, B. T. (2010). Does remediation work for all students? How the effects of postsecondary remedial and developmental courses vary by level of academic preparation. *Educational Evaluation and Policy Analysis*. Retrieved from http://www.postsecondaryresearch.org/i/a/document/17998_Aug2011Brief.pdf

Burdman, P. (2012). *Where to begin? The evolving role of placement exams for students starting college*. Retrieved from https://www.achievingthedream.org/sites/default/files/resources/Where_to_Begin.pdf.

Burley, H., Butner, B., & Cejda, B. (2001). Dropout and stopout patterns among developmental education students in Texas community colleges. *Community College Journal of Research and Practice, 25*(10), 767-782.

Calcagno, J. C., & Long, B. T. (2008). *The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance*. No. w14194. National Bureau of Economic Research. Retrieved from https://www.nber.org/system/files/working_papers/w14194/w14194.pdf

Camara, W., & Echternacht, G. (2000). *The SAT1 and high school grades: Utility in predicting success in college*. Research Notes RN-10, The College Board, Office of Research and Development. Retrieved from https://files.eric.ed.gov/fulltext/ED446592.pdf

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review, 42*(4), 41-57.

Central Carolina Community College (2017). *Required placement scores for curriculum courses*. Retrieved from http://www.cccc.edu/studentservices/placementtesting/scores/pdfs/Placement_Scores.pdf

College Board (2016a). *Concordance tables*. Retrieved from https://collegereadiness.collegeboard.org/pdf/higher-ed-brief-sat-concordance.pdf

College Board (2016b). *ACCUPLACER program manual*. Retrieved from http://professionals.collegeboard.com/profdownload/accuplacer-program-manual.pdf

Complete College America (2012). *Remediation: Higher education's bridge to nowhere*. Retrieved from http://www.completecollege.org/docs/CCA-Remediation-final.pdf

Complete College America (2013). *The game changers: Are states implementing the best reforms to get more college graduates?* Retrieved from http://completecollege.org/pdfs/CCA%20Nat%20Report%20Oct18-FINAL-singles.pdf.

Couturier, L., & Cullinane, J. (2015). *A call to action to improve math placement policies and processes: Six policy recommendations to increase STEM student aspirations and success while decreasing racial and income gaps*. Jobs for the Future, The Charles A. Dana Center, Achieving the Dream. Retrieved from https://files.eric.ed.gov/fulltext/ED559677.pdf

Denny, J. K., Nelson, D. G., & Zhao, M. Q. (2012). Creating and analyzing the effectiveness of a mathematics placement policy for new freshmen. *PRIMUS*, *22*(3), 177-185.

DesJardins, S. L., & Lindsay, N. K. (2008). Adding a statistical wrench to the "toolbox". *Research in Higher Education*, *49*, 172-179.

Dorans, N. (1999). *Correspondences between ACT and SAT I scores*. ETS Research Report Series 1999(1). Retrieved from https://www.ets.org/Media/Research/pdf/RR-99-02-Dorans.pdf

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101.

Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, *104*(2), 439–451.

Edgecombe, N. (2011). *Accelerating the academic achievement of students referred to developmental education*. CCRC Working Paper No. 30. Community College Research Center, Columbia University. Retrieved from https://ccrc.tc.columbia.edu/media/k2/attachments/accelerating-academic-achievement-students.pdf

Educational Testing Service (2015). *2014 Annual statistical report on the HiSET® exam*. Retrieved from https://hiset.ets.org/s/pdf/2014_annual_statistical_report.pdf

Fields, R., & Parsad, B. (2012). *Tests and Cut Scores Used for Student Placement in Postsecondary Education: Fall 2011*. Washington, DC: National Assessment Governing Board. Retrieved from http://files.eric.ed.gov/fulltext/ED539918.pdf

Fong, K. E., Melguizo, T., & Prather, G. (2014). Increasing success rates in developmental math: The complementary role of individual and institutional characteristics. *Research in Higher Education*, *56*(7), 719-749.

Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes.* Research & Occasional Paper Series: CSHE. 6.07. Center for Studies in Higher Education. Retrieved from https://files.eric.ed.gov/fulltext/ED502858.pdf

Geiser, S., & Studley, W. R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, *8*(1), 1-26.

Gerlaugh, K., Thompson, L., Boylan, H., & Davis, H. (2007). National Study of Developmental Education II: Baseline data for community colleges. *Research in Developmental Education*, *20*(4), 1-4.

Hanley, J., & McNeil, B. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*, 839-843.

Hasite, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction.* New York: Springer.

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression*. New York: John Wiley & Sons.

Hughes, K. L., & Scott-Clayton, J. (2011). Assessing developmental assessment in community colleges. *Community College Review, 34*(4), 327-351.

Jabareen, Y.R. (2009). Building a conceptual framework: Philosophy, definitions, and procedure. *International Journal of Qualitative Methods, 8*(4), 49-62.

Jaggars, S. S., & Hodara, M. (2011). *The opposing forces of that shape developmental education: Assessment, placement, and progression at CUNY community colleges*.

CCRC Working Paper No. 36. Community College Research Center, Columbia University. Retrieved from https://ccrc.tc.columbia.edu/media/k2/attachments/opposing-forces-shape-developmental.pdf

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *29*(2), 119-127.

Kuh, G., Cruce, T., Shoup, R., Kinzie, J., & Gonyea, R. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education*, *79*(5), 540-563.

Landis, J., & Koch, G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159-174.

Le, H., Casillas, A., Robbins, S., & Langley, R. (2005). Motivational and skills, social, and self-management predictors of college outcomes: Constructing the student readiness inventory. *Educational and Psychological Measurement*, *65*(3), 482-508.

Lewallen, W. C. (1994). *Multiple measures in placement recommendations: An examination of variables related to course success*. Lancaster, CA: Antelope Valley College. Retrieved from https://files.eric.ed.gov/fulltext/ED381186.pdf

Long Beach City College (n.d.). *Understanding your test results*. Retrieved from https://www.lbcc.edu/post/understanding-your-test-results

Long Beach City College (2014). *Overview of results for the fall 2013 promise pathways*. Long Beach City College Office of Institutional Effectiveness Research Brief. Retrieved from https://www.lbcc.edu/sites/main/files/file-attachments/first-term_results_of_f2013_cohort_promise_pathways_students_-_final_02122014.pdf?1495650985

Long Beach City College (2016). *Long Beach City College Promise Pathways*. Retrieved from https://www.lbcc.edu/post/lbcc-promise-pathways

Martorell, P., & McFarland, I. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *The Review of Economics and Statistics*, *93*(2), 436-454.

Maruyama, G. (2012). Assessing college readiness: Should we be satisfied with ACT or other threshold scores? *Educational Researcher*, *41*(7), 252-261.

Mattern, K. D., & Packman, S. (2009). *Predictive validity of ACCUPLACER scores for course placement: A meta-analysis.* Research Report No. 2009-2. New York: College Board. Retrieved from https://files.eric.ed.gov/fulltext/ED561046.pdf

Medhanie, A. G., Dupuis, D. N., LeBeau, B., Harwell, M. R., & Post, T. R. (2012). The role of the ACCUPLACER mathematics placement test on a student's first college mathematics course. *Educational and Psychological Measurement*, *72*(2), 332-351.

Melguizo, T., Bos, J., & Prather, G. (2001). Is developmental education helping community college students persist? A critical review of the literature. *American Behavioral Scientist*, *55*(2), 173-184.

Metropolitan Community College (2016). *FOCUS developmental education summary. Metropolitan Community College Office of Institutional Research and Assessment.* Retrieved from http://www.mcckc.edu/institutional-research-assessment/research/FOCUS%20Summary%20Success.pdf

Metropolitan Community College (2017). *Factbook 2017*. Metropolitan Community College Office of Institutional Research and Assessment. Retrieved from http://www.mccckc.edu/institutional-research-assessment/research/District%20Fall%20MCCKC%20Facts%20Information.pdf

Missouri Code of State Regulations 6 CSR 10-2 (2017). *Rules of Department of Higher Education, division 10 – commissioner of higher education, chapter 2 – student financial assistance program (6 CSR 10-2).* Retrieved from https://www.sos.mo.gov/cmsimages/adrules/csr/current/6csr/6c10-2.pdf

Missouri Department of Higher Education. (n.d.). *Principles of best practices in remedial education* (CBHE Publication). Jefferson City, MO. Retrieved from https://dhe.mo.gov/policies/documents/BestPracticesinRemedialEducationPolicy.pdf

Morrissey, S., & Liston, C. (2012). *Reconsidering how to place students enrolling in North Carolina's community colleges.* (North Carolina Community College System Office Academic and Student Services Policy Brief). Retrieved from https://www.nccommunitycolleges.edu/student-services/multiple-measures

Moss, B. G., Yeaton, W. H. & Lloyd, J. E. (2014). Evaluating the effectiveness of developmental mathematics by embedding a randomized experiment within a regression discontinuity design. *Educational Evaluation Policy and Analysis*, *36*(2), 170-185.

National Student Clearinghouse Research Center. (2013). *Completing college: A national view of student attainment rates – Fall 2007 cohort.* Retrieved from https://nscresearchcenter.org/wp-content/uploads/NSC_Signature_Report_6.pdf

New Mexico State University. (2016). *New Mexico State University English and mathematics placement.* Retrieved from https://www.math.nmsu.edu/msc/MPE/Placement.pdf

Ngo, F., & Kwon, W. W. (2015). Using multiple measures to make math placement decisions: Implications for access and success in community colleges. *Research in Higher Education*, *56*, 442-470.

Noble, J., & Sawyer, R. (2004). Is high school GPA better than admission test scores for predicting academic success in college? *College and University*, *79*(4), 17-22.

North Carolina Community College System. (2016). *NCCCS policy using high school transcript gpa and/or standardized test scores for placement.* North Carolina state board of community colleges. Retrieved from http://www.nccommunitycolleges.edu/sites/default/files/academic-programs/crpm/attachments/section26_16aug16_multiple_measures_of_placement.pdf

Parsad, B., Lewis, L., & Greene, B. (2003). *Remedial education at degree-granting postsecondary institutions in fall 2000* (NCES 2004-101). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2004/2004010.pdf

Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, *133*, 225-245.

Peng, C., Lee, K., & Ingersoll, G. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1), 3-14.

Peng, C., & So, T. (2002b). Logistic regression analysis and reporting: A primer. *Understanding Statistics*, *1*(1), 31-70.

Peng, C., So, T., Stage, F., & St. John, P. (2002). The use and interpretation of logistic regression in higher education journals: 1988–1999. *Research in Higher Education, 43*, 259–293.

Perry, M., Bahr, P., Rosin, M., & Woodward, K. M. (2010). *Course-taking patterns, policies, and practices in developmental education in the California community colleges.* Mountain View, CA: EdSource. Retrieved from https://mk0edsource0y23p672y.kinstacdn.com/wp-content/publications/FULL-CC-DevelopmentalCoursetaking.pdf

Porchea, S. F., Allen, J., Robbins, S., & Phelps, R. P. (2010). Predictors of long-term enrollment and degree outcomes for community college students: Integrating academic, psychosocial, sociodemographic, and situational factors. *The Journal of Higher Education*, *81*(6), 750–778.

Ramaswami, M. & Bhaskaran, R. (2010). A CHAID based performance prediction model in educational data mining. *International Journal of Computer Science Issues*, *7*(1), 10-18.

Reeves Bracco, K., Dadgar, M., Austin, K., Klarin, B., Broek, M., Finkelstein, N., . . . Bugler, D. (2014). *Core to college evaluation: Exploring the use of multiple measures for placement into college-level courses – Seeking alternatives or improvements to the use of a single standardized test.* WestEd. Retrieved from https://www.wested.org/wp-content/files_mf/1397164696product55812B.pdf

Robbins, S., Allen, J., Casillas, A., Peterson, C., & Le, H. (2006). Unraveling the differential effects of motivational and skills, social, and self-management measures from traditional predictors of college outcomes. *Journal of Educational Psychology*, *98*(3), 598-616.

Robbins, S., Lauyer, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 271-288.

Roska, J., Jenkins, D., Jaggars, S. S., Zeidenberg, M., & Cho, S. (2009). *Strategies for promoting gatekeeper course success among students needing remediation: Research report for the Virginia community college system.* Retrieved from http://ccrc.tc.columbia.edu/media/k2/attachments/strategies-promoting-gatekeeper-course.pdf

Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics*, *121*, 297-317.

Ruffalo Noel Levitz. (2014). *2014-24 projections of high school graduates by state and race/ethnicity, based primarily on data from WICHE.* Coralville, Iowa: Ruffalo Noel Levitz. Retrieved from www.RuffaloNL.com/Demographics

Sanchez, E. I. (2013). *Differential effects of using ACT college readiness assessment scores and high school GPA to predict first-year college GPA among racial/ethnic, gender, and income groups.* Retrieved from http://www.act.org/research/researchers/reports/pdf/ACT_RR2013-4.pdf

Sawyer, R. (2013). Beyond correlations: Usefulness of high school GPA and test scores in making college admissions decisions. *Applied Measurement in Education*, *26*, 89-112.

Schanzenbach, D. (2012). Limitations of experiments in education research. *Education Finance and Policy*, *7*(2), 219-232.

Schmitt, N. (2012). Development of rationale and measures of noncognitive college student potential. *Educational Psychologist*, *47*(1), 18-29.

Scott-Clayton, J. (2012). *Do high-stakes placement exams predict college success?* CCRC Working Paper No. 41. Community College Research Center, Columbia University. Retrieved from https://ccrc.tc.columbia.edu/media/k2/attachments/high-stakes-predict-success.pdf

Scott-Clayton, J., & Rodriguez, O. (2012). *Development, discouragement, or diversion? New evidence on the effects of college remediation.* National Bureau of Economic Research Working Paper No. 18328. Retrieved from https://www.nber.org/system/files/working_papers/w18328/w18328.pdf

Scott-Clayton, J., Crosta, P. M., & Belfield, C. R. (2014). Improving the targeting of treatment: Evidence from college remediation. *Educational Evaluation and Policy Analysis*, *36*(3), 371-393.

Sedlacek, W. E. (2004). *Beyond the Big Test: Noncognitive assessment in higher education*. San Francisco: Jossey-Bass.

Shapiro, D., Dundar, A., Ziskin, M., Yuan, X., & Harrell, A. (2013). *Signature report #6, Completing college: A national view of student attainment rates.* National Student

Clearinghouse Research Center. Retrieved from https://nscresearchcenter.org/wp-content/uploads/NSC_Signature_Report_6.pdf

Steyerberg, E., Bleeker, S., Moll, H., Grobbee, D., & Moons, K. (2003). Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology, 56*, 441-447.

Vandal, B. (2014). *Promoting gateway course success: Scaling corequisite academic support.* Complete College America. Retrieved from https://files.eric.ed.gov/fulltext/ED558791.pdf

Warner, R. M. (2012). *Applied statistics: from bivariate through multivariate techniques*. Sage.

Westrick, P. A., & Allen, J. (2014). *Validity evidence for ACT Compass placement tests.* ACT Research Report Series (2). Retrieved from https://www.act.org/research/researchers/reports/pdf/ACT_RR2014-2.pdf

Westrick, P. A., Le, H., Robbins, S., Radunzel, J., & Schmidt, F. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES. *Educational Assessment*, *20*(1), 23-45.

Wilkinson, L. (1992). *Tree structured data analysis: AID, CHAID, and CART*. Paper presented at the 1992 Sun Valley, ID, Sawtooth/SYSTAT Joint Software Conference. Retrieved from https://www.cs.uic.edu/~wilkinson/Publications/c&rtrees.pdf

Willett, T. (2013). *Student transcript-enhanced placement study (STEPS) progress report.* The Research & Planning Group for California Community Colleges. Retrieved from http://rpgroup.org/sites/default/files/STEPSTechnicalReport.pdf

VITA

William Parker Morgan IV was born on January 27, 1980 in Springfield, Missouri. Before completing his Ph.D. in Curriculum & Instruction and Mathematics at the University of Missouri – Kansas City, he attended Missouri State University, Springfield, where he earned a Bachelor of Science in Mathematics Education in 2003. He also attended the University of Arkansas, Fayetteville, where he earned a Master of Science in Mathematics in 2005.

After completing his master's degree, William taught mathematics at Monett High School in Monett, Missouri and at Benton High School in St. Joseph, Missouri. In 2008, he assumed a faculty teaching position at Metropolitan Community College – Maple Woods in Kansas City, Missouri. From 2014 to 2018, William worked as the Curriculum Developer/Activity Director for the U.S. Department of Education Title III grant at Metropolitan Community College – Blue River in Independence, Missouri. In 2018, he returned to Metropolitan Community College – Maple Woods as mathematics faculty and mathematics discipline coordinator. In 2019, William assumed the role of Division Chair of Mathematics, Physics, and Communication at Metropolitan Community College – Maple Woods. In 2020, he began teaching as an adjunct faculty in the mathematics department at the University of Missouri – Kansas City in addition to his work at Metropolitan Community College.