ABSTRACT

FRYE, BOBBIE JEAN. Using Propensity Score Matching to Model Retention of Developmental Math Students in Community Colleges in North Carolina. (Under the direction of Dr. James Bartlett II.)

Traditionally, modeling student retention has been done by deriving student success predictors and measuring the likelihood of success based on several background factors such as age, race, gender, and other pre-college variables, also known as the input-output model. Increasingly, however, researchers have used mediating factors of the student and thereby recognized that modeling student behavior requires not only an examination of pre-college characteristics, but also a study of the factors and behaviors that mediate between entry and successful completion of an academic program (Bahr, 2013). This study demonstrated the use of propensity score matching to model student retention of developmental math students in community colleges in North Carolina, specifically in terms of student-level and institutional-level predictors.

Participants were students who had been referred into one or more areas of developmental math coursework and who were enrolled in at least one developmental math course during the study period. The population consisted of 2007-2008 new student cohorts at North Carolina community colleges. The student record datasets used in this study included demographics; course enrollment and completion; developmental placement; financial aid data; and transfer information available through the National Student Clearinghouse. In order to demonstrate the practical use of propensity score matching in retention research, propensity score matching was used to create equivalent study and comparison groups in terms of predictors at the academic, student (Titus, 2007), and institutional levels. Multilevel propensity matching was used to create two equivalent groups

of students matched on the propensity to complete developmental math and to pass college-level math with a C or better.

The purpose of this study was to determine if there was a difference in outcomes between a comparison group of developmental math students who completed developmental math and attempted, but did not succeed, in college-level math, and a study group of developmental math students who completed developmental math and then attempted and succeeded in college-level math with a grade of C or better. The results indicated that, after matching, the average means differed between the two groups on key progress indicators. Specifically, the study group fared better than the comparison group and completed, on average, 25 more college credits at the community college. Completers of college-level math earned significantly more associate degrees than non-completers of college level math. Transfers to four-year and two-year institutions were common in both groups of students, and the study group (completers of college-level math) was twice as likely to transfer out of the institution. The results suggested that completion of developmental math and successful completion of college-level math were significant factors in student success outcomes in both student-level and institutional-level contexts. This study also confirmed variation between colleges, implying that administrators and policy makers need to strive to increase the number of students that are retained and complete college-level math with a grade of C or better. Future research examining institutional variation is needed to help explain the contextual variation that yielded different outcomes for students between institutions.

Using Propensity Score Matching to Model Retention of Developmental Math Students in Community Colleges in North Carolina

by
Bobbie Jean Frye

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Education

Adult and Community College Education

Raleigh, North Carolina

2014

APPROVED BY:

Dr. James E. Bartlett II Committee Chair	Dr. Timothy G. Hatcher
Dr. Dan B. Croom	Dr. Michelle E. Bartlett

DEDICATION

I would like to dedicate this dissertation to Eric, for his patience and support through the years. I would also like to dedicate my work to my daughters, Samantha, Jennifer, and Allison, and to my granddaughters, for all of your love and support.

BIOGRAPHY

Bobbie Jean Frye is a first-generation college student from North Carolina. Her educational journey began years ago at Central Piedmont Community College in Charlotte, NC. She later transferred to the University of North Carolina at Charlotte, where she graduated with both a Bachelor's and Master's degree in the field of Sociology.

Bobbie is currently the Director of Institutional Research at Central Piedmont Community College (CPCC) in Charlotte, NC. Her passion for community colleges stems from her desire to ensure all students have the opportunity to complete a degree in higher education and to grow and develop as productive citizens. She serves in several advisory and leadership roles for student success initiatives at the community college, and she is also actively involved in the Completion by Design and Achieving the Dream initiatives, both locally at CPCC and nationally. Her work has focused on the role of institutional research in a community college context, and her research interests include developmental education, low income, and underrepresented community college students. Her current research focuses on developmental education and the use of propensity matching research design in intervention and retention research.

ACKNOWLEDGEMENTS

I would like to thank my dissertation chair, Dr. James Bartlett for his guidance during the dissertation process and useful comments. I also acknowledge the contributions of my committee members, Dr. Timothy Hatcher, Dr. Michelle Bartlett and Dr. Dan Croom.

I extend a special thanks to Dr. Brad Bostian who introduced me to propensity score matching methods with his own dissertation topic. Dr. Bostian's willingness to share his work was invaluable to me.

Throughout this process I have had support and encouragement from my colleagues engaged in the student success work at Central Piedmont Community College. Thanks are extended to Dr. Pat West, Rita Dawkins, Denise Wells, Dr. Terri Manning and to all of the Planning & Research Team at CPCC, for your willingness to listen to my ideas and to offer suggestions for further investigation.

Finally, this project would not have been possible without the cooperation of the community colleges to share data sets derived during the Completion by Design work. The data sets provided the means to utilize a multilevel design and allowed my dissertation topic to come to fruition. I am grateful for your cooperation and approval of my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

In their examination of community college faculty, Twombly and Townsend (2008) noted that American society increasingly expects that individuals earn a bachelor's degree. The authors linked this expectation to the increasing demand on the community college to provide the first two-years of higher education to Americans seeking the bachelor's. Furthermore, the authors noted that approximately half of all college freshman and sophomores and about 37% of all undergraduates are educated in America's community colleges. According to the American Association of Community Colleges (2010), in the fall of 2007 more than 6.8 million students were enrolled in credit courses in community colleges. Additionally, more than 50% of U.S undergraduate students take or have taken courses at community colleges, and first generation, minority and low income students find in their local community college the most accessible path to postsecondary education. Two-thirds of community college students attend part-time and the average student is 29 years of age. Community colleges educate the majority of Black and Hispanic undergraduates, and since the 1980s, women comprise more than half of the community college student body.

"Though a primary point of access to higher education, community colleges struggle to ensure that all of their students earn a credential" (Kolenovic, Linderman, & Karp, 2013, p. 272). While there are varying definitions of student success in the community college (Wild & Ebbers, 2002), foundations, researchers, federal and state governments have focused national attention on persistence and retention of students in terms of certificate or degree completion and/or a successful transfer to a four-year institution (O'Banion, 2013). In a study of community college enrollment and completion, 45% of community college students had

earned a certificate or degree or had transferred to a four-year institution within four-years (Bailey, Jenkins, & Leinbach, 2005). Additional resources to improve student outcomes at the community college have been made possible in part by the philanthropic efforts of foundations (Burdman, 2009). The Bill and Melinda Gates Foundation' supports the student success work of community colleges in North Carolina utilizing a Completion by Design (CBD) framework that demonstrates the need to study student progression from entry to completion (CBD, 2011). The Gates' foundation, to date, provided 35 million dollars to support the work of the national college completion movement (CBD, 2011). The financial assistance can provide an impetus for educational reform efforts in the community college. In many respects, the assistance to community colleges provided needed resources to community colleges struggling with economic financial constraints and external demands from public taxpayers, federal and state governments to improve student success while simultaneously curtailing costs (Burdman, 2009).

One of the most well-known theories of persistence and retention is Tinto's theory of college departure. Tinto (1993) established a theory of student departure that is longitudinal and complex in design. The conceptual framework is based on the student and academic factors associated with retention in the educational system. A multi-campus student experience is understood in the examination of student persistence. Tinto's theory focuses on the extent to which the interactions of the individual characteristics such as family background, academic success, personality, values, and institutional characteristics like academic and social systems impact retention. Tinto's model of student retention takes into

account the shared responsibility of the student and the institutions in college retention (Tinto, 1993). Tinto (1993) noted that his framework was less applicable to students who combine work and schooling or to students who do not live on campus and Tinto's research focused on retention of university students. Moreover, the theory presented by Bean and Metzner (1985) has been used to study factors that impact decisions to stay or leave postsecondary education among non-traditional students.

In a community college context, students who begin work in developmental math courses are a significant proportion of students in the community college (Bailey, 2009; Bailey, Jeong, & Cho, 2010). Achieving the Dream (ATD), a national community college student success initiative addressing low income and under-represented populations, found that 62% of full-time students in community colleges need developmental math. In addition, 34% of full-time students need developmental English and 35% need developmental reading (Clery & Topper, 2008). In comparison to the ATD colleges, a National Educational Longitudinal Study (NELS) examined 8th grade students from 1988 to 2000. The researchers found 58% of those students who attended a community college took at least one remedial course, 44% took between one and three courses, and 14% took more than three remedial courses (Attewell, Domina, Lavin, & Levey, 2006). According to U.S. Department of Education statistics, in 1992, 61% of twelfth graders who first attended a community college required at least some remediation, compared with 64.5% of students who only attended a community college (Adelman, 2005).

The focus of this study was the developmental math student's progression and retention from developmental math courses to the college-level gateway math course, as well as course completion and credential completion outcomes. Developmental education is perhaps the most challenging and important problem facing community colleges (Bailey, 2009). Researchers have noted "remedial [developmental] courses have been a fixture in American community colleges since these institutions first appeared in post-secondary education at the turn of the 20th century" (as cited in Boylan & Saxon, 1999, p. 1). Developmental education programs are designed to provide students who lack prerequisite academic skills the means to remediate skills and progress to college-level coursework in the deficit subject areas such as reading, English, and math (Bailey et al., 2010). Community colleges have "open-door" admittance policies that allow for students not prepared for college-level work to enroll in college programs. As such, community college students often need remediation and academic assistance to succeed in college (Boylan & Saxon, 1999).

While it has long been recognized that one implication of the "open-door" institution is the enrollment of underprepared students (Roueche & Hurlburt, 1968) , the open door philosophy sparks an ethical debate in the community college regarding the underprepared student. Proponents posit that the institution of the community college provides hope for the "democratization" of education (Bahr, 2008). Despite its challenges in developmental education, community colleges provide an opportunity for students who did not do well in high school to improve labor market advantage (Cohen & Besharov, 2002). Critics charge that public taxpayers and federal and state governments are paying for information that was

covered, or should have been covered, in the high school course of study. According to critics, offering remediation in publicly-funded institutions like community colleges burdens taxpayers by having them pay twice to educate the student, thereby making remediation a poor use of scarce resources (Bahr, 2008).

Remediation is an expensive repair of the educational pipeline, and the cost has increased over time. Earlier studies indicated remediation cost public colleges and universities more than $1 billion dollars annually (Breneman & Harlow, 1998). A 2008 study indicated the annual cost of remediation is around $1.9-2.3 billion in community colleges and an additional $500 million at four-year colleges (Strong American Schools, 2008). Some states do not allow developmental education in the university (Oudenhoven, 2002). Clearly, such costs data indicate that developmental education programming is concentrated in the community colleges (Bettinger & Long, 2005b). In addition to financial costs, there are social and personal costs for students who move through the developmental education programming sequences (Bailey et al., 2010; Pretlow & Wathington, 2011).

Nature of Problem

The need to study developmental education is not a recent concern. Researchers have been researching the impact of developmental educational programs for over 30 years (Levin & Calcagno, 2008). Among the research, two national studies examined students using the NELS datasets (Adelman, 1999; Attewell et al., 2006) and state-level studies have examined a few noteworthy states in Ohio, Texas, Florida, California, and Virginia (Bahr, 2008, 2009, 2010a, 2010b; Bettinger & Long, 2005a; Calcagno & Long, 2008; Martorell & McFarlin Jr.,

2007; Roksa, Jenkins, Jaggars, Zeidenberg, & Cho, 2009). The ATD national data set composed of the initial ATD colleges has also been examined (Bailey et al., 2010). The studies varied in their approaches and methodologies to determine the nature, extent, and effectiveness of remediation in the community college.

Two national studies have investigated the incidence and impact of remediation in community colleges. A quantitative study used the national educational longitudinal 1988 dataset with a sample of 6,879 students who were followed since eighth grade. Using a propensity-matched study that simulates an experimental design, Attewell et al. (2006) found that attending a two-year college affected one's likelihood of remediation. The probability of remedial placement was 11% higher for students in two-year institutions compared to similar students in four-year institutions. The researchers also found an 11% higher probability of remediation for Black students compared to White students, but low income status was not a significant indicator of remedial placement. The association between institutional type, remedial status, and degree completion was also investigated using the matching technique. Overall, remedial students from two-year colleges had a 25% chance of graduating, while similar students from four -year colleges had a 50% chance of graduating but the propensity matched study only considered pre-college variables which is an important distinction to be made.

Adelman (1999) tested the impact of taking many remedial courses on college completion and found a negative impact of remediation. In this research, the National Educational Longitudinal Study (NELS) dataset was used to compare the graduation rates of

students who had taken remedial courses compared to students who had not. Adelman found lower graduation rates for the students who took more developmental courses. Additionally, it was found that low-income students were more likely to have a less rigorous curriculum in high school, a variable found to be a highly influencing factor in collegiate success (Adelman, 1999).

State-level studies have examined a few noteworthy states in Ohio, Texas, Florida, California, and Virginia. Bettinger and Long (2005a) conducted a statewide data analysis of community college students in Ohio. For a five-year period from 1998 to 2003, researchers examined 5,807 first-time, traditional aged, 18-20 years old students who had taken the ACT. Full-time and part-time remedial students were compared to full-time and part-time non-remedial students in terms of credits completed, transfers, and credentials earned. They found that full-time remedial students earned 5.4 fewer credits, were 15% more likely to have left college without a two-year degree and 3.6% more likely to have left college without a four-year degree, compared to full-time non-remedial students. At the same time, part-time remedial students, when compared to part-time non-remedial students, were less likely to complete two-year or four-year credentials or transfer by spring of 2003. The researchers thus concluded students in remediation had poorer outcomes than students not in remediation (Bettinger & Long, 2005a). Notably, however, the results were based on simple comparisons that failed to aptly control for selection bias and other student differences.

In a follow-up study designed to better control for selection bias, the researchers used differences among placement policies in Ohio and distance from high school to college to

compare outcomes of students placed into developmental education to similar students not placed into developmental education due to placement policies. The key is that similar students were compared to each other and differences among students were controlled in the study. The researchers found positive outcomes for students placed in math remediation. The students placed into math remediation were 15% more likely to transfer by spring of 2003 than similar students not placed into remediation. In addition, students in math remediation completed ten more credit hours than similar students not placed into math remediation (Bettinger & Long, 2005b).

Martorell and McFarlin Jr. (2011) examined the effects of college remediation on academic and labor market outcomes. Using a sample of entering freshman enrolling between 1991-92 and 1999-2000, students in both two-year and four-year colleges were tracked for up to six years. Using a regression discontinuity approach that examined marginal students, defined as those who barely pass and those who barely fail the placement tests, researchers found higher remediation rates for two-year college students. In addition, remediation for math is more common. At two-year colleges, 23% of students in remediation graduate from college compared to 38% among non-remediated students. In terms of retention, 86% of two-year versus 93% of four-year remedial students was retained the following semester after initial enrollment. In addition, remediation reduced first-year credits earned to 2.4 credits for two-year students and 1.5 credits for four-year students. Overall, the researchers found no positive effects of remediation on two-year and four-year college

students. They concluded remedial offerings in Texas may be ineffective and marginal students gain little academic or earnings benefits from remediation.

In Florida, researchers Calcagno and Long (2008) examined community college students who had taken a college placement test. They compared students placing just below the cutoff placement score to students who were right above the cutoff score, another utilization of regression discontinuity designs. In order to establish consistency in the study, Florida community colleges were excluded that did not use the placement test scores to determine remediation. Out of 130,000 community college students, around 98,146 students (approximately 75%) were tracked for a total of six years. The three cohorts consisted of first-time, associate degree seeking students from 1997-2000. Short-term outcomes examined were the completion of the subsequent college-level course in the remedial subject area and fall-to-fall persistence. Long-term outcomes examined were the completion of a credential, certificate, and or associate degrees, transferring to a four-year university, and total credits earned in remedial and college-level courses.

The researchers found students just below the cutoff score were more likely to persist from fall to fall than students who were right above the cutoff score, remedial and non-remedial students respectively. Also, developmental math students just below the cutoff score earned more total credits over six years than students who were right above the cutoff score. A closer examination of total credits revealed the differences in total credits among remedial and non-remedial students held for remedial credits, but did not hold for college-level credits or credits toward a degree; in other words, there was no difference between the

two groups in college credits earned. In addition, remediation in math had no effect on the completion of college-level math or the completion of subsequent English or math courses. Moreover, among the remedial group, researchers found a negative effect in associate degrees earned and transfer rates. Adult students enrolled in remedial courses experienced slight positive effects (Calcagno & Long, 2008).

Bahr (2008) examined the effectiveness of mathematics remediation among the fall 1995 cohort of 85,894 first-time freshman at 107 California community colleges. Four mutually exclusive groups were examined. Non-remedial college-level math completers and remedial college-level math completers formed two of the groups, and it was hypothesized that the remedial group and non-remedial group would attain credentials or transfer at comparable rates. The other two groups were non-remedial students not completing college-level math and remedial students not completing college-level math. The successful remedial students were hypothesized to attain outcomes that were superior to the non-successful remedial students. In order to be part of the remedial cohort, the remedial students' first math was at a remedial level. The results were reported and compared for the two groups of completers: non-remedial college math completers versus remedial college math completers. The researcher found that non-remedial completers were more likely to transfer without a credential, while remedial completers transferred with a credential. In addition, remedial completers were more likely to complete an associate's degree program without transferring, while both groups were equally likely to earn a certificate and transfer.

Next, the results were reported and compared for the two groups of non-completers: remedial non-completers of college-level math versus non-remedial non-completers of college-level math. The results indicated the remedial non-completers did not complete credentials and did not transfer at comparable rates to the non-remedial non-completers. On the other hand, while the non-remedial non-completers were more likely to transfer than remedial non-completers, their outcomes were "relatively poor" (Bahr, 2008, p. 435). Once college-level variables were added, the outcomes between both completers groups were very similar. Interestingly, remedial non-completers experienced the worst outcomes of the two non-completing groups. Bahr (2008) concluded math remediation works for those who remediate successfully. However, three out of four students did not remediate successfully, and four out of five earned no credential and did not transfer. The conclusion was that few remedial students remediate successfully, but remediation works for those who do succeed in remediation.

In a subsequent study the following year, Bahr (2009) reexamined the fall 1995 cohort of first-time freshman at 107 California community colleges. The final sample size was 68,884 community college students. Bahr (2009) examined whether the breadth and depth of remedial need moderated the extent to which students remediated successfully and obtained comparable outcomes to non-remedial students. The study investigated the impact of students placed into more than one subject area of remediation, defined as the breadth of remediation. In addition, the researcher examined the impact of depth of remediation, or being placed into more than one level below college-level in one or more English or reading

subject areas. The researcher found that remediation works for those who were successful remediators, even after taking into account the breadth and depth of deficiency. This finding was good news for the students who are faced with serious deficiencies in college reading and English subject areas.

However, successful reading remedial students were slightly less likely to transfer without a credential. In addition, successful remedial math students were slightly less likely than college-level math students to complete a terminal vocational degree. By adding the types of associate's degrees to the model, such as college transfer and terminal vocational, Bahr (2009) found successful remedial math students were less likely than college-level math students to attain a vocational credential, and posited that perhaps students who persisted through the developmental math sequence were more determined to earn the four-year degree than students who had not persisted through the developmental math sequence. Bahr (2009) concluded that more study is needed on how developmental education is executed at various colleges.

Next, Bahr (2010a) examined the relationship between retention and math remediation. Taking into account the level of math remediation, the researcher examined persistence, measured by the number of terms enrolled and successful completion of college-level math as a grade of C or better. Bahr (2010a) used a fall 2001 cohort of 28,089 first-time freshman community college students at 105 California community colleges who attended at least two semesters. The researcher found that success in remediation varied across levels of remediation and across levels of persistence. In addition, there was not a linear relationship

between persistence and the numbers of levels a student must complete in order to successfully remediate. Persistence was related to success but, for the students in need of the most remediation, it takes a longer time to remediate successfully, and successful remediation was less likely to occur. Bahr (2010a) thus concluded that remediation decreases the probability of successful outcomes even when taking into account persistence (Bahr, 2010a).

Bahr's (2010b) subsequent study examined racial disparities in postsecondary mathematics remediation. The researcher selected the four most populated racial groups among the fall 1995 first-time freshman cohort in 107 California community colleges. The four groups were White, Black, Asian, and Hispanic with a sample size of 63,147 students. Results indicated the odds of remediating successfully for White students were 3.1 times the odds for Black students and 1.6 times that of Hispanic students. The odds of remediating for Asian students were 1.2 times the odds for White students. Overall, less than 1 in 4 students, or 24.6%, successfully completed a college-level math course in six years. In addition, math skill deficiency was strongly related to successful math completion and substantial differences across math categories were observed across the categories of race. Slightly half (50.3%) of students that entered at the highest level of remedial math remediated successfully compared to 1 in 15 (6.9%) of those students in the lowest level of math remediation. Bahr (2010b) concluded that racial disparities reflect poorer resources in elementary school and its effect on student college readiness.

Data analyses of ATD colleges have revealed consistent patterns of entering students who require developmental education and who experience low completion rates in developmental courses ("Achieving the Dream," 2010; Bailey et al., 2010). In a study of over 256,000 students in 57 ATD colleges, Bailey et al. (2010) examined referral, enrollment, and completion in developmental math and reading. The researchers found that between 33-46% of students actually completed the required sequence of courses. However, as expected, developmental completion rates were negatively related to remediation need levels in math and reading. In addition, about one-third of students avoided developmental coursework by never enrolling in any developmental courses. Interestingly, the results also indicated more students exited their developmental course or sequence because they did not enroll in the first or subsequent course than because they failed or withdrew from a course in which they were enrolled (Bailey et al., 2010). Overall, men, older students, part-time students, and students in vocational programs were less likely to progress through developmental coursework.

In a 2006 study of 5,700 adult students, the Community College Research Center found that older students, those who enter college for the first time at age 25 or older, were more likely to need remediation as a refresher (Calcagno, Crosta, Bailey, & Jenkins, 2006), implying that adult students' remediation needs are different than those of younger students. The finding was especially significant in developmental math. In particular, the researchers suggested an application of Bean and Metzner's 1985 model to older community college students, a model which suggests:

> …in order to retain older students, colleges need to help mitigate the effects of the external pressures on them through flexible scheduling, evening and weekend courses, childcare, distance learning, and other means, rather than focus on providing the sorts of intensive advising, counseling, and other supports that are appropriate for younger students. Colleges should also consider offering accelerated programs and financial support to enable older students to attend full time and thus shorten the time it takes to reach the key milestones on the way to a degree and further education. (Calcagno et al., 2006, p. 28)

Use of Propensity Matching in Educational Research

Propensity score matching is a technique used to simulate an experimental study, with application in the medical field. Typically, randomized medical trials ensured that a medical treatment was provided in a controlled environment with at least two groups of interest; one group was given the treatment and the other group was given a placebo, and the differences in outcomes between the two groups were presumably associated with the treatment. Less expensive non-randomized trials afforded the medical profession a solution to costly randomization of patients, but results were subject to challenges of validity and biases inherent in non-randomized trials. Also, results of non-randomized trials were subject to biased results since patients choosing a treatment were potentially different from patients not exposed to the treatment, and reliance on such reported treatment effects could lead to negative consequences for patients and the medical profession as a whole. Propensity score

matching offered a resolution to the problems of bias and unobserved differences in estimating treatment effects (Rosenbaum & Rubin, 1985).

Propensity score estimation, operationalized by using a logistic regression or probit model, is used to compute a conditional probability that a person would choose a treatment given a composite profile of covariates or characteristics of the person in the sample. A conditional probability is calculated for each person in the sample, using a set of observed characteristics and an actual treatment indicator as the dependent variable, coded in the dataset as “1” for actually receiving the treatment and “0” for not receiving the treatment. Rosenbaum and Rubin (1985) posited that each individual in a treatment group typically has a matched partner in the control, or non-treatment, group. By matching the two individuals based on propensity scores, reliable treatment effects are estimated and can serve as alternatives to randomized trials. The matches proceed across both groups; if no match is found in both groups, the case is discarded and sample size is reduced. However, the alternative propensity matched groups can also be compared, with differences in outcomes attributed to the treatment and not the differences in the two groups, or explained by probabilities that one person would choose the treatment over another.

Analogous to the medical profession, educational researchers have been confronted with a similar dilemma. Educational reform, often characterized as interventions in the educational environment, has been frequently studied and assessed for the purpose of determining whether or not an intervention worked and assessing the potential of scaling up the intervention across the institution (Cho & Karp, 2013). In practice, a group of students

are targeted for the intervention; the use of randomized groups is not common, as they are costly and, in some instances, are even perceived as unethical (Bostian, 2008). More commonly, two groups of students are compared in retrospect in order to determine outcomes given different programming experiences, different demographics, or other relevant differences in the students. While such an examination and comparison of two groups is typically justified, as was the case in the current study's investigation of students completing a college-level math course versus students not completing a college-level math course, the results are nonetheless subject to bias and could potentially lead to costly decisions or interpretations based on erroneous information.

However, by simulating an experiment and matching the two groups of interest based on the conditional probability that a student will choose a treatment, or the likelihood a student is in one group versus another, the application of propensity score matching in the educational field can afford the researcher a solution to such frequently biased results (Rosenbaum & Rubin, 1985). The propensity score is a single score based on a summary of a set of covariates that describe students' characteristics or, in multi-campus data, can include institutional-level constructs (Kelcey, 2011; Kim & Seltzer, 2007).

Recent higher education studies have used propensity score methods to examine the following: the financial return on completing a graduate degree (Titus, 2007); the effect of initial enrollment on degree completion (Reynolds & DesJardins, 2009); the effectiveness of career academies on influencing career choices (Rojewski, Lee, & Gemici, 2010); and the propensity being placed into remedial coursework when comparing two-year to four-year

college students. The association between institutional type, remedial status, and degree completion was also investigated using the propensity matching technique (Attewell, Domina, Lavin, & Levey, 2006).

Bostian (2008) compared outcomes of remedial avoiders to non-avoiders who had transferred to a university and found no difference between the two groups after controlling for background and other academic variables. A study of minority students who avoided math remediation in the first year compared to a similar group of students using propensity matching revealed positive outcomes for the non-avoiders in college-level math completion when compared to the avoiders of remedial math in the first year (Frye, Bartlett, & Smith, 2013).

In a student success framework, the need to examine student retention and progression outcomes is relevant to practitioners in the community college who are increasingly called upon to demonstrate student outcomes in a context of limited resources, time constraints, and urgency to demonstrate returns on investments to funders, state and federal agencies, and taxpayers. Moreover, the institutional researcher is expected to produce practical research that can shape and influence policies, practices, and procedures in an institutional context (Padgett, Salisbury, An, & Pascarella, 2010).

Propensity matching has traditionally been used to study two groups of interest and to create equivalent groups who differ on an important treatment variable (Titus, 2007). In an educational setting, students have different characteristics and, upon entry, make choices about their academic pathways that if examined might provide insight about the likelihood

that the student will persist and complete a credential. Students are also exposed to policies and multiple interventions that reveal the complexity of the college experience. Similarly, community colleges set local policies and practices that reflect student demographics and needs of the community. The interplay of the two levels of analysis— the student and the institution— were modeled in the current research to demonstrate the use of propensity modeling in retention studies.

Problem Statement

Researchers acknowledge the limitations of comparing student performance, progression, or retention in a non-scientific study wherein participants are not randomly assigned nor even equivalent in terms of motivation, intentions, background, or skill level (Titus, 2007). While random selection is the "gold standard" (St. Pierre, 2006), random selection is often impractical, unethical, or resisted in educational settings. Propensity score matching is used to address the counterfactual; that is, what would have happened to a similar group not receiving the treatment through choice or self-selection (Titus, 2007). In fact, failure to control for selection bias is a common concern in the study of developmental education (Bettinger & Long, 2005b). The technique of propensity matching simulates an experimental design, controlling for selection bias and creating almost equivalent experimental and control groups on key indicators. Comparisons of student outcomes using propensity matching has been used to yield less biased results than are derived using simple raw comparisons (Rojewski et al., 2010). Proposed remedies to self-selection bias in educational research need to be explored in terms of feasibility and usefulness, as it is critical

that researchers apply methodologies that control for selection bias. Resources are scarce and monies need to be allocated to educational programs that are making a difference in students' success, retention rates, and long-term outcomes.

Purpose Statement

The purpose of this study was to determine if there is a significant difference in characteristics and college-level outcomes between developmental math students who complete college math coursework compared to students who do not complete such coursework. The study participants were students referred into one or more areas of developmental math coursework and enrolled in at least one developmental math course during the study period. The population consists of 2007-2008 new student cohorts at North Carolina community colleges. The student record datasets included demographics, course enrollment and completion, developmental placement, financial aid data, and transfer information available through the National Student Clearinghouse. All of the students completed college placement tests and were referred to developmental coursework prior to beginning their collegiate studies. In order to demonstrate the practical use of propensity score matching in retention research, propensity score modeling was used in this study to create equivalent study and comparison groups in terms of academic and student predictors.

Propensity score matching is a scientific method used in educational evaluation research to address selection bias and to simulate random selection between study and comparison groups. In order to expand the use of propensity matching methods to include community college contexts, multilevel propensity score modeling was used to create

equivalent study and comparison groups of students in terms of academic and student variables. In addition, a second level of data was examined, and propensity score modeling was used to study developmental math students within and between institutions. In developmental education, given the taxing extent of costs in time and resources on students, faculty, and the institutions, it is imperative that community colleges study the progression of developmental math students. This quantitative study examined students in developmental education at community college institutions, and the researcher created a matched group of students for study purposes. It is possible that institutional variation can provide an understanding of the college context and differences in the implementation of educational programming at various colleges. The identification of successful programming is critical information that colleges can use to drive decisions regarding needed changes in developmental education.

Theoretical Framework

Two theories informed the retention framework for developmental students in the community college: First, Tinto (1993) posited that colleges are similar to small societies or communities. A student becomes integrated in the community by membership in its social and academic systems. However, the society or institution as facilitator of persistence or departure comes into play. It is the interaction of individual characteristics (family background, academic success, personality, values) and institutional characteristics (academic and social systems) that impact retention. The institution is characterized with various members of the academic and social systems that mediate student persistence.

Although the academic and social systems of the college are presented as different conceptual constructs, there are interactions between the two systems that impact persistence (Tinto, 1993).

Integration is closely related to interactions with faculty, staff and students and subsequent impact. Academic integration can occur in the classroom, through contact with faculty, staff, and peer interactions and through extra activities such as joining clubs or establishing membership in professional organizations. Social integration occurs through social activities on and off campus that are connected to the overall goal of degree attainment. Academic and social integration are perceived as separate but interrelated components of the college experience (Tinto, 1993).

The second theory that will be used is the theory of non-traditional undergraduate student attrition that was developed by Bean and Metzner in 1985. Bean and Metzner's (1985) theory has been used to study factors that impact non-traditional students' decisions to stay or leave postsecondary education. Specifically, the theory indicates that an undergraduate student is considered *non-traditional* if the student is more than 24 years of age, enrolled part-time, and/or a commuter student. The rationale for using Bean and Metzner's theory of student attrition is that some non-traditional undergraduates attend college in a community college environment. Although Tinto (1993) recognized that environmental pull factors affect retention, Bean and Metzner (1985) contended that environmental pull factors are salient in understanding the persistence of non-traditional students.

Conceptual Framework

Modeling Student Retention

Retention of developmental math students in the community college was an important concept to consider in the current study for a number of reasons. From an institutional perspective, a majority of students are referred to developmental math (Bailey et al., 2010), and the performance rate of the institution is measured by whether or not these students complete developmental math and subsequently complete the required college-level math course with a grade of C or better (NCCCS, 2013b). Moreover, the financial stability of the institution is dependent on the success of the institution's programs that support students (NCCCS, 2013a). Performance-based funding, a budget approach based on performance, was implemented in North Carolina in 2013, and with it come financial impacts for institutions that do not meet, meet, or exceed performance criteria (NCCCS, 2013b). Public policy makers have advocated accountability, and national attention has been focused on the completion of credentials and/or successful transfers to four-year institutions (Kolenovic et al., 2013; O'Banion, 2013). Even more, community colleges' mission is to ensure that students have a positive college experience and complete their goals and objectives, thus understanding what student or institutional factors promote retention is critical if colleges are to improve and strengthen the college experience for students (Fike & Fike, 2008).

For years, Tinto's theoretical framework of academic integration informed higher education scholars about the importance of giving adequate attention to academic integration variables. Tinto (1993) stated that students come to college with social characteristics (family

background, socio-economic status, and parental educational level) and individual characteristics (age, gender, race, academic ability, and previous school experiences), all of which predict the student's level of commitment upon entry. In order to retain students, they must be integrated into the academic and cultural environment of the college. After 40 years of examining retention in higher education, more is known about different students (low income, class levels, ability levels) and different institutional settings, and context is known to matter (Tinto, 2006).

However, Tinto (1993) noted that his framework was less applicable to students who combine work and schooling or students who do not live on campus and most retention research focused on university students. Moreover, the theory presented by Bean and Metzner (1985) has been used to study factors that impact decisions to stay or leave postsecondary education among non-traditional students. This theory indicates that non-traditional undergraduates have at least one of the following factors: A student is considered non-traditional if the student is more than 24 years of age, enrolled part-time, and/or a commuter student. In contrast, it has been suggested that Tinto's theory is more focused on the traditional student than the non-traditional student. According to Tinto (1993), the social relationships that develop in college serve as important predictors of student retention. However, in the investigation of the application of Tinto's theory to non-traditional students, Ashar and Skenes (1993) found that the social integration that occurred in the classroom through learning experiences explained persistence among non-traditional students in

college. The researchers concluded that adults learn in a social environment that is conducive to social integration, such as smaller class sizes, career integration, and team work.

Pascarella and Chapman (1983) also investigated Tinto's theory of persistence in four- year residential, four-year commuter, and two-year commuter institutions. The effect of institutional commitment was stronger in four-year residential and commuter colleges than in two-year commuter colleges; on the other hand, the effect of goal commitment was stronger in two-year commuter colleges than in the other two types of institutions. In this study's analyses, integration was examined simultaneously with other factors and across different institutions in order to assess the explanatory power of Tinto's model, and its applicability to community college students.

Tinto (1993) posited that colleges are similar to small societies or communities. A student becomes integrated in the community by membership in its social and academic systems. However, the society or institution as facilitator of persistence or departure comes into play. Retention is impacted by the interactions of individual characteristics such as family background, academic success, personality, and values with institutional characteristics like academic and social systems. The institution is characterized with various members of the academic and social systems that mediate student persistence. Although the academic and social systems of the college are presented as different conceptual constructs, there are interactions between the two systems that impact persistence (Tinto, 1993).

Mediators and Covariates

Traditionally, modeling student retention has been done by deriving student success predictors and measuring the likelihood of success based on several background factors such as age, race, gender, and other pre-college variables also known as the input-output model (Bahr, 2013). However, in a study utilizing a quantitative regression model to determine predictors of first year student retention in the community college, several variables were found to be related to student retention: passing developmental courses, taking internet courses, participating in student support services program, receiving financial aid, parents' educational level, and number of hours enrolled/dropped in the first semester. The study population was 9200 first-time college students, from a Texas public urban community college, and data were analyzed over a four-year period (Fike & Fike, 2008). Similarly, Hoyt (1999) defined a regression model that accounted for demographic, goal commitment, academic, and financial support variables and measured the impact of these variables on retention rates. Hoyt (1999) examined retention among remedial students in terms of demographics and enrollment status. Fall 1993, 1994, and 1995 first-time college freshman cohorts with a total population of 7683 students were included in the study. About half of the new students required remediation at various levels.

The researcher concluded that Tinto's theory of attrition is applicable to retention in the community college in that academic integration measured as grade point average affected retention. The researcher also found the largest partial correlation was grade point average, while financial aid yielded the second largest partial correlation. Higher attrition rates were

related to minority status and students who were the first generation in their family to attend college. The study tracked three fall cohorts; results indicated that about 30-35% of students did not come back in the spring, and about 60% were gone by the subsequent fall. However, remediation level was found to be a significant predictor of attrition, in addition to minority status and age (Hoyt, 1999).

Bahr (2010a) examined the relationship between retention and math remediation. Taking into account the level of math remediation, the researcher examined persistence, measured by the number of terms enrolled and successful completion of college-level math as a grade of C or better. Bahr (2010a) used a fall 2001 cohort of 28,089 first-time freshman community college students at 105 California community colleges who attended at least two semesters. The researcher found that success in remediation varied across levels of remediation and across levels of persistence. In addition, there was not a linear relationship between persistence and the number of levels a student must complete in order to successfully remediate. Persistence was related to success but, for the students in need of the most remediation, it required a longer enrollment time to remediate successfully, and successful remediation was less likely to occur. Bahr (2010a) concluded remediation decreases the probability of successful outcomes even when taking persistence into account. In contrast, Umoh, Eddy, and Spaulding (1994) found a positive relationship to retention in developmental math and subsequent college persistence; however, the sample size was 41 students, so the study results cannot be generalized or considered useful (Bahr, 2010a).

Feldman (1993) examined factors associated with one-year retention in a community college utilizing remedial need as an explanatory variable in addition to gender, age, race, enrollment status, goals, and high school grade point average. A sample size of 1140 new students was selected for the study. The researcher concluded there were factors associated with likelihood of retention from fall 1989 to fall 1990. Using a logistic regression methodology, high school grade point average yielded an exponent of .459 or predicted retention. The odds were 1.770 that students 20-24 years old would not be retained. Black students were 1.75 times more likely to drop out than White students. In addition, minority students were more likely to drop out than White students. Among the significant variables, the strongest predictor of student drop out was high school grade point average.

Increasingly, researchers have used mediating factors of the student and recognized that modeling student behavior not only requires an examination of pre-college characteristics, but also the factors and behaviors that mediate between entry and successful completion of an academic trajectory (Bahr, 2013).Consequently, this study employed variables that are relevant to the community college context and the community college student experience. Hence, this study of student retention in developmental education utilized pre-treatment variables such as race, ethnicity, citizenship, age, remediation level, financial aid, and initial enrollment patterns (Bahr, 2013), as well as attainment of milestones to completion, such as first term grade point average, enrollment in developmental courses, enrollment in English and math college level courses, retention to the second major term (Leinbach & Jenkins, 2008), enrollment in student success courses (Cho & Karp, 2013; Fike

& Fike, 2008) and other factors such as institutional factors germane to the college's context that mediate between the input and output equation (Bahr, 2013; Tinto, 1993).

Community college students have a variety of educational goals that researchers take in to account when examining student outcomes and retention. Retention can be measured from an institutional perspective by determining, in a given number of students, how many were retained from the first semester to a subsequent semester. In addition, retention can be measured from a student-centered perspective which is concerned with student outcomes, regardless of whether or not a student transfers to another institution along the way (Attewell, Heil, & Reisel, 2011). The student-centered perspective is critical if student success is at the center of student success efforts (Bahr, 2013), so using propensity score matching to model retention and progression of developmental math students in community colleges from a student-centered perspective was the framework used in the current study.

Level 1: Student Variables

In this study, race categories were a set of dummy variables coded as White, Hispanic, American Indian, Asian, Black, and International Student. Age was a continuous variable; gender was dichotomous. Educational goals were defined using two categories of majors: a college transfer degree and vocational associate degree (Bahr, 2009). There were three types of enrollment status: enrollment status in the first term, first time in college, and late entry. Students enrolled in 12 or more credit hours are considered full-time. First time in college was dichotomous and defined as a first-time college student. Late entry was

dichotomous and defined as 24 years or older at time of entry into college (Bean & Metzner, 1985).

There were four categories of math proficiency in the current study:

- MAT050 – whole numbers, fractions, decimals (4 levels below college)
- MAT060 – applications of fractions, decimals, percents, ratio and proportion, order of operations, geometry (3 levels below college)
- MAT070 – signed numbers, exponents, simplifying expressions, solving equations, graphing, factoring (2 levels below college)
- MAT080 – factoring, rational equations, systems of equations, graphing quadratic and radical functions, complex numbers, variation (1 level below college)

There were two categories of reading proficiency:

- RED 080 - emphasis is placed on vocabulary, comprehension, and reading strategies, and focuses on teaching students to identify main ideas and supporting details
- RED 090 - improves reading and critical thinking skills, with topic including vocabulary enhancement; extracting implied meaning; analyzing purpose, tone, and style; and drawing conclusions and responding to written material (NCCCS, 2013a)

There were four categories of English proficiency:

- ENG 080 - focuses on standard conventions of written English, and works to teach students effective sentence structure, paragraph creation, and other writing skills.
- ENG 085 - develops proficiency with reading and writing at the college-level, and critical reading skills are developed; course is a combination of ENG 080/RED 080

- ENG 090 - focuses on the development of effective paragraphs and adhering to standard conventions of English
- ENG 095 - develops comprehension and analysis skills in relation to evaluating college texts; course is a combination of ENG 090 and RED 090

Academic integration (Tinto, 1993) was the next factor in the model and represented course enrollments, which in this model reflected commitment and course attempts that are conducive to persistence for developmental students. Four dichotomous variables represented course academic commitment or integration: enrollment in the student success course; enrollment in developmental reading; enrollment in developmental English; and enrollment in college-level English. The course commitment variables were dichotomous variables for each of the subject areas. Academic integration was also defined as academic grade point average and retention. Academic grade point average was defined as the first-term grade point average and was a continuous variable. Second term retention is defined as returned 2nd term and the variable was dichotomous. Retention was a continuous measure of the number of semesters the student persisted after the first term. There was one environmental pull factor (Bean & Metzner, 1985) in being a Pell recipient in the first term, a dichotomous variable that represents financial support in college.

There were two categories of outcomes in the model: short-term and long-term outcomes. The short-term outcomes reflected credit attempts, completions, and successful grades of C or better in developmental math (MAT-080, MAT-070, MAT-060, MAT-050); developmental reading; developmental English; and college-level English. In this model, the

short-term course completion outcomes led to the long-term outcomes. The long-term outcomes were highest credential attainment or transfer to another institution and retention.

Level 2: Institutional Variables

Since the current study involved more than one institution, institutional variables were used to provide information about the institutional context (Hofmann, 1997). The rationale was based on the belief that different contexts may yield different impacts of developmental education. Of note is the fact that variation among institutions as it relates to developmental education is not well documented in the literature (Bailey, 2009; Oudenhoven, 2002). Since this study involved multi-campus data, the multilevel composition of students nested within colleges provided the rationale for a multilevel propensity score model (Kelcey, 2011). A multilevel model enabled the researcher to capture institutional variation and provided information about contextual variables that can impact developmental educational programming (Pascarella & Chapman, 1983).

In order to account conceptually for institutional membership and to analyze institutional effects, terms involving institutional-level covariates (Rosenbaum & Rubin, 1985), such as the percent of the fall 2007 student population that was of minority race, were the same for each individual in the institution; in other words, the institutional variables were constant across individuals within the same institution. Finally, given the multilevel structure of the data, an alternate model examined institutional effects within individuals of the same school, utilizing constant categorical institutional-level covariates, dummy coded for each institution.

The following institutional variables were used in the retention model:

1) Percent of math, English and reading remedial students – Martorell & McFarlin Jr. (2007) found that high remedial institutions were less effective in developmental education for students in Texas.
2) Percent of White, Asian, Black, and Hispanic students – Racial concentration has been associated with success in remediation (Bahr, 2010b).
3) Percent of students on Pell assistance – Income level of students has been associated with resources and was an environmental pull factor in the model (Bean & Metzner, 1985).
4) Dummy-coded institutional membership variable -1 represented the student being a member of the institution; the variable 0 indicated the student was not a member of the institution (Kelcey, 2011).

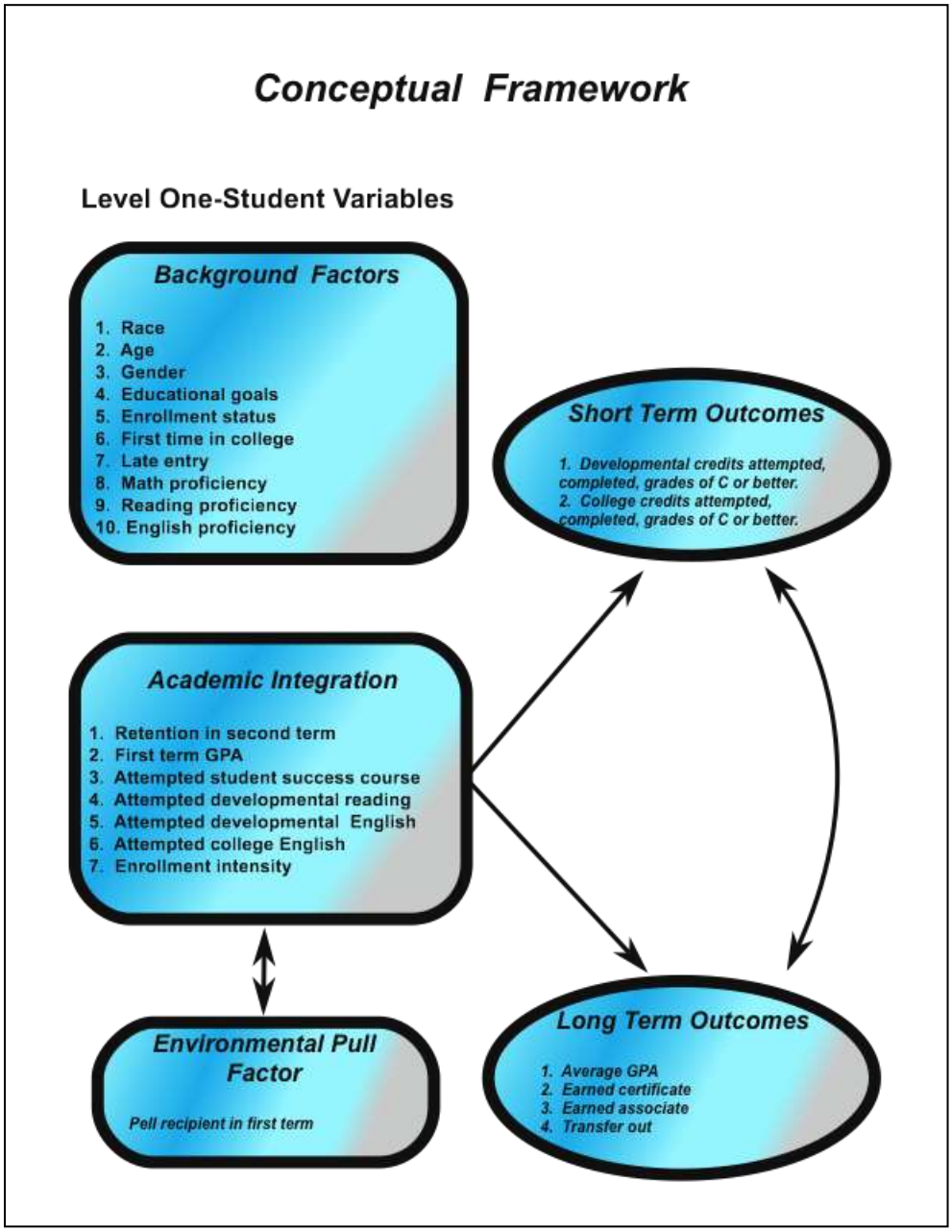

Figure 1: Conceptual Framework Level 1— Student Variables

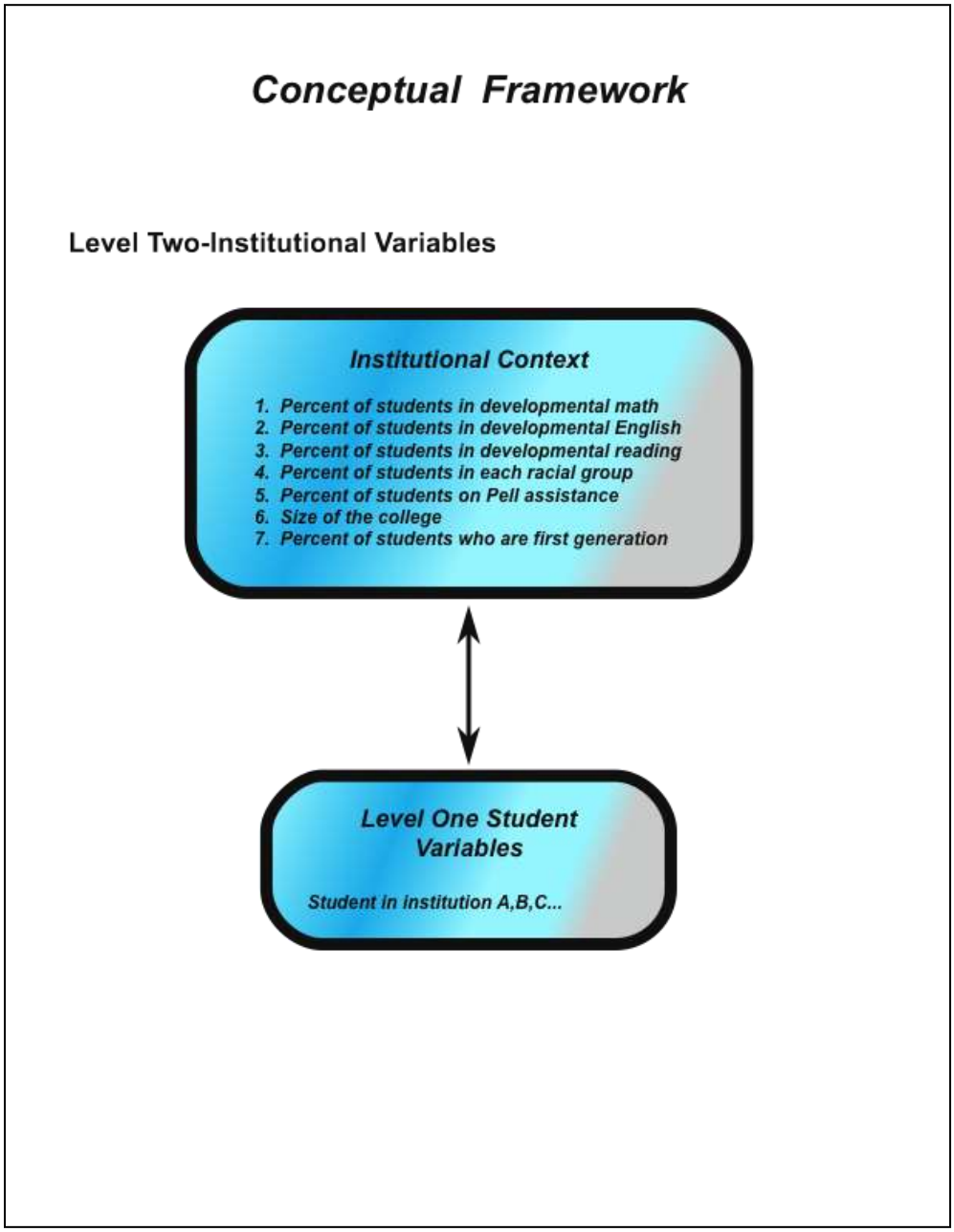

Figure 2: Conceptual Framework Level 2—Institutional Variables

Research Questions and Hypotheses

The study examined the following research questions and hypotheses:

1) What are the demographics and academic characteristics of the study population?

2) What are the demographics and academic characteristics of the two groups in the study?

3) Is there a difference in demographics and academic characteristics of the two study groups prior to propensity score matching?

Hypothesis 1

Ho: There is no difference between students who first enroll in developmental math and the conditional probability of completing with a grade of C or better one college-level math course.

Ha: There is a statistically significant difference between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

4) Is there a difference in demographics and academic characteristics of the two study groups after propensity score matching?

Hypothesis 2

Ho: After propensity score matching, there is a statistically significant difference between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

Ha: There is no difference in between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

5) Is there a difference between the two study groups in college outcomes after propensity score matching?

Hypothesis 3

Ho: There is no difference in student outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

Ha: There are statistically significant outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

Hypothesis 4

Ho: There is no difference in student outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course between institutions.

Ha: There are statistically significant different outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course between institutions.

Significance of the Study

Developmental education is perhaps the most challenging and important problem facing community colleges (Bailey, 2009). In response to this important problem, there have been a few recent sophisticated studies designed to address selection bias and to ensure a rigorous approach to the study of developmental education (Bahr, 2008) in states such as Ohio, Florida, Virginia, and California. The study examined a state that has received little attention in the literature. This study contributed to the rigorous studies by using a propensity score matching, multivariate approach that addresses self- selection bias and accounts for different student characteristics such as low-income, first time in college, or race/ethnicity. In addition, the study examined institutional variance utilizing a multilevel propensity matching methodology.

The purpose of using propensity score modeling was that the method controls for self-selection bias and to simulate experimental comparison and study groups in order to study developmental education in a community college setting. Propensity score modeling was used to create equivalent study and comparison groups in terms of academic and student predictors (Bostian, 2008; Titus, 2007). After a propensity score was derived, matched groups were derived and outcomes compared to determine the average differences between the two groups. The study was significant in that if developmental education research is not conducted effectively and extensively, then researchers, practitioners, and policy makers might make programmatic decisions on incomplete or biased information. In addition, the study can help to inform the ongoing reform efforts in developmental education at the state

level that have led to the redesign of developmental math. The ongoing reform efforts will also lead to the redesign of developmental English and reading, making the study relevant to the broader developmental education topic.

Limitations

There were four limitations to the study. First, no research methodology is free of bias (Bostian, 2008). One objective of propensity matching is to reduce the selection bias in intervention research when researchers compare two different groups and do not control for initial and intermediate differences between the students. There were unobserved differences that were not explained in the model. In other words, the models did not explain or account for all of the variance in the two groups (Titus, 2007). Some important variables such as marital status, number of children, and socio-economic status – variables found to be significant indicators of student success – were not available in the dataset; thus, the missing data introduced the potential for unobserved differences.

Second, the propensity score method reduced the number of cases through the matching procedure. Concern has been raised that the propensity score method eliminates a good number of students who are not part of the analysis. The elimination of cases occurs during the matching process when a match is not found based on the propensity score (Caliendo & Kopeinig, 2008). Multi-level analyses required adequate matched sample sizes of at least 300 cases in each institution. Smaller institutions with fewer than 300 cases were a challenge for the multilevel analyses and sample size was a limitation to the study (Caliendo & Kopeinig, 2008).

Third, timing of developmental courses and timing of the subsequent college-level course were not part of this study. The timing of each sequence of developmental math and the subsequent college math success rates were likely impacted by timing of the courses and the students' ability to schedule courses in sequence and without an interruption of the series of math courses.

Fourth, the research study examined initial student differences that were not explored in the research. Since one of the major objectives of the study was to create equivalent study and comparison groups using propensity matching, there was less focus on the exploration of the initial differences between the groups. Although the intent of the study was to simulate a scientific experimental design, student and academic factors associated with initial membership in the two groups are important to explore in future research.

Delimitations

The study was limited to seven community colleges in North Carolina. The ability to generalize results is questionable, since the community colleges were not randomly selected and the selection process did not follow the rules of scientific research sampling procedures (Bartlett, Kotrlik, & Higgins, 2001). The colleges were selected based on their participation in the Completion by Design Initiative. While the colleges ranged in size from small, medium, to large student populations, there was no indication that the colleges are representative of most community colleges. Studies based on one or a few colleges, while interesting, may lack the methodological rigor of representative samples and sound quantitative research techniques (Creswell, 2008).

Developmental math success and the success of the college-level math course are complex phenomena in the community college. The decision to examine two groups was based on the desired methodology and limits the dependent variable in the initial logistic regression to two values. Research of other groups has been conducted in an attempt to examine and understand the efficacy of developmental math programs. Bahr (2008, 2009, 2010b) studied developmental programs and examined several pairs of developmental math students: 1) takers and completers versus takers and non-completers of developmental math; 2) completers of college-level math between developmental math completers and students who started in college-level math; and 3) racial disparities in developmental program outcomes. The purpose of this study was to examine the success in college-level math for students whose first math course was developmental math. Some important study groups and sub-populations were not examined in this research.

Definitions of Terms

Propensity Score Method (PSM). PSM examines grouped data and the dependent variable is categorical instead of quantitative. Covariate variables are the independent variables with the highest degree of influence on the dependent variable. The use of logistic regression permits the researcher to identify the covariate by using multiple quantitative independent variables to predict the probability of group membership, a dependent variable (Caliendo & Kopeinig, 2008).

Propensity Score. The covariates identified in the logistic regression before PSM are combined into a single summary score whose value ranges between 0.0 and 1.0. A variety of

estimation techniques have been used to determine propensity scores; the appropriate technique depends on the number of study groups being examined (Caliendo & Kopeinig, 2008).

Developmental Education. Developmental education is designed to provide students who lack prerequisite academic skills the means to remediate skills and progress to college-level coursework in the deficit areas (Bailey et al., 2010). Developmental education is often referred to as holistic support and assistance of the developmental student. Remedial developmental courses have been defined as "a series of courses designed to remediate students in specific subject areas (as cited in Boylan & Saxon, 1999, p. 1). Throughout this paper, the terms *developmental education* and *remedial courses* are used interchangeably.

Placement Tests. Developmental students are placed into developmental coursework through a placement test. The cutoff point for developmental versus college-level placement is determined by a cut score or range of acceptable scores.

Remediation Level. The required number of developmental courses in each subject area is referred to as the remediation level. For example, Remediation Level 1 in math would indicate a student requirement to enroll in and pass one developmental course in math (the highest level developmental math course) before proceeding to college-level math. However, some colleges vary in terms of developmental levels and offer one level of developmental math while some offer as many as four levels of developmental math (Bailey et al., 2010).

Program Effectiveness. A measure of developmental program effectiveness is the extent to which passing the developmental course or sequence leads to successful completion of the gatekeeper or gateway course (Bailey et al., 2010).

Course Completion. Course completion means that the student completed the course or did not withdraw from the course.

Credits- Attempted. The total number of credits a student was enrolled in at the ten percent census date.

Credits –Completed. The total number of credits associated with a course in which the student did not withdraw.

Credits -Completed A-C. The total number of credits associated with an official grade of C or better.

Retention. A student is retained from one term to the next if after first enrolling, was enrolled at the ten percent point in the term being evaluated. For example, fall to spring retention means the student was enrolled in the fall term and the student was subsequently enrolled at the ten percent point in the spring. Fall to fall retention indicates that an enrolled student in the fall term was enrolled at the ten percent point in the subsequent fall term.

Study Groups. The two study groups are the dichotomous dependent variable in the initial logistic regression and two values (0,1) are used to indicate group membership. The groups are categorical and can be represented as male/female, successful/non-successful, etc. and form the basis of the dependent variable.

Late Entry. A term referred to a student who entered higher education at 24 years or older.

Summary

A criticism of early 1990s evaluations of developmental education was that many were plagued with selection bias or other methodological problems (Bahr, 2008; Grubb, 2001; O'Hear & MacDonald, 1995). In response to early criticisms, there have been a few sophisticated studies designed to address selection bias and to ensure a rigorous, comprehensive approach to the study of developmental education (Bahr, 2008). Yet of the surprisingly few studies, the results are mixed and inconclusive. Moreover, the studies have methodological limitations to particular states such as California, Texas, Florida, or Ohio and sub-populations of students such as younger, full-time students, or students in ATD colleges. Studies that compare students who barely pass and barely fail placement tests were common among the research. These regression discontinuity studies yield information about students on the cusp of remediation, or marginal students, which may not reflect the entire picture. Failure to control for selection bias in program evaluation was a common concern in the study of developmental program impact and student outcomes (Bettinger & Long, 2005b). In order to make a substantive contribution to the evidence provided to date, empirical research should be rigorous, comprehensive, multi-institutional, and control for selection bias (Bahr, 2010b).

CHAPTER II: LITERATURE REVIEW

Chapter II begins with a review of two prominent theories of retention that were used by the researcher to frame a retention model. Tinto's theory of college departure is well-known and closely connected to academic integration and persistence. Bean and Metzner's theory of retention has been applied to explain the retention of non-traditional students. Both theories informed the retention model in this study.

The next section reviews the developmental education literature. Background information and policy characteristics are provided to set the stage for the context of developmental education as a topic of research. The effectiveness of developmental education is explored in a community college setting and the following factors were examined: demographics, depth and breadth of remedial need, racial disparities, and completion of the gateway courses of college-level math and college-level English, academic attainment, and labor market outcomes.

The chapter concludes with a discussion of propensity matching in educational research and the application of propensity matching in this study. Key findings from the literature examined are provided in the summary.

Theoretical Perspectives on Student Retention

Tinto's Theory of College Departure

One of the most well-known theories of persistence and retention is Tinto's theory of college departure. Tinto's theory of student departure is longitudinal and complex in design. The conceptual framework is based on the factors associated with retention in the educational

system. The extent to which the interactions of the individual characteristics such as family background, academic success, personality, values, with institutional characteristics such as academic and social systems impact retention, is the focus of Tinto's theory. Tinto's model of student retention takes into account the shared responsibility of the student and the institution in college retention (Tinto, 1993).

Tinto's (1993) theory of student departure originated in Van Genepp's anthropological studies in tribal societies in that Van Genepp identified stages that individuals moved through in the passage of adolescence to adulthood. Three stages are associated with movement and transition. First, tribal members separated themselves from their families of origin and from childhood relationships. Second, students transitioned to the new role of adult along with the rights and responsibilities adulthood entails. Last, students became part of the society as a different person, being welcomed in their new membership by rituals and ceremonies indicative of adulthood (Tinto, 1993).

Analogous to the college experience, Tinto (1993) posited that students move through three separate yet overlapping stages if they are to successfully establish membership in the college environment. The first stage of the college career requires students to separate themselves from past communities, families and friends. The student becomes a part of a new environment and culture with norms, values and expectations characteristic of a tribe or community. Students can facilitate the separation by integrating themselves into the social and academic systems of the college. However, for students in non-residential schools, the separation stage is not as pronounced. While students in non-residential schools may not

experience the separation stage to the extent that students in residential schools do, the downside is the students may not enjoy the rich experiences of college life. Hence, students in different institutions experience the first stage differently, and a host of factors can impact the explanatory power of the separation stage in its ability to predict or explain college persistence (Ashar & Skenes, 1993).

In addition, the student's ability or capacity to experience the separation stage successfully is impacted by family expectations and cultural norms of the family unit. It is possible students from disadvantaged backgrounds are from families where college is less valued than is working after high school to help with family expenses. In contrast, a student with college-educated parents is encouraged to attend college and to experience the independence and separation of adult college life. Early on, the message is conveyed that college attendance is valued and expected. The student is encouraged to take full advantage of the gains in expected earnings through college completion and to prepare for membership in a higher standard of living than non-college graduates (Tinto, 1993).

In light of the various factors that interact with the college experience, the second transition stage occurs during and after the separation stage and is typically associated with the first year of college. In the transition stage, the student is learning the new roles and responsibilities of the college student but is in a state of flux. In that the stress of the new environment and uncertainty associated with the new experience can lead to departure from the institution, the transition stage is fraught with challenges. In addition, the transition stage

can be influenced by how prepared the student is for the new role and the degree to which the student is supported and encouraged to persist (Tinto, 1993).

After passing through the stages of separation and transition, persistence is not a certainty. Having moved away from past associations, behaviors, and cultural norms, the student situates into the college life. In other words, the student must integrate themselves into the functional areas of the college. Still, just how a student is able to navigate college life is not fully understood. It is likely that some institutions provide an array of student support services to facilitate the transition to college and the inoculation of students' experiences. Other students navigate through college on their own or become integrated in a haphazard manner. Nonetheless, Tinto (1993) speculated those students who become fully integrated in the social and academic systems of the college remain committed to the goals of completion and are convinced that the institutional environment is congruent with their values and expectations. Students establish connections and bonds with members of the college community and align commitment and goals with those of the institutional administrators, students and faculty. The inability of students to integrate and situate themselves into the social and academic systems of the college is theorized to explain why students stay or leave (Tinto, 1993).

Building on the work of Van Genepp and the three stages of college life, Tinto turned to Emile Durkheim's study of suicide to further explore the causes of student departure from a college community. Tinto (1993) noted that colleges are similar to small societies or communities, as a student becomes integrated in the community by membership in its social

and academic systems. However, the society or institution as facilitator of persistence or departure comes into play. It is the interaction of individual characteristics (family background, academic success, personality, values) and institutional characteristics (academic and social systems) that impact retention. The institution is characterized with various members of the academic and social systems that mediate student persistence. Although the academic and social systems of the college are presented as different conceptual constructs, there are interactions between the two systems that impact persistence (Tinto, 1993).

Integration is closely related to interactions with faculty, staff and students and subsequent impact. Academic integration can occur in the classroom, through contact with faculty, staff, and peers, and through extra activities such as joining clubs or establishing membership in professional organizations. Social integration occurs through social activities on and off campus that are connected to the overall goal of degree attainment. Academic and social integration are perceived as separate but interrelated components of the college experience (Tinto, 1993).

The Non-Traditional Student

Bean and Metzner (1985) put forth a theory that has been used to study factors that impact the decisions to stay or leave postsecondary education among non-traditional students. This theory defines non-traditional undergraduates as such: a student is considered *non-traditional* if the student is more than 24 years of age, enrolled part-time, and/or a commuter student. The conceptual model also includes: 1) background variables: age,

enrollment status, residence, personal goals, educational goals, high school performance, ethnicity, gender; 2) academic variables: study habits, advising experiences, absenteeism, major certainty, course availability; 3) environmental variables: finances, hours of employment, outside encouragement, family responsibilities, opportunity to transfer; 4) Academic Outcomes: grade point average (GPA); 5) Psychological Outcomes: utility, satisfaction, goal commitment, stress; 6) intent to leave; 7) dropout; and 8) social integration (Bean & Metzner, 1985).

Bean and Metzner (1985) delineated a relationship between academic and environmental factors. When both factors are favorable, the students are likely to stay in college; on the other hand, if both factors of academic and environmental factors are poor, students leave college. In addition, they argued that positive environmental factors outweigh poor academic outcomes and students persist even when facing academic hurdles if externally supported. Similarly, there is a relationship between grade point average and psychological factors for non-traditional students. If both grade point average and psychological factors are favorable, the student persists; on the other hand, if both grade point average and psychological factors are poor, students will leave college. In addition, they claimed that poor psychological factors outweigh a poor grade point average, and that students will more likely persist than not persist if satisfied with their educational experience (Bean & Metzner, 1985).

To illustrate the focus that Bean and Metzner (1985) placed on environmental factors as more important to non-traditional students, it is useful to examine a study of non-

traditional students in remedial education. In a 2006 study of 5,700 adult students, the Community College Research Center found older students— those that enter college for the first time at aged 25 or more— were more likely to need remediation as a refresher (Calcagno et al., 2006) and implied adult students' remediation needs are different than those of younger students. The finding was especially significant in developmental math. In particular, the researchers suggested an application of the Bean and Metzner (1985) model to older community college students, which suggests:

> …in order to retain older students, colleges need to help mitigate the effects of the external pressures on them through flexible scheduling, evening and weekend courses, childcare, distance learning, and other means, rather than focus on providing the sorts of intensive advising, counseling, and other supports that are appropriate for younger students. Colleges should also consider offering accelerated programs and financial support to enable older students to attend full time and thus shorten the time it takes to reach the key milestones on the way to a degree and further education. (Calcagno et al., 2006, p. 28)

Thus, in Bean and Metzner's (1985) theory, there is an expected relationship between the external pressures or environmental pull factors and academic integration for the non-traditional student. In contrast, it has been suggested that Tinto's theory is more focused on the traditional student than the non-traditional student. According to Tinto (1993), the social relationships that develop in college serve as important predictors of student retention. However, in their investigation of the application of Tinto's theory to non-traditional

students, Ashar and Skenes (1993) found that the social integration that occurred in the classroom through learning experiences explained persistence among non-traditional students in college. The researchers concluded that adults learn within a social environment that is conducive to social integration, such as smaller class sizes, career integration and team work (Ashar & Skenes, 1993).

Pascarella and Chapman (1983) also investigated Tinto's theory of persistence in four-year residential, four-year commuter, and two-year commuter institutions. The effect of institutional commitment was stronger in four-year residential and commuter colleges than in two-year commuter colleges. On the other hand, the effect of goal commitment was stronger in two-year commuter colleges than in the other two types of institutions. Disaggregating by residential and commuter types, the effect of social integration was significant in four-year residential schools, while the effect of academic integration was more significant in four-year and two-year commuter schools than in residential schools. In commuter schools, academic integration directly impacted institutional commitment and predicted persistence. On the other hand, institutional commitment in residential institutions was largely a function of student interactions with the social system (Pascarella & Chapman, 1983).

Two Theories of Choice

Drawing on persistence theory, one theory that will be used is the model of non-traditional undergraduate student attrition developed by Bean and Metzner (1985), which has been used to study factors that impact the decisions to stay or leave postsecondary education among non-traditional students. This theory indicates that *non-traditional* undergraduates are

those students who are more than 24 years of age, enrolled part-time, and/or a commuter student.

The community college population is non-traditional and students are more likely to attend part-time than full-time. The American Association of Community Colleges (2008) reported that fall 2008 enrollment reflected a diverse adult learner population. In 2007-2008, the average age was 28; 45% of students were ages 22 to 39, while 15% of students were 40 or older. Forty-two percent of students were the first generation in their families to attend college, 13% were single parents, 6% were non-U.S. citizens, 12% were students with disabilities, and 3% were veterans. The enrollment data also indicated, when employment status was reported, that most students were attending part-time while working full-time. Minority students comprise 45% of the community college population, with 13% Black, 16% Hispanic, 6% Asian/Pacific Islander, and 1% Native American students. In fall 2008, community college students were 44% of all United States undergraduates in college.

In addition, "in the last 50 years, there has been a dramatic growth of adult learners (aged twenty-five years and older) in credit and non-credit higher education programs" (Kasworm, Sandmann, & Sissel, 2000, p. 450). Furthermore, 67% of adult learners attend formal education settings part-time due to adult responsibilities such as work and family roles (Kasworm et al., 2000). From 2006-2017, the Department of Education projects a rise of 19% in enrollments of people over age 25 compared to a rise of just 10% in enrollments of people under age 25 (U.S. Department of Education, 2009).

This study also used one of the most well-known theories of persistence and retention, Tinto's theory of college departure. Tinto's theory of student departure is longitudinal and complex in design. The conceptual framework is based on the factors associated with retention in the educational system. The extent to which the interactions of the individual characteristics such as family background, academic success, personality, values, with institutional characteristics such as academic and social systems impact retention is the focus of Tinto's theory. Tinto's model of student retention takes into account the shared responsibility of the student and the institution in college retention (Tinto, 1993) and is thus appropriate in a multilevel, student-centered framework.

Developmental Education

Developmental education is perhaps the most challenging and important problem facing community colleges (Bailey, 2009). In response to criticisms of selection bias and poor research design, there have been several sophisticated studies designed to address selection bias and to ensure a rigorous approach to the study of developmental education (Bahr, 2008). Yet, among the rigorous studies of developmental education, the results are mixed and inconclusive. Moreover, the studies have methodological limitations to particular states such as California (Bahr, 2008, 2009, 2010a, 2010b); Texas (Martorell & McFarlin, Jr, 2007); Florida (Calcagno & Long, 2008); Ohio (Bettinger & Long, 2005a); and Virginia (Roksa et al., 2009). In addition, the findings are limited to sub-populations of students such as younger students (Bettinger & Long, 2005a), or students in ATD colleges (Bailey et al., 2010).

Researchers utilize studies that compare students who barely pass and barely fail placement tests, known as regression discontinuity studies. These regression discontinuity studies yielded information about students on the cusp of remediation, or marginal students, which may not reflect the entire picture (Bailey, 2009a). Failure to control for selection bias in program evaluation is a common concern in the study of developmental program impact and effectiveness (Bettinger & Long, 2005b). This review will analyze the contradictory findings in developmental education and use the literature to critique the findings examined. The major scholars in developmental education and their contributions are explored. The scholars that have been investigating developmental education have impacted the field in various ways, but the fiercely contested debate regarding the effectiveness of developmental education is at the center of fundamental changes that are occurring in community colleges today, such as a statewide math redesign. Nonetheless, there are significant challenges facing community colleges in developmental education, and researchers can contribute and inform the field to address the challenges.

The Effectiveness of Developmental Education Research

Researchers investigating the effectiveness of developmental education have used a variety of outcomes as indicators that the developmental student has remediated successfully; in general, outcomes of interest vary from short-term to long-term outcomes. Developmental education programs are designed to provide students who lack prerequisite academic skills the means to remediate skills and progress to college-level coursework in the deficit subject areas such as reading, English and math (Bailey et al., 2010) . Therefore, one measure of

program effectiveness is the extent to which students remediate skills, progress to college-level coursework, and pass the gatekeeper course(s) in the areas of deficiency. Credits earned while the student is in college are another outcome of interest among researchers. The tendency is to report college credits earned and compare them to remedial credits in an attempt to examine the impact of being placed into remediation in the first place (Bettinger & Long, 2005a). Recognizing that credits earned are a function of enrollment status, researchers Bettinger and Long (2005b) also compared part-time to full-time students (Bahr, 2008, 2009, 2010a, 2010b). Retention in college is another recognized outcome of interest examined by researchers and measured as the number of terms enrolled or enrollment intensity (Bahr, 2010a). Ideally, and in alignment with the student success agenda, the long-term outcomes of interest are graduation and completion in a program of study with a credential that leads to sustainable earnings (Martorell & McFarlin Jr, 2007) or successful transfer (Bettinger & Long, 2005b). Sustained progress and completion of short-term outcomes ostensibly should lead to the completion of long-term outcomes. Due to the nature of the college experience, longitudinal studies are analyzed to answer the long-term questions regarding the effects of effective remediation.

Background: Assessment, Placement and Lack of Consistency

Typically, developmental students are determined by a placement test score; the cutoff point for developmental versus college-level placement is determined by departments that varies from institution to institution or by statewide policies (Bailey et al., 2010; Oudenhoven, 2002; Perin, 2005). As a result, there is little or no consensus among different

community colleges on what constitutes being a “college-ready” student. The number of developmental courses required in each subject area is sometimes referred to as the *remediation level*. For example, Remediation Level 1 in math would indicate a student enrolls in and passes one developmental course (the highest level developmental math) in math before proceeding to college-level math. However, some colleges vary in terms of developmental levels and offer one level of developmental math, while some offer as many as four levels of developmental math (Bailey et al., 2010). In addition, some programs of study exempt students from higher level courses in the recommended remediation levels, while vocational certificates may completely exempt students from developmental math. Moreover, there is variation among developmental programs in terms of voluntary versus mandatory attendance. Although Alfred and Lum (1988) found a negative effect between remediation and subsequent course grades for students in voluntary remedial programs, the researchers, along with Boylan, Bliss, and Bonham (1997), also found a positive effect for persistence and graduation rates for students in mandatory remedial programs.

The lack of consistency in developmental policies and practices led Bailey et al. (2010) to refer to developmental requirements as “mazes” through which students navigate. In practice, the accumulation of developmental credits does not accrue to college credits earned and can delay college completion (Martorell & McFarlin Jr., 2007). As such, the value of developmental courses is not clear to students, and developmental courses are avoided by some students (Grubb & Cox, 2005). Also, the developmental course programmatic requirements can become frustrating for students who must meet them but

who may be surprised to learn they are not earning college credits towards their degree, nor are they even earning credits that are recognized at most four-year institutions (Deil-Amen & Rosenbaum, 2002).

Bailey (2009) recommended that community colleges stop classifying students as *developmental* and *non-developmental*, and instead strive to appropriately assess and diagnose students' remedial and curriculum needs. Placement tests, the classification mechanisms most commonly used by community colleges, are under scrutiny; there is an ongoing examination of the validity of placement testing, and recent studies have revealed high school grade point averages were better predictors than placement tests in predicting success in college-level work (Belfield & Crosta, 2012; Hughes & Scott-Clayton, 2011; Scott-Clayton, 2011). As a result, policies involving multiple measures associated with placement are currently being considered in North Carolina. The move toward multiple measures is likely to challenge the entire placement referral system and how placement operates in community colleges.

Enrollment Status

Researchers Bettinger and Long (2005a) examined over 5,807 first-time traditional aged, 18-20 years old community college students who had taken the ACT, for a five-year period from 1998 to 2003. Full-time and part-time remedial students were compared to full-time and part-time non-remedial students in terms of credits completed, transfers, and credentials earned. The researchers found that full-time remedial students earned 5.4 fewer credits, were 15% more likely to have left college without a two-year degree and 3.6% more

likely to have left college without a four-year degree, compared to full-time, non-remedial students. Also, part-time remedial students were less likely to complete two-year or four-year credentials or transfer by spring of 2003, when compared to part-time non-remedial students. The researchers concluded students in remediation had poorer outcomes than students not in remediation. However, the results were based on simple comparisons that failed to control for selection bias and other student differences (Bettinger & Long, 2005a).

In a follow-up study designed to control for selection bias, researchers used differences among placement policies in Ohio and distance from high school to college to compare outcomes of students placed into developmental education to similar students not placed into developmental education due to placement policies. The key is that similar students were compared to each other and differences among students were controlled in the study. The researchers found positive outcomes for students placed in math remediation, as the students placed into math remediation were 15% more likely to transfer by spring 2003 than similar students not placed into remediation. In addition, students in math remediation completed ten more credit hours than similar students not placed into math remediation (Bettinger & Long, 2005b).

Depth and Breadth of Remediation Need

Bahr (2009) examined the fall 1995 cohort of first-time freshman at 107 California community colleges, with a final sample size of 68,884 community college students. The researcher examined whether the depth and breadth of remediation need moderated the extent to which students remediated successfully and obtained outcomes comparable to non-

remedial students. In addition, Bahr (2009) examined the impact of depth of remediation, or placing more than a level below college-level in one or more English or reading subject areas.

Bahr (2009) hypothesized four relationships to explore the extent to which breadth and depth of remediation need impacted credentials attained. With this in mind, the researcher first hypothesized, at each level of initial English deficiency, that students who remediated successfully in English (attaining college-level English skill) experienced academic outcomes (credential completion and credential transfer) comparable to those students who place into college-level English, after adjustment for attainment in math. Second, at each level of math deficiency, students who remediated successfully in math or attained college-level math skill experienced academic outcomes comparable to those of students who did not require remediation in math, after adjustment for attainment in English. Third, students who remediated successfully in both English and math experienced academic outcomes that are comparable to those of students who require remediation in math or English and attain college-level skill in both English and math. Finally, students who remediated successfully in English and math experienced similar outcomes that are comparable to students who do not require remediation in either area. The researcher found remediation effective for students who were successful remediators, even after taking into account the breadth and depth of deficiency. This finding was good news for the students who were faced with serious deficiencies in college reading and English subject areas.

Successful reading remedial students were slightly less likely to transfer without a credential. In addition, successful remedial math students were slightly less likely than college-level math students to complete a terminal vocational degree. By adding categories of associate degrees to the model, such as *college transfer* and *terminal vocational*, Bahr (2009) discovered successful remedial math students were less likely than college-level math students to attain a vocational credential. The researcher posited that students who persisted through the developmental math sequence were more determined to earn the four-year degree than students who had not persisted through the developmental math sequence.

In Florida, researchers Calcagno and Long (2008) examined community college students who had taken a college placement test and compared students placing just below the cutoff placement score to those right above the cutoff score. In order to establish consistency in the study, Florida community colleges that did not use the placement test scores to determine remediation were excluded. Out of 130,000 community college students, around 75%, or 98,146 students, were tracked for a total of six years, in three cohorts consisting of first-time, associate degree-seeking students from 1997-2000. Short-term outcomes examined were the completion of the subsequent college-level course in the remedial subject area and fall-to-fall persistence. Long-term outcomes examined were the completion of a credential, certificate, and or associate degree, transferring to a four-year university, and total credits earned in remedial and college-level courses.

The researchers found that students just below the cutoff score, or remedial students, were more likely to persist from fall to fall than students who were right above the cutoff

score, or non-remedial. However, developmental math students just below the cutoff score earned more total credits over six years than students right above the cutoff score. A closer examination of total credits revealed the differences in total credits among remedial and non-remedial students held for remedial credits, but did not hold for college-level credits or credits toward a degree. That is, there was no difference between the two groups in college credits earned. In addition, remediation in math had no effect on the completion of college-level math or the completion of subsequent English or math courses. Moreover, among the remedial group, researchers found a negative effect in associate degrees earned and transfer out rates. Adult students enrolled in remedial courses experienced slight positive effects.

Demographics-Racial Disparities

Bahr (2010b) examined racial disparities in postsecondary mathematics remediation. The researcher selected the four most populated racial groups among the fall 1995 first-time freshman cohort in 107 California community colleges. The four groups were White, Black, Asian, and Hispanic, with a sample size of 63,147 students. Results indicated that the odds of remediating successfully, defined as completing college-level math, were 3.1 times higher for White students than Black students, and 1.6 times higher than Hispanic students. The odds of remediating for Asian students were 1.2 times the odds for White students. Overall, less than 1 in 4 students (24.6%) successfully completed a college-level math course in six years. In addition, math skill deficiency was strongly related to successful math completion, and substantial differences across math categories were observed across the categories of race. Of students that entered at the highest level of remedial math, slightly half (50.3%) remediated

successfully compared to 1 in 15 (6.9%) students in the lowest level of math remediation. Also, 26% of White students entered at the highest level of remedial math, compared to 1 in 9 Black students (11.5%) and 1 in 7 Hispanic students (15%). By contrast, 1 in 6 White students (17.4%) entered at the lowest level of remedial math, compared to 2 in 5 Black students (40.8%) and 1 in 3 (31%) of Hispanics. Of these remedial students, 43.8% received an “A” in the course and remediated successfully, 11.8% received an “F” and failed, and 11.2% withdrew before completing the course. Notably, White students (15.3%) were more than twice as likely to receive an “A” as Black students (6.6%), and Black students (48.3%) were 1.4 times more likely than White students (35%) to earn an “F” or withdraw from the course. Asian students exhibited the highest rate of successful remediation, although more likely to be an ESL student, and Hispanics were more likely to persist than White students.

After investigating two mediators of the gap between Black and White students’ results, two variables emerged as important mediators: remedial math level and the grade received in first math. In addition, average differences in English deficiency contributed modestly while the other variables were of little consequence. Similar mediators were observed with the Hispanic/White gap, but enrollment patterns were also of significance for the Hispanic group, who had higher persistence rates despite their lower rates of successful remediation.

However, once college controls were added, Black students did not experience a disadvantage in terms of racial concentration of Black students. On the other hand, White, Asian, and Hispanic students experienced a noted disadvantage in institutions with high

concentrations of Black students, and Hispanic students experienced a disadvantage in institutions with higher enrollments of Hispanic students. Overall, students of all races benefitted from successful remediation, and there were racial disparities in successful remediation that reflected racial differences in remedial need. Bahr (2010b) concluded that such racial disparities reflect poorer resources in elementary school and the subsequent effect on students' college readiness.

Passing the Gatekeeper Course

One measure of program effectiveness in developmental education is the extent to which students remediate skills, progress to college-level coursework, and pass the gatekeeper course(s) in the areas of deficiency. Roksa, Jenkins, Jaggars, Zeidenberg, and Cho (2009) looked at progression through gatekeeper courses for students in Virginia community colleges. The study examined student characteristics, course-taking patterns, and other factors associated with higher probabilities that students in remediation will take and pass math and English gatekeeper courses. In order to be included in the study, students had to begin college work in summer or fall of 2004, and to be part of the cohort they must have been enrolled in fall 2004. The sample of 24,140 students was followed for four-years. The data analysis used a logistic regression to examine factors associated with higher probabilities of success in developmental courses. The descriptive results indicated that half of the cohort enrolled in at least one developmental course over the four-year period. In addition, over one third of students did not take any developmental courses in the recommended area.

College transfer students were more likely to enroll in developmental math compared to career and technical education students. Most college transfer students did not complete the sequence, both because they did not take the courses and, to a lesser extent, because they did not pass the courses in which they were enrolled. Among students referred to the highest level of developmental math, 60% did not enroll in developmental math, 16% enrolled but did not pass, and 25% enrolled and passed. At the highest end of developmental math, students did not complete the developmental sequence, largely because they did not enroll in the course. However, those students who skipped the developmental sequence and enrolled in the gatekeeper course did about as well as those students who enrolled in the developmental courses. Gatekeeper course enrollments were low, especially in math. A substantial number of students with high test scores did not take gatekeeper courses. In addition, 10% to 15% of high scorers did not take college-level English. Among students in the highest quartile of pre-algebra, only 31% took gatekeeper math, but students in highest quartiles of algebra and college algebra were more likely to take college-level math (Roksa et al., 2009).

Only 47% of students in the cohort entering in the summer or fall of 2004 completed gatekeeper English within four-years, and a mere 26% completed gatekeeper math. In general, rates in gatekeeper course enrollments were low especially for students in developmental education. The researchers concluded that, in community colleges in Virginia and elsewhere, a key challenge is motivating students to enroll in gatekeeper courses. However, the researchers' finding that students who took developmental courses did about as well in gatekeeper courses as those who did not take developmental courses might be

interpreted as an indication that developmental education enables success in gatekeeper courses for those who complete it. Yet the finding that students who were recommended to developmental education but skipped it nonetheless performed as well as those who did take developmental courses suggests that developmental instruction does not make a difference. Although the results do not imply that developmental instruction caused success in gatekeeper courses, it does suggest that some students do not need developmental coursework.

In a study of over 256,000 students in 57 ATD colleges, Bailey et al. (2010) examined developmental math and reading enrollment patterns in order to investigate if students completed the developmental sequence of course work. They found that 21% of students placed in developmental math and 33% of students placed in developmental reading had skipped the developmental sequence in reading and math. Of the developmental math students, 12% enrolled in and successfully completed the gatekeeper course, compared to 32% of the developmental reading students. Gatekeepers typically are the highest enrolled general education courses with low success rates.

Given that these students skipped the developmental course sequence and successfully completed the gateway course, Bailey et al. (2010) suggested that students have a better idea of their abilities and skills than placement tests indicate. The researchers concluded that, in some instances, skipping developmental coursework was a wise strategy. However, for some students, especially for those referred to math remediation who avoided the math course sequence, only 61% enrolled in another class and 42% earned no other

credits in the three years after the first term. The researchers mined an Achieving the Dream national dataset and found between 33% and 46% of students actually completed the required sequence of courses. However, as expected, developmental completion rates were negatively related to remediation level in the math and reading subject areas examined. In addition, about one third of students avoided developmental coursework by never enrolling in any developmental courses. Interestingly, the results also indicated that more students exited their developmental course or sequence because they did not enroll in the first or subsequent course, rather than because they failed or withdrew from a course in which they were enrolled. Overall, men, older students, part-time students, and students in vocational programs were less likely to progress through their full remedial sequence (Bailey et al., 2010).

Effectiveness of Math Remediation as Academic Attainment

Bahr (2008) examined the effectiveness of mathematics remediation among the fall 1995 cohort of 85,894 first-time freshman at 107 California community colleges. Bahr (2008) used a hierarchical, multinomial logistic regression technique to measure data at two levels, student and college-level data. The dependent variable of interest was academic attainment using five categories of outcomes.

Once college-level variables were added, the outcomes among both completer groups were very similar. Interestingly, remedial non-completers experienced the worst outcomes of the two non-completing groups. Bahr (2008) concluded math remediation works for those who remediate successfully. However, 3 out of 4 students did not remediate successfully and

4 out of 5 earned no credential and did not transfer. The conclusion was that few remedial students remediate successfully, but remediation works for those who do succeed in remediation. Bahr (2008) noted that future research should address the impact regarding the depth and breadth of remedial need.

In Attewell et al.'s (2006) study of remedial effects and academic attainment, remediation had a negative effect for four-year college entrants in degree completion; four-year college entrants in remediation took 2 to 3 months longer to complete the Bachelor's degree, and over half of students who took remedial courses in four-year colleges graduated within eight years of leaving high school. However, four-year college entrants experienced negative time to degree effects by taking three or more remedial courses. Despite the negative effects, 1 in 3 of those students who took many remedial courses still graduated within eight years. In addition, four-year college entrants had a clear negative effect in reading remediation and graduation rates.

In two-year colleges, taking two or more remedial math courses lowered likelihood of graduation by 3%. However, remediation in writing and reading improved two-year college entrants' chances of graduating. In sum, there was evidence that remediation successfully helped students in two-year colleges but had negative effects for students in four-year colleges. Most students took one or two remedial courses in the first year of college. Interestingly, students in public colleges were more likely to be enrolled in remediation than students in private colleges. The researchers thus concluded that public institutions placed more hurdles in the way of students' progress than did private colleges.

Overall, remedial students from two-year colleges had a 1 in 4 (25%) chance of graduating while students from four-year colleges had a 50% chance of graduating (Attewell et al., 2006).

Effectiveness of Remediation and Labor Market Outcomes

Martorell & McFarlin Jr. (2007) examined the effects of college remediation on academic and labor market outcomes in Texas. The researchers used several independent variables and compared outcomes to students in math remediation, reading remediation, and writing remediation who had taken the state placement test (TASP) in Texas. Entering freshman enrolling in the 1991-1992 and 1999-2000 school years, and students in both two-year and four-year colleges, were followed for up to six years after first enrolling. For the main analysis, the sample was composed of 255,878 two-year college students and 197,502 four-year college students. In order to be part of the sample, students needed valid scores on all three subject areas of reading, English, and math, and were excluded if they failed the writing section due to methodological requirements of the regression discontinuity technique. Writing scores did not meet the criteria of a continuous variable and were excluded for the main analysis, but students who failed the writing section were added to the analyses that examined student outcomes in the separate areas of math and reading.

The variables examined were: pass rates in the TASP; attempted academic credits in first year; total attempted academic credits; transferring up to a four -year university or down to a community college; completing at least one year, two-years, or three years; graduating within 4 to 6 years; and earnings in years five, six, and seven. Other variables included were:

White, economically disadvantaged, economically disadvantaged missing, ages 21 years and up in first semester, distance from high school being less than 55 miles, distance from high school being more than 50 miles, distance from high school missing, receive in-district tuition, enrolled in 1995 or earlier, started college during the fall semester, rescaled max (reading, math) score, and rescaled min (reading, math) score (Martorell & McFarlin Jr., 2007).

Using a regression discontinuity approach, the descriptive statistics revealed higher remediation rates for two-year college students and a more common occurrence of remediation for math, and showed that remedial students were older, economically disadvantaged, more likely to be non-White, and have much lower test scores than non-remedial students. At two-year colleges, 23% of students in remediation graduated from college compared to 38% among non-remediated students. In addition, 86% of two-year versus 93% of four-year remedial students were retained the following semester after initial enrollment. Remediation reduced first-year credits to 2.4 credits for two-year students and 1.5 credits for four-year students. There was no evidence of transferring up or transferring down. The estimates indicated that starting at a two-year college lowers the probability of completing a year in college by 6%. The results also supported the hypothesis that remediation increases the time to degree completion. However, there was little or no effect on earnings (Martorell & McFarlin Jr., 2007).

Higher remedial percentages of students by institution led to a 12% point reduction in probability of completing at least one year in college. Math remediation had smaller negative

effect on attempted credit hours and completing at least one year in college. The researchers concluded that remedial offerings in Texas were ineffective, cut scores were not at the right cut-offs, and marginal students gained little benefit from remediation. The researchers also determined that future studies should examine why the effects of remediation differ across high versus low remedial institutions. In general, remediation made no difference in terms of academic and labor market outcomes in Texas (Martorell & McFarlin Jr., 2007).

Propensity Score Matching and Educational Research

Using a propensity matching methodology that controls for background factors, Attewell et al. (2006) examined and compared outcomes for two groups of similar students. The background factors included demographics and student-level academic factors such as high school grade point average. The two groups were two-year and four-year college entrants from the National Educational Longitudinal 1988 high school cohort study population, with a sample size of 6,879 students. The researchers found that the probability of remedial placement was 11% higher for two-year versus four-year students with similar backgrounds, and that there was an 11% higher probability that remedial students were African-American (Black) than White. Moreover, SES was not a significant indicator of remedial placement (Attewell et al., 2006).

Bostian (2008) hypothesized that some students transfer from a community college to a four-year university ostensibly in order to avoid remedial coursework, which in practice carries no college credit and can delay completion (Bettinger & Long, 2005b; Martorell & McFarlin Jr., 2007). The purpose of Bostian's (2008) research study was to determine the

extent of "remedial avoidance" for two, first-time cohorts of 910 community college students. He selected two groups of students for analysis, the avoider group and the non-avoider group, and matched them based on the conditional probability that they were members in one group versus another (i.e., propensity score matching). Another objective of the study was to determine, after matching, whether there were differences in the two groups of students' outcomes of transferring to a local four-year university (Bostian, 2008).

Bostian's (2008) study of remedial avoidance found that 14% (n=127) of the total 910 students in the study population avoided remediation and transferred to the four-year university. However, 178 of the original cohort did not avoid remediation, transferred to the 4-year institution, and composed the non-avoider group. The student background and academic characteristics differed for the two groups among several independent variables such as placement test scores, age, and grade point average. However, once the differences in students were controlled using propensity score matching, there were no detectable differences among the 89 matched students in the non-avoider and avoider groups. The groups fared about the same in terms of degree completion and grade point averages. The researcher concluded that students do avoid remediation, but those that have the tenacity and intention to transfer to the university achieved comparable results to the non-avoider group. While the sample size was small (Bartlett et al., 2001) and may be inadequate, the study filled an important gap in the literature by following students after transfer to an institution. However, we know little about the extent to which outcomes vary in significant ways among

different student groups based on student background, academic characteristics, goals, and intentions.

Summary

Most of the research is converging towards the finding that remediation proficiency levels matter in terms of remediation and college success (Roksa et al., 2009). Bahr (2009) initially found that successful remediators, even those students faced with serious deficiencies, can succeed at comparable levels to non-remedial students; however, this finding contrasts with other studies, including Bahr's (2010) own subsequent study examining racial disparities. For example, Roksa et al. (2009) found that students in the need of the most remediation do not succeed at comparable rates to those students with fewer remedial needs, especially in math. Bailey (2009) found, as expected, that developmental completion rates were negatively related to remediation levels in the math and reading subject areas examined.

Bailey (2009) and Roksa et al. (2009) also found that one third of students placed into remediation did not attempt developmental courses. Since Bahr (2009) excluded non-attempters or "avoiders" of developmental math from all studies examined, a full one third of the students who may have needed remediation would not be in the study group. Students who take and complete a developmental sequence may be especially motivated or otherwise different in ways that would explain the fact that they do as well in gatekeeper math courses as students not placing into developmental coursework (Roksa et al., 2009).

Typically, developmental students are determined by a placement test score, and the cutoff point for developmental versus college-level placement is determined by departments and varies from institution to institution or by statewide policies (Bailey et al., 2010; Oudenhoven, 2002; Perin, 2005). Bahr (2008) acknowledged that one of the limitations of his study was the assumption that students were perfectly placed into their developmental or college-level coursework. In other words, placement test scores were not used to determine remediation need levels, but a student's first enrollment in remedial coursework was indicative of their remedial status. The rationale for using the first course enrolled rather than some other measure was that placement policies and practices varied across the community colleges examined. Bettinger and Long (2005a) have demonstrated that differences in placement policies impacted the likelihood of placement into remediation even after controlling for differences in student backgrounds. Since the methodology the researcher used controlled for student differences but lacked proficiency scores on placement tests, the results need to be interpreted with an understanding of the students not included in the study.

However, the research revealed placement into developmental education as an indicator of remedial need also has its limitations. Two studies found that students placed into math remediation skipped developmental courses and successfully completed the math gatekeeper course (Bailey et al., 2010; Roksa et al., 2009). The successful math students who skipped developmental math would not be in Bahr's analysis because the students did not initially enroll in developmental math.

Unfortunately, the ability to determine the nature of skipping courses is limited in the ATD national data set and in the Virginia community college study. If a student places in developmental education, he or she is treated as a student who must move through a sequence of courses. To what extent skipping is actual behavior or a result of varying program of study requirements cannot be determined. The phenomenon of students skipping developmental education courses has been used to argue that a good number of students did not need developmental coursework (Bailey et al., 2010). However, students who actually skipped courses, as well as those who were not required to take some of them, cannot be teased out of the data. Completion of the developmental sequences often does not reflect how developmental education is executed at community colleges.

Moreover, the classification mechanisms most commonly used by community colleges, placement tests, are under scrutiny. There is an ongoing examination of the validity of placement testing, and recent studies have revealed high school grade point averages were better than placement tests in predicting success in college-level work (Belfield & Crosta, 2012; Hughes & Scott-Clayton, 2011; Scott-Clayton, 2011). As a result, policies involving multiple measures associated with placement are currently being considered in North Carolina. The move toward multiple measures is likely to challenge the entire placement referral system and how placement operates in community colleges.

Regression discontinuity studies have demonstrated that marginal students would probably do just as well as if they did not take remedial courses. Using a regression discontinuity approach Martorell and McFarlin Jr. (2007) concluded remedial offerings in

Texas are ineffective, cut scores are not at the right cut-offs, and marginal students gain little benefit from remediation. In general, remediation made no difference in terms of academic and labor market outcomes in Texas. In Florida, researchers Calcagno and Long (2008) examined community college students who had taken a college placement test. They compared students placing just below the cutoff placement score to those who were right above the cutoff score, another utilization of regression discontinuity designs. The researchers found students just below the cutoff score were more likely to persist from fall to fall than students who were right above the cutoff score, remedial and non-remedial students respectively. There was no difference between the two groups in college credits earned. In addition, remediation in math had no effect on the completion of college-level math or the completion of subsequent English or math courses. Moreover, among the remedial group, researchers found a negative effect in associate degrees earned and transfer out rates. Adult students enrolled in remedial courses experienced slight positive effects (Calcagno & Long, 2008).

After a review of studies which examined students at the margins of remediation, Bailey (2009) recommended community colleges stop classifying students as developmental and non-developmental and instead strive to appropriately assess and diagnose students' remedial and curriculum needs. However, regression discontinuity studies yielded inconclusive and contradictory findings. In a follow-up study designed to control for selection bias, researchers used differences among placement policies in Ohio and distance from high school to college to compare outcomes of students placed into developmental

education to similar students not placed into developmental education due to placement policies. The key is that similar students were compared to each other and differences among students were controlled in the study. The researchers found positive outcomes for students placed in math remediation, who were 15% more likely to transfer by Spring 2003 than similar students not placed into remediation because of placement policies. In addition, students in math remediation completed ten more credit hours than similar students not placed into math remediation (Bettinger & Long, 2005a).

Taken together, an inverse relationship to remedial need and successful completion of the course sequence or gatekeeper course, the skipping behavior observed in Virginia and in the ATD colleges, limited and inconclusive results from the regression discontinuity studies and questionable placement test validity, researchers have shaped and influenced developmental education programming in community colleges across the country. The shift to a focus on student success in community colleges is paramount to understanding the educational landscape in developmental programming. Specifically, researchers argue far too many students enter the developmental education “gauntlet” (Bostian, 2008) without ever progressing to college-level work or completing a program of study. Given that some research indicates those students on the margin gain little or no benefit from remedial instruction and that developmental education yields inconclusive benefits at best and makes no difference in the worst cases, developmental education is under major revision to accelerate and shorten the length of time students spend in developmental education programming.

Similar to his previous study, Bahr (2010b) found that when examining developmental students as a group, differences in terms of the extent to which students need remediation affected the results of successful remediation for those groups of students. However, Bahr (2010b) did not compare the results to non-remedial students as in previous studies. The focus of the study was specifically on remedial students and their progress to college-level gatekeeper courses in math. His examination of racial disparities is important, but the findings were not that different from other research (Roksa et al., 2009).

The research did yield two findings that are important to the discussion of remedial effectiveness: Hispanics were more likely to persist than White students, despite having relatively lower rates of success in remediation. In addition, the study found that racial concentrations of remedial students affected students differently. Again, the study was limited, since enrollment in remedial math is the group of study rather than remedial need as determined by placement tests or other assessments. Institutional variation is not well examined in the literature, and the researcher's discovery of racial concentration impacting the likelihood of effective remediation is important to the discussion but not well-addressed in the literature (Bahr, 2010b).

Once college controls were added, Black students did not experience a disadvantage in racial concentration of Black students. In contrast, White, Asian, and Hispanic students did experience a disadvantage in institutions with high concentrations of Black students. Hispanic students experienced a disadvantage in institutions with higher enrollments of Hispanic students. Overall, students of all races benefitted from successful remediation and

there were racial disparities in successful remediation that reflected racial differences in remedial need. Bahr (2010b) thus concluded that racial disparities reflect poorer resources in elementary school and its effect on student college readiness. Martorell and McFarlin Jr. (2007) found that higher remedial percentages of students by institution led to a 12% point reduction in probability of completing at least one year in college. The researchers concluded future studies should examine why the effects of remediation differ across high versus low remedial institutions. It was to this body of work regarding institutional contexts and the progress of developmental math students that this study made a significant contribution.

While the developmental education literature is growing and seeks to examine the effectiveness of developmental education, an equally important question to address should regard educational research methodology as a topic of significance. Indeed, Levin and Calcagno (2008) critiqued evaluation studies in the community college and argued that ideal studies utilize an experimental design. The second choice among these researchers would be quasi-experimental designs. The researchers argue that community college research offices are under-staffed and evaluation studies are not appropriate in providing effective information. Institutional researchers need to provide information that institutions can use to improve services to developmental students but spend most of their time generating reports for state and federal agencies. (Levin & Calcagno, 2008). However, in developmental education, given the extent of costs in time and resources taxed on students, faculty, and the institutions, community colleges need to evaluate developmental education using the best, most feasible and practical approach. While research has expanded at the national level,

quantitative studies need to examine students in developmental education at the institutional levels.

It is common to encounter student unit record data in the community college and to analyze the impact of educational interventions using two groups of students, those exposed to the intervention and those not exposed. Yet results are limited in that the students are not typically randomly selected into experimental and control groups. Non-random selection implies that the two groups of students may be very different on key factors that affect the results of analyses through self-selection bias and other differences. In addition, Bettinger & Long (2005b) argued that comparing remedial students with non-remedial students introduced the potential for bias:

> Although a simple comparison suggests that remedial placement has a negative impact on students, it masks the fact that students are not randomly placed in remediation. Better-prepared students are less likely to be placed in remediation and they also do better in college. Thus, simply comparing remedial students with non-remedial students is an unsatisfactory way to establish the true effects of remediation. (Bettinger & Long, 2005b, p. 23)

Propensity matching is a technique designed to simulate an experimental design, controlling for selection bias and creating almost equivalent experimental and comparison groups on key indicators. Comparisons of student outcomes using propensity matching has been used to yield less biased results than are derived using simple comparisons (Rojewski et al., 2010). It is critical that researchers apply methodologies that control for selection bias. Resources are

scarce and monies need to be allocated to educational programs that are making a difference in students' success and long-term outcomes.

Multilevel propensity matching, that is, matching across multiple institutions in order to account for variation in community colleges, was relevant to the study of developmental education. Multilevel studies can be used to exploit the contextual variation in developmental education across multiple institutions and to determine what institutional factors impact developmental programs. Researchers have pointed to institutional variables that needed further examination such as the percent of racial concentration and percent of remedial needs as institutional constructs. In addition, institutional disparities in income level as measured by Pell grant recipients were important constructs to examine.

To the researcher's knowledge, no researchers have used propensity-matching methods to study developmental math across multiple institutions. The researcher's study applied propensity score matching to contribute to developmental education studies. Multilevel propensity matching was used in order to illuminate the methodology's use in the evaluative literature and to connect the study to one of the primary purposes for evaluation in educational settings: to strive for continuous improvement of educational programs. As Grubb (2001) pointed out in his review of the developmental education literature, "colleges must evaluate remedial education using sophisticated techniques such as treatment and control groups; program evaluation and improvement is central to improving remedial student's outcomes" (p. 10). Moreover, researchers have noted:

As policymakers intensify their calls for increasing graduation rates, it becomes important to consider the relative effect sizes of various predictors, to determine which factors have direct effects and which are mediating variables, and to determine whether mechanisms vary across types of college—so that policies can be well targeted and potential interventions be appropriately prioritized. (Attewell et al., 2011, p. 537)

CHAPTER III: METHODS

Overview

The methods used for this study is known as the propensity score matching technique. The study population consisted of the 2007-2008 new student cohorts at community colleges that are part of the Completion by Design Student Success Initiative in North Carolina. All of the students completed college placement tests and were referred to developmental math coursework prior to beginning their studies. In the study population, students were assessed in three different areas; English, reading, and math, and students often placed in more than one area of developmental course requirements.

The selection of independent variables in the study was aligned with previous studies in developmental education research. The initial use of logistic regression permitted the researcher to identify covariates by using multiple quantitative independent variables to predict the probability of group membership, the dependent variable. In general, researchers agree appropriate independent variables are those variables that explain differences in group membership which are constant over time (Rosenbaum & Rubin, 1983, 1984). In an educational environment, constants include student demographics and academic variables (Jenkins, Zeidenberg, & Kienzl, 2009; Rojewski et al., 2010). Student demographics, academic variables, and institutional variables were the independent variables considered for initial inclusion in the propensity score analyses.

Pre-screening was used to ensure multi-collinearity of student and college-level variables was minimized. Multi-campus studies are used to examine the institutional context

associated with developmental education programming. Gelman & Hill (2007) recommend that propensity studies utilizing multi-institutional data examine a minimum of five institutions in order to justify the use of multi-campus data.

Methods

Study Population

The study population consisted of the fall 2007-08 new student cohorts at 7 community colleges. Student progress and outcomes were examined for 5 years. The student record dataset included demographics, course completion data, and transfer information available through the National Student Clearinghouse. Through participation in the Completion by Design Initiative, a project of the Tides Center funded by the Bill and Melinda Gates Foundation, the colleges were provided with a raw data extract. The extract included individual college submitted data provided to the Completion by Design partners and a series of derived measures provided by the Columbia University Teachers College Community College Research Center (CCRC), a data partner in the Completion by Design Initiative. In addition, the raw data extract was analyzed in order to create additional measures for this study.

Because Integrated Postsecondary Integrated Data Systems (IPEDS) cohorts represent a select group of degree seeking students (Leinbach & Jenkins, 2008), the study did not restrict the analysis to IPEDS classifications (first-time, full-time, certificate, or degree declared), but instead examined all full-time and part-time students that began work at the colleges in the academic year fall 2007 through summer 2008. In order to be part of the study

group, a student must have first enrolled in a developmental math class and subsequently enrolled in a college-level math course during the five-year time frame from 2007 to 2012 (Bahr, 2008). Most (92%) of the developmental math students completed college placement tests and were referred to developmental math coursework prior to beginning their studies. Since some test scores were missing from the datasets, students were also placed in the study group based on first enrollment in developmental math. The rationale was that, despite missing test scores, the population of interest was the group that first enrolled in developmental math, and initial placement levels were determined based on the first enrolled course (Bahr, 2008). In the study population, students were assessed in three different areas, English, reading, and math and were often placed in more than one area of developmental course requirements.

Students choose to take the developmental coursework; there is no mandatory policy in place, but students are blocked from classes in the developmental need areas until meeting the developmental requirement. General Education courses also have requirements that will block students from enrolling in them. As a result, mandatory placement is somewhat enforced through course prerequisites and specifically in the general education courses.

Dataset Construction

Dataset construction was done using SAS 9.2. Each college uploaded a set of files to the researcher that were provided by JBL Associates during the spring of 2013. The data files were created as part of the colleges' participation in the Completion by Design data submission process. In order to use the data file, the encrypted student identifier was used to

merge the data files, creating multiple record files for use by the researcher for each college. For this analysis, the researcher used the student-derived and course enrollment files. The student-derived file contains demographics and other student indicators provided by colleges and subsequently derived by CCRC. The course enrollment files were provided by the colleges and included course enrollments and grades received in the courses for each student, for each course by the terms enrolled. The dataset construction combined all the data during the final procedure and yielded an unduplicated count of 26,871 students for the analysis.

Student-Level Derived and Recoded Variables

Variables not provided as part of the data files were derived by the researcher using a PROC summary feature in SAS 9.2. Summary course enrollments were derived by dummy coding an occurrence of an event, such as enrollment in the course, as indicated by the student's record. Further coding used the student's final grade to determine if the student persisted in the course during the semester and whether a student earned a grade of C or better during the semester. Final grades were coded according to a grading scheme outlined by the data submission requirements and subsequently translated by the researcher. A code of *4, 3, 2, 1, 0, and 2* was translated to a grade of *A, B, C, D, F,* and *I,* respectively, and a code of *-3* indicated the student withdrew from the course. Audits were coded as -4 and were retained in the analysis. Missing grades, grades coded *-5* or *-1*, were not used and were dropped from the analysis during the dataset creation phase (Bahr, 2008). The variables in the dataset *numcreditattempt* and *numcreditcomplete* were validated using the grading schemes and subsequently used to derive enrollment and course completion variables.

Outcomes in developmental courses were derived using a three-digit prefix and course numbering system of less than 100. College-level courses were derived using a three-digit prefix and a course numbering system of 100 or greater (NCCCS, 2013a). The colleges use a common course number system which facilitated the selection of courses used in the study. Aggregated course enrollments were derived using all the courses in the datasets meeting the criteria of a specific developmental or college-level course in a certain subject area, or also categorized in a category of courses such as total college or developmental level credits attempted, completed, and earning a recorded grade of C or better. Table 1 shows the course enrollments analyzed and used in this study.

Table 1: Coding Scheme for Outcomes: Developmental Credits, Developmental Math, Reading and English Credits, College Credits, College Math and English Credits

Outcome Variable	Variable Type	Coding
Developmental Credits		
Attempted	Continuous	
Completed	Continuous	
A - C Grades	Continuous	
Developmental Math Credits		
Attempted	Continuous	
Completed	Continuous	
A - C Grades	Continuous	
MAT 050, A - C Grades	Continuous	
MAT 060, A - C Grades	Continuous	
MAT 070, A - C Grades	Continuous	
MAT 080, A - C Grades	Continuous	
Developmental English Credits		
Attempted	Continuous	
Completed	Continuous	
A - C Grades	Continuous	
Developmental Reading Credits		
Attempted	Continuous	
Completed	Continuous	
A - C Grades	Continuous	

Table 1 (continued)

Outcome Variable	Variable Type	Coding
College Credits		
Attempted	Continuous	
Completed	Continuous	
A - C Grades	Continuous	
College Math Credits		
Attempted	Continuous	
Completed	Continuous	
A - C Grades	Continuous	
College English Credits		
Attempted	Continuous	
Completed	Continuous	
A - C Grades	Continuous	

Demographic variables were initially provided by the colleges and coded or categorized by the researcher for this study. Female was derived using the variable coded as gender in the original data and dummy coding female as 1 and male as 0. Race and ethnicity variables were translated and dummy coded, collapsing Other race (Multiple, Unknown, and American Indian (n=13)) as the reference group and dummy coding White, Black, Hispanic, Asian American, and International students at 1 and 0, respectively. A continuous age was

derived using the first term of enrollment and the birth year to compute the age at enrollment. Students preparing to transfer to a four-year institution were coded as 6 in the original data and dummy coded to reflect the intent to transfer or not transfer. Pell grant recipients in the first term were coded as 1 in the original data and dummy coded to reflect Pell status in the first term of enrollment. A student's enrollment status in the first term was dummy-coded to reflect full-time status of attempting 12 or more credit hours in the first term. A student was coded as late entry if the student's computed age was 24 years of age or more (Bean & Metzner, 1985). Table 2 shows the coding scheme used for demographic covariates.

Table 2: Coding Scheme for Demographic Covariates

Variable Name	Variable Type	Coding
Female	Dummy	1, Yes 0, Male
Other Race (Unknown, American Indian, Multiple)	Dummy	
White	Dummy	1, Yes 0, Other Race
Black	Dummy	1, Yes 0, Other Race
Hispanic	Dummy	1, Yes 0, Other Race
International	Dummy	1, Yes 0, Other Race
Asian American	Dummy	1, Yes 0, Other Race
Age	Continuous	
Transfer Program	Dummy	1, Yes 0, Non-Transfer
Pell Recipient in First Term	Dummy	1, Yes, 0 No
Enrolled Full Time* in First Term	Dummy	1, Yes, 0 Part-time
Late Entry $\geq$ 24 years old	Dummy	1, Yes, 0 No

**Full Time = 12 or more credit hours*

Since the data were longitudinal, several variables were derived and used as indicators of retention and persistence. Timing variables of first term enrolled in developmental math and first term enrolled in college-level math were constructed to address

that the student's first enrolled math course was at a developmental level (Bahr, 2008). Using the array feature in SAS, terms were arrayed in order and, subsequently, first terms of enrollment in the two course areas were translated to years 1-6 or 2007-2012. The timing variables were used to ensure the developmental math enrollment preceded the subsequent college-level math enrollment. The first developmental math course in which the student enrolled was translated into a developmental level with the lowest number: MAT-050 coded as 4; MAT-060 coded as 3; MAT-070 coded as 2; and MAT-080 coded as 1. Two colleges did not offer MAT-050 and the first developmental math level was re-coded. Levels 3 and 4 were combined to reflect three levels of developmental math, 1-3. Finally, first-term GPA and total terms retained were derived using the PROC summary feature in SAS. Table 3 outlines the coding scheme for first-year and first-term covariates for the academic years 2007-2012.

Table 3: Coding Scheme for First-Year and First-Term Covariates, 2007-2012

Variable Name	Variable Type	Variable Range
First Year Dev. Math	Ordinal	1-6
First Year College Math	Ordinal	1-6
First Dev. Math Level	Ordinal	1-3
First Term GPA	Continuous	
Terms Retained	Continuous	

Other variables indicating persistence and enrollment behavior were dummy-coded to indicate whether the student returned or enrolled in courses and were used as proxies of behavior. A returned second major term was derived to indicate persistence to the second term. If the student first enrolled in fall 2007, the next major term was the spring of 2008. Students first enrolled in the spring or summer of 2008, were assessed in the fall, the next expected major term. In the study population, students are assessed in three different areas: English, reading, and math. Students are often placed in more than one area of developmental course requirements. Enrollments in reading and/or English developmental courses were dummy-coded at 0 and 1. In addition, enrollment in a student success course was dummy-coded in the dataset. Notably, student success courses have yielded positive outcomes, especially for developmental students (Cho, Karp & Mechur, 2013). Table 4 displays the coding scheme used for academic progress indicators.

Table 4: Coding Scheme for Academic Progress Indicators

Variable Name	Variable Type	Coding
Returned 2nd major term	Dummy	1, Yes, 0 No
Developmental English Attempted	Dummy	1, Yes, 0 No
Developmental Reading Attempted	Dummy	1, Yes, 0 No
Success Course Attempted	Dummy	1, Yes, 0 No

A series of completion outcomes as of Year 5 were provided in the original data and subsequently collapsed to create five derived outcomes of success pertaining to this community college study. The measures used are also aligned with accountability measures for community colleges in the state (NCCCS, 2013a). Using an original set of 13 outcomes, the researcher derived the following mutually exclusive categories: *no outcome*, *completion*, *transfer*, or *persistence*. No outcome or certificates, or not being enrolled in Year 5, was coded as 1; associate degrees were coded as 2; transferring to a two-year institution was coded as 3; transferring to a four-year institution was coded as 4; and being still enrolled with 30 or more credits was coded as 5. Table 5 indicates the coding scheme for five-year completion outcome variables used in this study.

Table 5: Coding Scheme for Five-Year Completion Outcome Variables

Outcome Variable	Variable Type	Coding
Complete, Transfer, or Still Enrolled with 30 plus credits	Categorical	1 No Degree Outcome 2 Associate 3 Transfer to Two-year 4 Transfer to Four-Year 5 Still enrolled with 30 plus credits

Institutional-Level Variables

A multi-campus study was used to examine the institutional context associated with developmental education programming. Gelman & Hill (2007) recommend propensity studies utilizing multi-institutional data examine a minimum of five institutions in order to justify the use of multi-campus data; the current study used seven institutions in the same

state. The researcher derived one set of institutional variables using the official Integrated Post-secondary Educational Data System (IPEDS) enrollment data for each institution during the fall 2007 term (NCES, 2007). Institutional variables were computed as the percentage of fall enrollment divided by the total number of students that fell into the category measured. Another set of institutional variables used the critical success factors from the North Carolina Community College System for the year 2007-2008 (NCCCS, 2009). The developmental metrics measured the percent of students enrolled in the developmental math, English, and reading classes as a function of full-time equivalent for that year. Results indicated that one institution did not offer developmental reading and, due to potential problems with missing values, developmental reading indicators were therefore not used in the current study. An institutional membership variable was used to examine institutional variance. Table 6 indicates the institutional variables used in the study.

Table 6: Coding Scheme for Institutional Indicators

Variable Name	Variable Type	Coding
Institutional ID	Dummy	A,B,C,D,E,F,G
% of minority students (non-White)	Continuous	0-100
% of students Pell-awarded	Continuous	0-100
% of White students	Continuous	0-100
% of students in Developmental Math	Continuous	0-100
% of students in Developmental English	Continuous	0-100

Dependent Variable

Because propensity score matching (PSM) examines grouped data, the dependent variable is categorical instead of quantitative. Two groups were used in the propensity model, and logistic regression was used to determine the propensity to complete college-level math with a grade of C or better. The two groups analyzed in the study were students who completed college-level math with a grade of C or better and students attempting and not completing college-level math with a grade of C or better, referred to as a study and comparison group, respectively. A student's initial membership into the group was based on first enrollment of developmental math and a subsequent enrollment of a college-level math course (Bahr, 2008). Table 7 shows the dependent variables in the study.

Table 7: Dependent Variable in Logistic Regression

Variable Name	Variable Type	Coding
College-level Math Completer	Categorical	1, Yes 0, No

Research Design

Students have different characteristics, and institutions vary in terms of characteristics that influence a student's success. Students' progression through a developmental math intervention and subsequent enrollment in a college-level math course is generally measured at the institutional level. Since this study involved multi-campus data, the multilevel

composition of students nested within schools provided the rationale for a multilevel propensity score model (Kelcey, 2011). Multilevel research, in general, is an area that needs investigation (Huneycutt, 2010), and the community college context is an area of research that particularly lacks the multilevel analyses provided in this study. To the researcher's knowledge, no multilevel propensity models are available that model retention of developmental math students.

There are two types of multilevel models available to researchers: random intercept (RI) and random intercept and slope (RIS). RI views the contribution of student-level characteristics on the probability of study assignment as fixed across schools. In other words, school-level variation comes from school membership. RIS views the study assignment as a combination of the interaction between school characteristics and student characteristics, in addition to school membership alone. According to Kelcey (2011), variation between schools in student-level slopes must be considered in propensity score estimation.

In this study, student-level characteristics were considered Level 1 variables. Two propensity models aggregated the students in all of the institutions and measured the contribution of student-level covariates in two phases. Phase 1 was a propensity model that matched students based on student-level background characteristics at the beginning of the term, while Phase II retained the initial student-level background characteristics and added the academic retention and persistence indicators to the model. The second phase added variables that demonstrated the importance of considering more than just student background

characteristics in a propensity model estimation or in a retention model that measured the propensity to complete a college-level math course.

Level 2 variables were the institutional-level variables for each college represented in the study. There were two different models utilizing institutional variables, and the models were analyzed separately. To account for institutional membership, terms involving school level covariates (Rosenbaum & Rubin, 1985) are the same for each individual in the school; in other words, the institutional membership variables are constant across individuals within the same school. The original multilevel structure is preserved and the model controls all observed and unobserved school level covariates. Finally, given the multilevel structure of the study, an alternate model examined student-level variables across individuals of the same school and constant institutional-level predictors.

Assumptions

A recent comprehensive review of propensity studies in the social science literature concluded that well-designed and rigorous propensity studies exhibited the following characteristics: adequate sample sizes based on the number of covariates selected; a methodological discussion of covariates that reflect theoretically-bound characteristics and include more than age, race, and gender for controls; and the examination of results before and after propensity matching, including effect sizes due to large sample size influences (Thommes & Kim, 2011). Hence, the review suggested the same rigorous methods be applied as would be needed in any good and valid research study. In this study, the aggregate data were used to implement propensity score matching with the full sample. However, one

objective of this study was to examine variance across institutions. Notably, Rosenbaum & Rubin (1983) pointed out that propensity score estimation across institutions requires adequate sample sizes of at least 300 cases in all of the institutions used in the study. The smaller institutions with less than 300 cases were challenging for this type of multilevel analysis. A match must be found within each institution and across multiple institutions or the case was dropped.

Researchers acknowledge the limitations of comparing student performance, progression, or retention in a non-scientific study where participants are randomly assigned or equivalent in terms of motivation, intentions, background, or skill level (Titus, 2007). While random selection is the "gold standard" (St. Pierre, 2006), random selection is often impractical, unethical, or resisted in educational settings. Propensity score matching is used to address the counterfactual; that is, what would have happened to a similar group not receiving the treatment through choice or self-selection (Titus, 2007). Indeed, failure to control for selection bias is a common concern in the study of developmental education (Bettinger & Long, 2005b). Propensity matching is a technique designed to simulate an experimental design, controlling for selection bias and creating almost equivalent experimental and control groups on key indicators. Comparisons of student outcomes using propensity matching has been used to yield less biased results than are derived using simple comparisons (Rojewski et al., 2010).

Since propensity score matching (PSM) is a multivariate statistical technique, there are multiple steps involved in the analysis. There are six essential steps to PSM: data pre-

screening, covariate identification, propensity score estimation, matching of propensity scores, determination of matching success, and presentation of results. Texts such as Mertler and Vannatta (2010) or Green and Salkind (2011) offer step-by-step SPSS procedures for many of the analytical steps, while texts by Agresti and Finlay (1997) and Hair, Black, Babin, Anderson, and Tatham (2006) have provided theoretical background on many of the statistical techniques. General outlines of the components of the essential steps are presented in the following sections.

Pre-screening

During dataset construction, missing grades coded -5 were removed from the study datasets. During pre-screening, the missing grades were reanalyzed in the raw data extracts to determine the extent and impact of missing grades in the courses that the study concerned. Specifically, courses in English, reading, and math subject areas were examined among each individual institution. The analysis measured the percent of missing grades in respect to the study population of developmental math students. The results indicated the missing data eliminated 20% of the grades in the English, reading, and math subject areas in one institution. Both developmental and curriculum course credits were affected. There was no method available to translate the grades to actual credits and, due to the extent of the missing observations which impacted 634 students, the institution was removed from the analysis. Missing grades also resulted in the removal of 102 grades or 48 students in another institution. The missing grades did not affect curriculum courses in the subject areas. Overall, 224 students were retained in the analysis and a subsequent analysis indicated the impact on

the study population was minimal, given that the study group was reduced considerably by retaining only the students attempting college-level math. As a result, the researcher determined that the second institution would remain in the dataset.

Using the STATA software, correlations were computed to measure the pairwise correlation between pairs of variables while controlling for the effect of the other variables. First, Level 1 variables were examined for possible collinearity problems. Initial results indicated that age and late entry were correlated at .879, $p < .01$. Since the correlation coefficient did not reach the threshold of .900 (Hair et al., 2006), both variables were retained for possible inclusion in the analyses. Level 2 variables were also examined for potential issues of multi-collinearity; the results of the analyses are displayed in Table 8. Percent minority and percent White were highly correlated with a coefficient of -.973, $p < .01$. Percent White was also negatively correlated with percent of Pell-awarded students with a coefficient of -.897, $p < .01$. Due to an interest in the percent of Pell-awarded students in this study, percent White was removed from the analysis and percent minority was retained. Percent of students in developmental English and percent of minority students were positively correlated with a coefficient of .731, $p < .01$. Both variables were retained in the analyses.

Table 8: Correlations of Institutional Covariates Used in Propensity Model

	White	Minority	Pell Awarded	Dev. English	Dev. Math
White	1.000	-0.973*	-0.897*	-0.745*	-0.571*
Minority	-0.973*	1.000	0.851*	0.731*	0.708*
Pell-awarded	-0.897*	0.851*	1.000	0.435*	0.482*
Dev. English	-0.745*	0.731*	0.435*	1.000	0.498*
Dev. Math	-0.571*	0.708*	0.482*	0.498*	1.000

*(Note: All numbers represent percentages; Dev. = Developmental; *p< .001.)*

In order to address multi-collinearity, the researcher chose to exclude the variable of least importance to the study and retain the most important variable. Once a decision was made to exclude a variable based on collinearity problems, a sensitivity analysis was performed using probit regression and the PScore function in STATA to validate issues which, due to collinearity, were captured and lacked balance when generating the PScore (Leuven & Sianesi, 2003). One advantage of using the STATA PScore function is that the researcher was provided with analyses of propensity score estimations within blocks (Leuven & Sianesi, 2003). In the event of model imbalance due to collinearity or categorical imbalance, the researcher was warned that the model was unbalanced and a different specification of the propensity score was needed. The PScore function ensured the researcher that the model was balanced before proceeding.

Selection of Pre-treatment Covariates

The researcher began the study by identifying, coding, and deriving independent variables that can bias comparison studies. The initial use of logistic or regression permitted the researcher to identify covariates by using multiple quantitative independent variables to predict the probability of group membership (dependent variable). Although the selection of pre-treatment variables considers those variables that explain the variance in-group membership, there is some debate in the literature concerning the appropriate use of pre-treatment covariates (Titus, 2007). In general, however, researchers agree that appropriate independent variables are those variables which explain differences in group membership that are constant over time (Rosenbaum & Rubin, 1983, 1984). In an educational environment, constants include student demographics and academic variables (Jenkins et al., 2009; Rojewski et al., 2010; Xu, Jaggars, & Smith, 2012). In an educational setting, students are found in academic programs, academic programs are found in departments or divisions, divisions are found in institutions, and institutions are found in state systems. Thus, the various levels of data that would be available for each group or entity are “nested” within lower groups of levels of data (Hofmann, 1997, p. 724). Student demographics, academic variables, and institutional variables were the independent covariate variables considered for initial inclusion in the propensity score analyses.

Covariate variables are the independent variables with the highest degree of influence on the dependent variable. Because propensity score matching (PSM) examines grouped data, the dependent variable was categorical instead of quantitative, and logistic or probit

regressions were used to facilitate the selection of variables in the final analyses, after the independent variables were identified and coded logistic regression was used to generate a final propensity score for matching.

Propensity Score Matching

This study explored the use of propensity score matching as a tool to model retention of developmental math students in seven community colleges. The model is complex and utilizes a variety of student, academic and institutional level covariates. Propensity score matching allows the researcher to simplify the analysis by creating a one-number composite of all the covariates and then using the propensity score to match students. Propensity scores represent the "conditional probability of a person being in one condition rather than another given a set of observed covariates used to predict a person's condition" (Rosenbaum & Rubin, 1984, p. 4). Propensity scores range from 0.0 to 1.0, and these scores are used to match students from a large database of a potential comparison group to produce a comparison group that is similar to the study group on the significant covariates. Propensity scores must be assessed to ensure that the distributions are similar across the two groups and that outliers are not present in the propensity scores that could affect the analysis. Box plot examinations are useful for determining whether outliers should be addressed. In some cases, outliers can be eliminated using a minimum maxima technique of common support. Common support implies that if propensity scores fall in the range of 0.14-0.94 for the study group and 0.09-0.79 for the comparison group, then the region of common support using minima-maxima criteria is defined as the interval (0.14-0.79). In no instance is the minimum or

maximum value of the propensity score present in one group and not present in another. The PSMatch2 in STATA allows the researcher to select the option *comsup* to ensure that the region of common support is selected (Leuven & Sianesi, 2003). However, if a small number of outliers are detected or the outliers are believed to represent a portion of the population, the researcher may decide not to eliminate the outliers and continue with the analysis. One option is to run both analyses and determine the differences for both analyses (Rojewski et al., 2010).

Matching refers to a variety of functions that capture students who are similar to each other and create a subset of data that, on average, are balanced in terms of relevant variables. The STATA PSMatch2 function allows the researcher to limit the number of matches and to control the standard deviation or distance from the propensity scores of the student in the comparison group and the corresponding study group. There are several matching algorithms available in STATA but with large sample sizes, such as are often found in student unit record data, the outcomes are similar (Rojewski, et al., 2010)..

The most common matching algorithm used is nearest neighbor matching. Within nearest matching, a few options are available to researchers, specifically, matching with replacement and without replacement. *With replacement* means an individual is considered more than once in the matching procedure. Matching *without replacement* means the case is removed from further consideration for matching. Both types of nearest neighbor matching will affect the variance explained by the model and the bias on key indicators. With replacement is preferred when there are many cases in the treated group with high propensity

scores but only a few matching cases in the comparison group (Caliendo & Kopeinig, 2008). However, tests using the “no replacement” option in STATA, yielded almost equivalent numbers of students in both groups and balanced distributions. A nearest neighbor match with replacement led to bad matches and biased results.

In addition, nearest neighbor matching can lead to bad matches when the nearest neighbor is far away from the matched partner. In other words, the propensity scores are distanced from each other and lead to biased matches. Caliper matching is used to impose a common support range when running the matched procedure. The caliper default in STATA PSMatch2 is .01. The downside is that the researcher does not know beforehand which caliper setting yields the best match. However, in this study, the default setting of caliper of .01 was not sensitive to adjustments when different settings were used. In addition, prior to the imposition of caliper matching with common support, the matching procedure in STATA did not yield reasonable or appreciable differences in bias before and after matching on most of the covariates. Since the sample size in this study is large with 2,102 cases, the imposition of common support and caliper settings of .01 were used to ensure reasonable estimates of study effects and balanced matches (Caliendo & Kopeinig, 2008). Some of the analyses reported in this study were generated multiple times to test the effect of different matching algorithms; in general, the results were similar but the best balance between variance and bias utilized the “no replacement” option. Matching can be thought of as a method of eliminating cases so that the remaining cases show good balance and overlap. (Gelman & Hill, 2007).

After matching, the quality of the match should be assessed and measured statistically using t-tests or chi-square as appropriate. *P* values and effect sizes also should be reported in the results sections (Rojewski, Gemici, & Lee, 2009). Outcomes of interest are reported after the matching procedure and measure the treatment effect or difference in outcomes between the two groups. In theory, any differences in outcomes between the two groups, is associated with the treatment or intervention, since PSM has eliminated the variation between the two groups (Rojewski et al., 2009).

Research Questions and Hypotheses

The study used propensity score methods to examine the following research questions and hypotheses:

1) What are the demographics and academic characteristics of the study population?

2) What are the demographics and academic characteristics of the two groups in the study?

3) Is there a difference in demographics and academic characteristics of the two study groups prior to propensity score matching?

Hypothesis 1

Ho: There is no difference between students who first enroll in developmental math and the conditional probability of completing with a grade of C or better one college-level math course.

Ha: There is a statistically significant difference in between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

4) Is there a difference in demographics and academic characteristics of the two study groups after propensity score matching?

Hypothesis 2

Ho: After propensity score matching, there is a statistically significant difference between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

Ha: There is no difference in between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

5) Is there a difference between the two study groups in college outcomes after propensity score matching?

Hypothesis 3

Ho: There is no difference in student outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

Ha: There are statistically significant outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

Hypothesis 4

Ho: There is no difference in student outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course between institutions.

Ha: There are statistically significant different outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course between institutions.

Data Analysis

In order to answer the research question regarding the demographics and academic characteristics of the study population, descriptive data including frequency tables were reported, and the demographics and academic characteristics of the two groups were described. The two groups in the propensity study were the study group of students completing college-level math and the comparison group not completing college-level math. Descriptive data including frequency tables were reported.

To analyze whether there is a difference in demographics and academic characteristics of the two study groups prior to propensity score matching, t-tests will be reported. Means, standard deviation, significance levels of $p \leq .05$, t values and Cohen's d (effect size) were reported using t-tests analyses. Logistic regression was also used to determine the background and academic factors that explain membership in the two groups of study, as well as to create a propensity score used in the matching technique. Logistic stepwise regression yielded significant independent variables that predicted membership in the comparison or study groups. An analysis of Nagelkerke R-Squared, chi-squared, beta

coefficients, and independent variables with a value of $p \leq .05$ indicated biased predictors, retained in the model (Rojewski et al., 2009).

Using the propensity score generated through the logistic regression technique, equivalent groups of students were matched using a matching procedure in STATA software (Titus, 2007). The technique, known as ps-match2, creates two matched groups based on the propensity score reflecting the propensity to complete college-level math. After matching, to determine if there is a difference in college-level outcomes between developmental math students who complete college math coursework compared to students who do not complete math coursework between institutions, t-tests and a percentage of bias reduction due to matching is appropriate. Means, standard deviation, significance levels of $p \leq .05$, t values, and Cohen's d are analyzed and reported using t-tests analyses (Rojewski et al., 2009).

If the results indicate that there was significant bias in the groups prior to matching, the matched group was used in the final outcome analyses. T-tests analyses were used to predict the average treatment effect of completing college-level math on terms retained, and chi-square is appropriate for the categorical outcomes (Gelman & Hill, 2007; Titus, 2007). Beta coefficients and independent variables with p values of $< .05$ will be reported. The value of R- squared, a measure of explained variance, was also reported (Gelman & Hill, 2007; Titus, 2007).

Summary

Chapter 3 described the research methodology appropriate to the two-level nature of the analysis. The research methodology that the study utilized is known as multilevel

propensity score matching, and the study simultaneously examined student and institutional variance as explanatory variables in developmental math student progression at the community college.

The study population consisted of the fall 2007-08 new student cohorts at seven community colleges. Student progress and outcomes were tracked for five years. The student record dataset included demographics, course completion data, and transfer information available through the National Student Clearinghouse. Through participation in the Completion by Design Initiative, a project of the Tides Center funded by the Bill and Melinda Gates Foundation, the colleges were provided with a raw data extract. The extract included individual college submitted data provided to the Completion by Design partners and a series of derived measures provided by the Columbia University Community College Research Center (CCRC), a data partner in the Completion by Design Initiative. In addition, the raw data extract was analyzed in order to create additional measures for this study. Pre-screening data assessed the potential for bias and observed differences that are a function of similar independent variables. The goal of propensity matching was not to create groups that are equal across all independent variables but to create a purposeful sample of students who are equivalent on average (Gelman & Hill, 2007). Multi-collinearity was minimized in this study and outliers were assessed to determine the impact that outliers have on the subsequent analysis of college outcomes. The selection of pre-treatment covariates was described and was aligned with previous studies in developmental education programming, with a specific focus on developmental math student progression.

CHAPTER IV: RESULTS

Overview

This study explores the use of propensity score matching as a tool to model retention of developmental math students in seven community colleges. The model is complex and utilizes a variety of covariates at the student, academic, and institutional levels. Propensity score matching allows the researcher to simplify the analysis by creating a one-number composite of all the covariates and then using the propensity score to match students. Propensity scores represent the "conditional probability of a person being in one condition rather than another given a set of observed covariates used to predict a person's condition" (Rosenbaum & Rubin, 1984, p. 4).

While there are varying definitions of student success in the community college (Wild & Ebbers, 2002), recent attention is focused on persistence and retention of students in terms of certificate or degree completion and/or a successful transfer to a four-year institution. In a community college context, students who begin work in developmental math courses are a significant proportion of students in the community college (Clery & Topper, 2008). The developmental math student's progression and retention from developmental math courses to the college-level gateway math course and subsequent success were the focus of this study.

Data Analyses

The data analyses began with an examination of the merged multi-campus student record data. After data cleansing, as described in Chapter III, 20,879 student records

(77.70%) were kept in the final data set. The 20,879 records included all students in the 2007-2008 cohorts with an official grade in one or more credit bearing courses. Since CBD colleges report all course enrollments, including curriculum and extension courses, the researcher removed students from the dataset who did not enroll in credit-bearing courses during the first term of enrollment, yielding a reduced dataset of 17,060 students (81.71%) of the original 20,879 student records. Next, the researcher selected the study group of interest. A student was considered a developmental math student if the first record of enrollment in math coursework was in a developmental math course (Bahr, 2008). There are students who are referred to such developmental math coursework that do not take the courses, but these students could not be reliably determined from the dataset. Math placement scores were missing on 20% of the students. In total, 5,548 students (33%) were identified as developmental math students and selected for potential inclusion in the study group.

Due to the large percentage of students that take developmental courses in math at the community college but do not progress to college-level math, the next selection criteria involved selecting those students who had progressed from the developmental math coursework and attempted to complete college-level math. From the attempted college-level math group, two groups were identified and used as the group variable in the initial logistic regression that yielded the propensity score estimation. The two groups were students who completed developmental math at a sufficient level to enroll in college-level math coursework at MAT-101 or higher. The initial group of 5,548 developmental math students

was reduced to a final sample size of 2, 102 developmental math students (38%) in seven institutions.

The students in the 2,102 case sample dataset defined the two groups of interest in the study. Two groups were determined of students who met the criteria of attempting college-level math after successful completion of developmental math. The group dependent variable was coded as 1 for students who passed at least one college-level math course and 0 for students who failed to pass at least one college-level math course with a grade of C or better. Of the 2,102 students, 1,563 students (74%) passed at least one college-level math course with a grade of C or better.

The final analytical sample used in the propensity score model consisted of 2,102 students. Females outnumbered males in the sample, as 56.52% of students in the sample were females (n=1188). The majority of students, 64.61% (n=1358), were 19 years of age and under in their first term of enrollment. However, 21.02% (n=442) of students were 24 years of age and older. Over half of students, 57.71% (n=1213) were White, and Black students represented the next largest group of students at 21.98% (n=462). International students represented 6.57% (n=138) of the sample and Hispanics represented 4.28 % of students (n=90). Asian Americans were represented but only slightly, at 2.57% (n=54), and the American Indian group was so small (n=13) that the students were re-coded as “Other” and combined with that race category, yielding a total of 145 students (6.90 %) sampled. The other race category was used as the reference group in the race categories due to the number

of students in the group and the decision to retain the students in the analysis. Table 9 shows the demographic information for the 2,102 students in the study database.

Table 9: Demographic Information for 2,102 Students in Study Database

Category	*N*	Percent
Gender		
Male	914	43.48
Female	1,188	56.52
Age Range		
19 and Under	1,358	64.61
20 – 23	302	14.37
24 – 62	442	21.03
Race		
White	1,213	57.71
Black	462	21.98
Hispanic	90	4.28
International	138	6.57
Asian	54	2.57
Other	145	6.90
Late Entry ($\geq$ 24 years old)		
Yes	442	21.03
No	1,660	78.97

Less than half of students, 44.55% (n=936), declared a college transfer program and the majority, 65.60% (n=1379), were enrolled full-time in their first semester. Almost all of

the students, 91.48% (n=1923), were retained to the second major term. Students who passed the college-level math course earned a higher grade point average in the first term compared to students who did not pass the college-level math course, with GPAs of 2.62 and 2.07, respectively. Table 10 and Table 11 show the first-term achievement of students in the study group and the mean and standard deviation of the study population's first-term GPA, respectively.

Table 10: Frequency and Percentage of First-Term Achievement (Enrollment, Pell Grant, and Retention after term) of 2,102 Students in Database

Category	*N*	Percent
Enrollment		
Transfer Program	936	44.53
Non-transfer Program	1,166	55.47
Full-time	1,379	65.60
Part-time	723	34.40
Pell Grant Recipient		
Yes	631	30.02
No	1,471	69.98
Retention after First Term		
Yes	1,923	91.48
No	179	8.52

Table 11: Mean and Standard Deviation of First-Term GPA of 2,102 Study and Comparison Group Students in Database

	M	*SD*
GPA after First Term		
Study Group	2.62	1.41
Comparison Group	2.07	1.47

Coursework selections for students in the database indicated that 39.63% (n=833) also enrolled in developmental English coursework, and one third of students also enrolled in a student success course. Over half of the student group, 56.71% (n=1192), enrolled in the highest-level developmental math course, MAT-080, and 8.42% (n=177) enrolled in the lowest level developmental math course, MAT-050. Not all students taking college-level math completed MAT-080 and, in general, the study population was taken from about half, or 56.71% (n=1192), of students completing MAT-080, with the remainder of students, 64.80% (n=1362), completing at least MAT-070. Table 12 indicates coursework selections of the students in the study population.

Table 12: Coursework Selections for 2,102 Students in Database

Category	*N*	Percent
Developmental English		
Yes	833	39.63
No	1,269	60.37
Student Success Course		
Yes	751	35.73
No	1,351	64.27
Developmental Math (080)		
Yes	1,192	56.71
No	177	8.42
Unknown / Missing Data	733	34.87
Developmental Math (070)		
Yes	1,362	64.80
No	57	2.71
Unknown / Missing Data	683	32.49
Developmental Math (060)		
Yes	463	22.03
No	30	1.43
Unknown / Missing Data	1,609	76.56
Developmental Math (050)		
Yes	170	8.09
No	7	0.33
Unknown / Missing Data	1,925	91.58

The study found high completion rates in a student's first developmental math course, ranging from 96.05% (n=170) in MAT-050 to 87.71% (n=514) in MAT-080. However, a student's likelihood of completion in subsequent developmental math courses varied. Less than half of students, 45.76% (n=81), who started in MAT-050 completed MAT-080. In addition, a smaller percentage, 38.02% (n=138), who started in MAT-060 completed MAT-080. Also, 47.05% (n=459) of students who first enrolled in MAT-070 completed MAT-080. In other words, while the study population was comprised of students from each level of developmental math, fewer students who started at the lowest level of developmental math represented the study population: 8.42% of students (n=177) started in MAT-050; 17.27% (n=363) started in MAT-060; 46.43% (n=976) started in MAT-070; and 27.88% (n=586) started in the highest-level math course, MAT-080. However, pass rates in the college-level math course were similar and not related to the level of the course in which a student was first enrolled. Overall, 1,563 (74%) of the total sample population passed at least one college-level math course. Table 13 indicates the progression of math courses in the study population.

Table 13: Frequency and Percent of Developmental Math Progression to Completion of College-Level Math in Relation to First Developmental Math Course in which Student was Enrolled

Category		N	Percent of Total Students in Category
First Developmental Math Course	MAT-050	177	100.00
	Completed MAT-050	170	96.05
	Completed MAT-060	126	71.19
	Completed MAT-070	146	82.49
	Completed MAT-080	81	45.76
	First Dev. MAT-050 & Passed College Math	132	74.58
First Developmental Math Course	MAT-060	363	100.00
	Completed MAT-060	337	92.84
	Completed MAT-070	273	75.21
	Completed MAT-080	138	38.02
	First Dev. MAT-060 & Passed College Math	265	73.00
First Developmental Math Course	MAT-070	976	100.00
	Completed MAT-070	943	96.62
	Completed MAT-080	459	47.03
	First Dev. MAT-070 & Passed College Math	729	74.69
First Developmental Math Course	MAT-080	586	100.00
	Completed MAT-080	514	87.71
	First Dev. MAT-080 & Passed College Math	437	74.57

After data exploration, the data analysis procedure began with running logit analyses on the selected independent variables in STATA and estimating and testing propensity scores to be used in the final models. The logit analyses were executed in STATA using the PScore function. The PScore function checked for propensity score balance, and the researcher was instructed to check the model if balanced propensity scores were not achieved. During the initial analyses, age and late-entry variables were problematic and lacked balance. During pre-screening, the two variables were correlated but not in the threshold of .95 or higher and retained. However, the propensity model failed to converge entering both variables and, as a result, late entry was removed from the analysis. The propensity models achieved balance on the remaining covariates.

Propensity scores ranged from 0.0 to 1.0 and these scores were used to match students from a database of a potential comparison group to produce a comparison group that was similar to the study group on the significant covariates. Propensity scores were assessed to ensure that the distributions were similar across the two groups and that outliers were not present in the propensity scores that could affect the analysis. Outliers were eliminated using a minimum maxima technique of common support. Common support implied that if propensity scores fell in the range of 0.14-0.0.94 for the study group and in the range of 0.09-.0.79 for the comparison group, then the region of common support using minima-maxima criteria was defined as the interval, 0.14-0.79. In no instance was the minimum or maximum value of the propensity score in one group and not present in another (Rojewski, et al., 2009). The PSMatch2 function in STATA allowed the researcher to select the option *comsup* to

ensure that the region of common support and overlap was selected (Leuven & Sianesi, 2003). After the propensity scores were estimated, PSMatch2 was executed in STATA to create the matched groups for the final analyses.

The researcher chose the most common matching algorithm used, which is nearest neighbor matching. Within nearest matching, a few options are available to researchers, specifically, matching with replacement and without replacement. *With replacement* means an individual is considered more than once in the matching procedure. Matching *without replacement* means the case is removed from further consideration for matching. Both types of nearest neighbor matching affected the variance explained by the model and the bias on key indicators. With replacement was preferred, since there were many cases in the treated group with high propensity scores, but only a few matching cases in the comparison group. The researcher attempted a nearest neighbor match with replacement, but the procedure led to bad matches and biased results. Analyses using the “no replacement” option in STATA, yielded almost equivalent numbers of students and balanced distributions in both groups and matching without replacement was chosen for the study.

Caliper matching was also used to impose a common support range when running the matched procedure. The caliper default in STATA PSMatch2 is .01. The challenge is the researcher does not know beforehand which caliper setting yields the best match (Caliendo & Kopeinig, 2008). However, in this study, the default setting of caliper of .01 was not sensitive to adjustments when different settings were used. In addition, prior to the imposition of caliper matching with common support, the matching procedure in STATA did not yield

reasonable or appreciable differences in bias before and after matching on most of the covariates. Since the sample size in this study was large with 2,102 cases, the imposition of common support and caliper settings of .01 were used to ensure reasonable estimates of study effects and better-balanced matches (Caliendo & Kopeinig, 2008; Titus, 2007). Some of the analyses reported in this study were generated multiple times to test the effect of different matching algorithms; in general, the results were similar, but the best balance between variance and bias utilized the “no replacement” option. Matching can be thought of as a method of eliminating cases so that the remaining cases show good balance and overlap (Gelman & Hill, 2007).

Analyses yielded the best balance in bias and variance utilizing nearest neighbor matches, common support, without replacement, and caliper thresholds of .01. Logistic regressions were also used to estimate a propensity score based on each model configuration. In total, four models were operationalized in this study, and in each model the steps were repeated to create different matched groups for analyses. Table 14 describes the model characteristics.

Table 14: Grouping of Independent Variables According to Model Iterations

Variable Name	Model I Student Demographics	Model II Academic Integration	Model III Individual & Institution	Model IV Individual & Institution
Gender	X	X	X	X
Age	X	X	X	X
Race*	X	X	X	X
Transfer Program	X	X	X	X
Pell Recipient	X	X	X	X
Full-Time in 1st Term	X	X	X	X
First Dev. Math Level	X	X	X	X
Retained 2nd term		X	X	X
Dev. English		X	X	X
Success Course		X	X	X
First Term GPA		X	X	X
% Minority**			X	
% Pell Recipients**			X	
% Dev. English ***			X	
% Dev. Math ***			X	
Institution Member				X

**Race categories: White, Black, Hispanic, International, Asian, Other*
*** Number of students as a percentage of fall 2007 enrollment*
****Number of students as a percentage of 2007-08 institutional FTE (full-time equivalents).*

Logistic regression analyses were used to determine the background and academic factors that explain membership in the two groups of study and to create a propensity score that was used in the matching technique. Logistic stepwise regression yielded significant independent variables that predicted membership in the comparison or study groups (Rojewski et al., 2010). An analysis of Nagelkerke R-Squared, chi-squared, beta coefficients,

and independent variables with *p* value $\leq$ .05 indicated significant predictors and covariates that were retained in the model (Hair et al., 2010).

Stepwise logistic regression was conducted to determine which independent variables in Model I were associated with the dependent variable— completing or not completing college-level math with a grade of C or better. Logistic regression results indicated that the overall model of four predictors (Gender, Age, Black Race, Pell Recipient) were statistically reliable in predicting membership in the dependent variable (-2 Log Likelihood= 2328.79), chi-squared = 64.471, p< .001, Nagelkerke R Squared= .044).

The model correctly classified 74.3% of the cases and explained 4.4% of the variance in the dependent variable. Regression coefficients are presented in Table 15. Wald statistics indicated that four variables significantly predict group membership in the dependent variable. Female students were 41% more likely to pass college-level math. Age was a positive predictor of passing college-level math, but only slightly at 1.8%. Black students were 47% less likely to pass college-level math and Pell recipients in the first term were 27% less likely to pass college-level math than non-Pell recipients with a grade of C or better.

Table 15: Model I—Results of Logistic Regression Predicting Membership in Groups Utilizing Demographics, Beta Coefficients, Standard Error, Wald, Degrees of freedom, Significance Levels, and Odds Ratios

Covariate	*B*	S. E.	Wald	df	Sig.	Exp(*B*)
Gender	0.344	0.104	11.003	1	0.001	1.411**
Age	0.017	0.008	5.183	1	0.023	1.018*
White	0.058	0.209	0.077	1	0.781	1.060
Black	-0.631	0.220	8.266	1	0.004	0.532**
Asian American	0.197	0.384	0.265	1	0.607	1.218
Hispanic	-0.197	0.307	0.412	1	0.521	0.821
International	0.274	0.292	0.885	1	0.347	1.316
Transfer	-0.107	0.104	1.052	1	0.305	0.899
Pell Recipient in First Term	-0.306	0.116	6.979	1	0.008	0.736**
Full Time in First Term	0.136	0.110	1.521	1	0.217	1.146
First Dev. Math Level	-0.018	0.072	0.060	1	0.806	0.983
Constant	0.698	0.301	5.393	1	0.020	2.010

**p <.05 **p < .01 ***p < .001*
(-2 Log Likelihood= 2328.79, chi-squared = 64.471, p <.001, Nagelkerke R Squared= .044)

Logistic regression was conducted to determine which independent variables in Model II were significantly associated, $p < .05$, with the dependent variable— completing or not completing college-level math with a grade of C or better. Model II includes the independent variables used in Model I and additional variables used to indicate a student's progress. Logistic regression results indicated that the overall model of four predictors (Gender, Black Race, Pell Recipient, First Term GPA) were statistically reliable in predicting

membership in the dependent variable (-2 Log Likelihood= 2282.99, chi-squared = 110.26, $p < .001$, Nagelkerke R Squared= .075).

The model correctly classified 74.3% of the cases and explained 7.5% of the variance in the dependent variable. Regression coefficients are presented in Table 16. Wald statistics indicated that four variables significantly predict group membership in the dependent variable. Female students were 37% more likely to pass college-level math. Age was a positive predictor of passing college-level math, but only slightly in Model I and was not significant in Model II. Black students were 40% less likely to pass college-level math and Pell recipients in the first term were 27% less likely than non-Pell recipients to pass college-level math with a grade of C or better. First-term grade point average entered the second model as a significant positive predictor of success in college-level math, and students with higher grade point averages were 24% more likely to pass college-level math with a grade of C or better.

Table 16: Model II—Results of Logistic Regression Predicting Membership in Groups Utilizing Demographics and Academic Integration Covariates, Beta Coefficients, Standard Error, Wald, Degrees of freedom, Significance Levels, and Odds Ratios

Covariate	*B*	S. E.	Wald	df	Sig.	Exp(*B*)
Gender	0.314	0.106	8.866	1	0.003	1.369*
Age	0.009	0.008	1.354	1	0.244	1.009
White	0.035	0.211	0.027	1	0.869	1.036
Black	-0.499	0.224	4.958	1	0.026	0.607*
Asian American	0.233	0.389	0.359	1	0.549	1.263
Hispanic	-0.075	0.311	0.059	1	0.809	0.927
International	0.415	0.297	1.943	1	0.163	1.514
Transfer	-0.142	0.106	1.792	1	0.181	0.868
Pell Recipient in First Term	-0.307	0.118	6.796	1	0.009	0.736**
Full Time in First Term	0.058	0.115	0.252	1	0.616	1.059
First Dev. Math Level	0.031	0.074	0.169	1	0.681	1.031
Retained 2nd	0.319	0.177	3.247	1	0.072	1.376
Enrolled in Dev. English	-0.133	0.118	1.275	1	0.259	0.876
Enrolled in Success Class	-0.047	0.113	0.172	1	0.678	0.954
First Term GPA	0.211	0.038	31.607	1	0.000	1.235***
Constant	0.12	0.337	0.126	1	0.723	1.127

**p < .05 **p < .01 ***p < .001*
(-2 Log Likelihood= 2282.99, chi-squared = 110.26, p <.01, Nagelkerke R Squared= .075)

Logistic regression was conducted again, entering the institutional- level variables to determine which independent variables in Model III were significantly associated with the dependent variable—completing or not completing college-level math with a grade of C or better. Model III includes the independent variables used in Models I and Model II and institutional variables to capture institutional variation. Logistic regression results indicated that the overall model of seven predictors (Gender, Black Race, Pell Recipient, First Term GPA, Percent Minority, Percent Pell, Percent Developmental English) were statistically reliable in predicting membership in the dependent variable (-2 Log Likelihood= 2246.950), chi-squared= 146.309, $p < .001$, Nagelkerke R Squared= .099).

The model correctly classified 74.9% of the cases and explained 10%, or an additional 2.7%, of the variance in the dependent variable. Regression coefficients are presented in Table 17. Wald statistics indicated that seven variables significantly predict group membership in the dependent variable. In the multilevel model, female students were 38% more likely to pass college-level math. Black students were half (51%) as likely to pass college-level math, and Pell recipients in the first term were 25% less likely than non-Pell recipients to pass college-level math with a grade of C or better. First-term grade point average also remained in the second model as a significant positive predictor of success in college-level math, and students with higher grade point averages were 26% more likely to pass college-level math with a grade of C or better. Students in institutions with higher percentages of minorities were slightly (5%) more likely to pass college-level math with a grade of C or better. Students in institutions with higher percentages of Pell recipients and

developmental English enrollments were slightly less likely to pass college-level math, 5% and 10%, respectively. Table 17 outlines the results of the logistic regression.

Table 17: Model 1II—Results of Logistic Regression Predicting Membership in Groups Utilizing Demographics, Academic Progress, and Institutional Membership Indicators, Beta Coefficients, Standard Error, Wald, Degrees of Freedom, Significance Levels, and Odds Ratios

Covariate	*B*	S. E.	Wald	df	Sig.	Exp(*B*)
Gender	0.320	0.106	9.034	1	0.003	1.377**
Age	0.010	0.008	1.692	1	0.193	1.010
White	-0.037	0.158	0.055	1	0.815	0.964
Black	-0.720	0.172	17.471	1	0.000	0.487***
Hispanic	-0.245	0.277	0.782	1	0.376	0.783
Transfer	-0.078	0.110	0.505	1	0.478	0.925
Pell Recipient in First Term	-0.279	0.120	5.410	1	0.020	0.756*
Full Time in First Term	0.067	0.118	0.321	1	0.571	0.935
First Dev. Math Level	0.053	0.075	0.504	1	0.478	1.054
Retained 2nd	0.212	0.180	1.391	1	0.238	1.236
Enrolled in Dev. English	-0.090	0.118	0.582	1	0.446	0.914
Enrolled in Success Class	0.031	0.119	0.066	1	0.797	1.031
First Term GPA	0.226	0.038	35.069	1	0.000	1.254***
Percent Minority	0.053	0.015	12.696	1	0.000	1.055***
Percent Pell Recipient	-0.057	0.019	9.137	1	0.030	0.945*
Percent Dev Eng	-0.090	0.034	6.771	1	0.009	0.914**
Percent Dev Math	0.021	0.013	2.400	1	0.121	1.021
Constant	0.623	0.624	0.996	1	0.318	1.865

**p < .05 **p < .01 ***p < .001*
(-2 Log Likelihood= 2246.950, chi-squared= 146.309, p < .001, Nagelkerke R Squared= .099)

Logistic regression was conducted for the final time, entering the institutional-level variables as membership variables to determine which independent variables in Model III were significantly associated with the dependent variable—completing or not completing college-level math with a grade of C or better. Model III includes the independent variables used in Models I and Model II and institutional membership variables used to capture institutional variation. Logistic regression results indicated that the overall model of seven predictors (Gender, Black Race, Pell Recipient, First Term GPA, College B, College F, College G) were statistically reliable in predicting membership in the dependent variable.

The multilevel model correctly classified 75.5% of the cases and explained 11.6%, or an additional 4.1%, of the variance in the dependent variable compared to the single-level model. Regression coefficients are presented in Table 18. Wald statistics indicated that seven variables significantly predict group membership in the dependent variable. In the multilevel model, female students were 35% more likely to pass college-level math. Black students were half (51%) as likely to pass college-level math and Pell recipients in the first term were 25% less likely than non-Pell recipients to pass college-level math with a grade of C or better. First-term grade point average also remained in the second model as a significant positive predictor of success in college-level math, and students with higher grade point averages were 24% more likely to pass college-level math with a grade of C or better. With College A as the reference group, students in College B were 32% more likely to pass college-level math, and students in College F were 197% more likely to pass college-level math. However, students in College G were 54% less likely to pass college-level math than

students in the reference group. Overall, moving from a single-level to multilevel model explained more variance in the two groups, from 4.4% to 11.6%, and suggested that within institutional variance was evident in the student data.

Table 18: Model 1V—Results of Logistic Regression Predicting Membership in Groups Utilizing Demographics, Academic Progress and Institutional Membership Indicators, Beta Coefficients, Standard Error, Wald, Degrees of Freedom, Significance Levels, and Odds Ratios.

Covariate	*B*	S. E.	Wald	df	Sig.	Exp(*B*)
Gender	0.297	0.107	7.680	1	0.006	1.345**
Age	0.008	0.008	1.068	1	0.301	1.008
White	-0.089	0.159	0.311	1	0.577	0.915
Black	-0.715	0.173	17.506	1	0.000	0.484***
Hispanic	-0.267	0.278	0.921	1	0.337	0.665
Transfer	-0.079	0.110	0.512	1	0.474	0.924
Pell Recipient in First Term	-0.282	0.121	5.445	1	0.020	0.754*
Full Time in First Term	0.004	0.120	0.001	1	0.972	1.004
First Dev. Math Level	0.038	0.076	0.255	1	0.613	1.039
Retained 2nd	0.213	0.181	1.375	1	0.241	1.237
Enrolled in Dev. English	-0.136	0.120	1.282	1	0.258	0.873
Enrolled in Success Class	0.060	0.122	0.243	1	0.622	1.062
First Term GPA	0.211	0.039	29.885	1	0.000	1.235***
College B	0.300	0.128	5.497	1	0.019	1.350*
College C	-0.263	0.265	0.987	1	0.320	0.769
College D	0.892	0.575	2.410	1	0.121	2.440
College E	0.033	0.275	0.014	1	0.906	1.033
College F	1.086	0.242	20.110	1	0.000	2.963***
College G	-0.779	0.197	15.679	1	0.000	0.459***
Constant	0.231	0.316	0.534	1	0.465	1.260

$*p < .05$ $**p < .01$ $***p < .001$
(-2 Log Likelihood= 2221.665, chi-squared = 171.595, $p < .001$, Nagelkerke R Squared= .115)

To further analyze whether there was a difference in demographics and academic characteristics of the two study groups prior to propensity score matching, t-tests were provided and analyzed. Means, standard deviation, significance levels, T-values and Cohen's d (effect size) were reported using t-tests analyses. Bias percentages (standardized effect sizes) before and after matching were also reported (Oakes & Johnson, 2006). The standardized percent bias is the percent difference of the sample means in the treated and non-treated (full or matched) sub-samples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups (Rosenbaum & Rubin, 1985). Given the number of students in the initial sample, significance levels in t-tests analyses are less reliable than effect sizes as indicated in Cohen's d and reported biased predictors (Rojewski, et al., 2010).

Since the study examined retention of developmental math students and subsequent success in college-level math, four iterations of analyses were constructed to conceptualize the retention model. Each time logistic regression was utilized to create a propensity score, an analysis examined the results of before and after matching on the covariates. The tables below report the results of propensity score matching and the logical movement of the individual level to multilevel analyses in this study.

Model I

The first model examined the student demographics of gender, age, race, transfer goals, enrollment status, Pell recipient status, and the first developmental math course in which a student enrolled. The initial t-tests yielded significant differences in gender,

t(2,100)= 2.69, $p < .01$ and Pell recipients, t(2,100)= -3.53, $p < .01$. The results also indicated that students passing college-level math with a C or better differed significantly for White, t(2,100)=4.37, $p < .01$ and Black students, t(2,100)= -6.52, $p < .01$ compared to the other race categories. Before matching, the absolute values in bias ranged from 3.0% to 31%, with an average bias of 10.4%.

After matching, the absolute value in bias ranged from 3.2% to 23%, and average bias remained high at 10.3%. In short, the model did not achieve good balance. Caliendo and Kopeinig (2008) recommended that predictors after matching yield 5% or less bias. Two predictors, Black race and Transfer student, met the criteria of bias $\leq$ 5%. Gelman and Hill (2007) contended that the goal of propensity matching is to achieve equivalent groups on average of students in both the study and comparison group. Failure to achieve balance after propensity matching suggested a different specification of the model was needed, and few predictors were equivalent on average. Table 19 and Table 20 outline the Model I findings before and after propensity matching, respectively.

Table 19: Model I— Number in Study and Comparison Groups, Group Means, Standard Deviation, T-tests, *p* values, Cohen's d, and Standardized Bias of Covariates Prior to Propensity Matching

Note: First line of each variable is the comparison group; second line of each variable is the study group.

Covariate	*N*	*M*	*SD*	*t*	*p*	*d*	% bias
Gender	539	0.516	0.500	2.687	0.007**	0.133	13.400
	1,5	0.582	0.493				
Age	539	21.453	6.909	1.840	0.066	0.093	9.400
	1,5	22.141	7.705				
White	539	0.497	0.500	4.370	0.000***	0.218	21.700
	1,5	0.605	0.489				
Black	539	0.319	0.200	-6.520	0.000***	-0.309	-31.100
	1,5	0.186	0.009				
Hispanic	539	0.046	0.211	-0.470	0.636	-0.049	-2.300
	1,5	0.042	0.199				
International	539	0.051	0.218	1.690	0.091	0.088	8.800
	1,5	0.710	0.257				
Asian	539	0.022	0.147	0.580	0.560	0.032	3.000
	1,5	0.027	0.161				
Transfer	539	0.466	0.499	-1.100	0.270	-0.003	-5.500
	1,5	0.438	0.496				
Pell Recipient	539	0.359	0.480	-3.530	0.000***	-0.172	-17.300
	1,5	0.280	0.449				
Full-Time	539	0.657	0.475	-0.040	0.967	-0.002	-0.200
	1,5	0.656	0.475				
First Dev. Math	539	1.989	0.736	-0.400	0.692	-0.027	-2.000
	1,5	1.974	0.730				
Average Bias							10.400
R-Squared							0.027

Cohen's d = Mt – Mc / σpooled where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < 0.001*

Table 20: Model I— Number in Study and Comparison Groups, Group Means, Standard Deviation, T-tests, *p* values, Cohen's d and Standardized Bias of Covariates After Propensity Matching

Note: First line of each variable is the comparison group; second line of each variable is the study group.

Covariate	*N*	*M*	*SD*	*t*	*p*	*d*	% bias
Gender	537	0.516	0.500	-1.190	0.234	-0.080	-8.100
	534	0.479	0.500				
Age	537	21.376	6.697	-1.108	0.268	0.067	6.000
	534	21.824	6.522				
White	537	0.497	0.500	-3.185		-0.201	-20.200
	534	0.401	0.491				
Black	537	0.318	0.466	0.389	0.696	0.036	3.800
	534	0.330	0.470				
Hispanic	537	0.047	0.211	0.974	0.330	0.053	6.300
	534	0.059	0.238				
International	537	0.051	0.219	0.178	0.075	0.106	11.000
	534	0.077	0.267				
Asian	537	0.023	0.148	1.291	0.197	0.078	8.400
	534	0.036	0.185				
Transfer	537	0.467	0.499	0.393	0.695	0.032	3.200
	534	0.479	0.500				
Pell Recipient	537	0.359	0.480	2.080	0.038*	0.138	14.400
	534	0.421	0.494				
Full-Time	537	0.657	0.475	-1.529	0.126	-0.091	-9.300
	534	0.613	0.487				
First Dev. Math	537	1.987	0.737	3.765	0.002**	0.229	23.000
	534	2.155	0.728				
Average Bias							10.300
R-Squared							0.041

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Model II

Model II was the second model examined and included the remaining individual student predictors. The model retained the predictors from Model I and added academic integration variables that, in theory, predict student retention. The academic progress variables that were added were retained to second major term, first-term grade point average, and enrollment in developmental English and student success coursework. Standardized difference tests indicated, prior to matching, gender remained significant in Model II and bias was 13%. Black and White race also remained significant. Prior to matching the bias for White students was 21% and for Black students the bias was -31.1%. A negative bias indicated that the mean for the comparison group was higher than the mean for the study group or that Black students were less likely to complete college-level math with a grade of C or better. Pell recipients in the first term was also negatively biased, -17.3%, indicating that Pell recipients were less likely to complete college-level math compared to non-Pell recipients. The mean difference in first-term grade point average was also significant and the bias prior to matching was positive and high, at 38.3%. Students in the study group earned an average first-term grade point average of 2.62 compared to 2.07 for students in the comparison group. The average bias was 12.5% prior to matching and 6 covariates yielded bias estimators greater than the recommended threshold of less than or equal to 5%.

After matching, the reduction in bias was 66.4% and the initial average bias of 12.5% was 4.2% after matching. Since the goal of propensity matching was to create two equivalent groups on average, Model II yielded a stronger matched group of students for analyses.

Gender, Black, and Hispanic race were biased predictors that failed to meet the threshold of less than 5% after matching. However, the reduction in bias in Black race was large from -31% prior to matching to 5.9% after matching. Prior to matching, retention to the second major term was significant and positive with an average mean of .924 for the study group and .889 for the comparison group. After matching, the comparison group had a slightly higher mean than the study group, but the two groups were balanced and not significantly different in terms of retention to the second major term. Table 21 and Table 22 outline the Model II findings before and after propensity matching, respectively.

Table 21: Model II— Number in Study and Comparison Groups, Group Means, Standard Deviation, T-tests, *p* values, Cohen's d, and Standardized Bias of Covariates Prior to Propensity Matching

Note: First line of each variable is the comparison group; second line of each variable is the study group.

Covariate	*N*	*M*	*SD*	*t*	*p*	*d*	% bias
Gender	539	0.516	0.500	2.687	0.007**	0.133	13.400
	1,563	0.582	0.493				
Age	539	21.453	6.909	1.840	0.066	0.093	9.400
	1,563	22.141	7.705				
White	539	0.497	0.500	4.370	0.000***	0.218	21.700
	1,563	0.605	0.489				
Black	539	0.319	0.200	-6.520	0.000***	-0.309	-31.100
	1,563	0.186	0.009				

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p <.05 **p <.01 ***p <.001*

Table 21 (continued)

Covariate	N	M	SD	*t*	*p*	*d*	% bias
Hispanic	539	0.046	0.211	-0.470	0.636	-0.019	-2.300
	1,563	0.042	0.199				
International	539	0.051	0.218	1.690	0.091	0.088	8.800
	1,563	0.710	0.257				
Asian	539	0.022	0.147	0.580	0.560	0.032	3.000
	1,563	0.027	0.161				
Transfer	539	0.466	0.499	-1.100	0.270	-0.056	-5.500
	1,563	0.438	0.496				
Pell Recipient	539	0.359	0.480	-3.530	0.000***	-0.172	-17.300
	1,563	0.280	0.449				
Full-Time	539	0.657	0.475	-0.040	0.967	-0.002	-0.200
	1,563	0.656	0.475				
First Dev. Math	539	1.989	0.736	-0.400	0.692	-0.027	-2.000
	1,563	1.974	0.730				
Retained to 2nd	539	0.889	0.314	2.530	0.120	0.120	12.100
	1,563	0.924	0.265				
Dev. Eng	539	0.471	0.499	-4.140	0.000***	-0.206	-20.500
	1,563	0.370	0.483				

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p <.05 **p <.01 ***p < .001*

Table 21 (continued)

Covariate	N	M	SD	*t*	*p*	*d*	% bias
Success	539	0.365	0.482	-0.460	0.645	-0.023	-2.300
	1,563	0.354	0.478				
1st Term GPA	539	2.07	1.473	7.750	0.000***	0.382	38.300
	1,563	2.62	1.406				
Average Bias							12.500
R-Squared							0.046

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 22: Model II—Number in Study and Comparison Groups, Group Means, Standard Deviation, T-tests, *p* values, Cohen's d and Standardized Bias of Covariates After Propensity Matching

Note: First line of each variable is the comparison group; second line of each variable is the study group.

Covariate	*N*	*M*	*SD*	*t*	*p*	*d*	% bias
Gender	531	0.522	0.500	-0.859	0.391	-0.058	-5.700
	529	0.495	0.500				
Age	531	21.478	6.946	-0.757	0.449	-0.049	-4.400
	529	21.170	6.286				
White	531	0.505	0.500	-0.553	0.580	-0.038	-3.800
	529	0.487	0.500				
Black	531	0.309	0.462	0.766	0.444	0.054	5.900
	529	0.331	0.470				
Hispanic	531	0.047	0.212	0.967	0.334	0.058	6.400
	529	0.060	0.239				
International	531	0.051	0.219	-0.717	0.473	-0.043	-4.000
	529	0.042	0.199				
Asian	531	0.023	0.149	0.009	0.993	0.000	-0.100
	529	0.023	0.149				
Transfer	531	0.465	0.499	-0.932	0.352	-0.053	-5.300
	529	0.437	0.496				
Pell Recipient	531	0.351	0.478	0.876	0.381	0.052	5.400
	529	0.376	0.484				

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 22 (continued)

Covariate	*N*	*M*	*SD*	*t*	*p*	*d*	% bias
Full-Time	531	0.655	0.476	-0.367	0.714	-0.025	-2.500
	529	0.645	0.479				
First Dev. Math	531	1.991	0.738	0.040	0.967	-0.005	-0.500
	529	1.992	0.759				
Retained to 2nd	531	0.898	0.303	-0.998	0.319	-0.070	-7.700
	529	0.879	0.326				
Dev. English	531	0.463	0.499	0.858	0.391	0.054	5.400
	529	0.490	0.500				
Success	531	0.363	0.481	0.745	0.456	0.049	4.800
	529	0.386	0.487				
1st Term GPA	531	2.097	1.464	-0.106	0.916	-0.016	-1.700
	529	2.087	1.540				
Average Bias							4.200
R-Squared							0.044

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p <.01 ***p <.001*

Model III

Model III expanded the retention model of student-level predictors to include – institutional-level predictors. College indicators included percent minority and percent Pell-awarded students as a percentage of fall 2007 official enrollments. Remedial costs were examined by comparing the percent of institutional budget FTE as an indicator of costs associated with developmental math and English courses at the various institutions.

T-test analyses indicated that, prior to matching, percent minority was significantly different among the two groups, $t(2,100)= 2.85$, $p< .001$ and a positive biased predictor of success in college-level math of 13.7%. In addition, percent developmental math was also a positive predictor with a high level of bias of 20.4%. T-test analyses indicated that, prior to matching, the two groups differed significantly, $t(2,100)= 4.15$, $p< .001$. In addition, prior to matching, the average amount of bias was 12.3%.

After matching, the average bias was 3.2%. However, standardized bias measures revealed predictors were significantly different on an institutional predictor, percent of institutional full-time equivalent allocated to developmental math. Overall, effect sizes were small and the reduction in bias was over 10%, from 13.0% prior to matching to 2.8% after matching. Table 23 and Table 24 show the findings from Model III before and after propensity matching, respectively.

Table 23: Model III— Number in Study and Comparison Groups, Group Means, Standard Deviation, T-tests, *p* values, Cohen's d, and Standardized Bias of Covariates Prior to Propensity Matching

Note. First line of each variable is the comparison group; second line of each variable is the study group.

Covariate	*N*	*M*	*SD*	*t*	*p*	*d*	% bias
Gender	539	0.516	0.500	2.687	0.007**	0.133	13.400
	1,563	0.582	0.493				
Age	539	21.453	6.909	1.840	0.066	0.093	9.400
	1,563	22.141	7.705				
White	539	0.497	0.500	4.370	0.000***	0.218	21.700
	1,563	0.605	0.489				
Black	539	0.319	0.200	-6.520	0.000***	-0.309	-31.100
	1,563	0.186	0.009				
Hispanic	539	0.046	0.211	-0.470	0.636	-0.019	-2.300
	1,563	0.042	0.199				
Transfer	539	0.466	0.499	-1.100	0.270	-0.056	-5.500
	1,563	0.438	0.496				
Pell-awarded	539	0.359	0.480	-3.530	0.000***	-0.172	-17.300
	1,563	0.280	0.449				
Full-Time	539	0.657	0.475	-0.040	0.967	-0.002	-0.200
	1,563	0.656	0.475				
First Dev. Math	539	1.989	0.736	-0.400	0.692	-0.027	-2.000
	1,563	1.974	0.730				

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < 0.05 **p < 0.01 ***p < 0.001*

Table 23 (continued)

Covariate	N	M	SD	*t*	*p*	*d*	% bias
Retained to 2nd	539	0.889	0.314	2.530	0.120	0.120	12.100
	1,563	0.924	0.265				
Dev. Eng	539	0.471	0.499	-4.140	0.000***	-0.206	-20.500
	1,563	0.370	0.483				
Success	539	0.365	0.482	-0.460	0.645	-0.023	-2.300
	1,563	0.354	0.478				
1st Term GPA	539	2.07	1.473	7.750	0.000***	0.382	38.300
	1,563	2.62	1.406				
% Minority	539	40.675	13.871	2.850	0.004**	0.136	13.700
	1,563	42.425	11.667				
% Pell-awarded	539	33.293	7.447	0.056	0.574	0.027	2.700
	1,563	33.481	6.412				
% Dev. Eng	539	14.046	5.264	1.660	0.098	0.048	8.300
	1,563	14.280	4.441				
% Dev. Math	539	27.006	12.278	4.150	0.000***	0.015	20.400
	1,563	27.177	11.123				
Average Bias							13.000
R-Squared							0.061

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < 0.05 **p < 0.01 ***p < 0.001*

Table 24: Model III—Number in Study and Comparison Groups, Group Means, Standard Deviation, T-tests, *p* values, Cohen's d, and Standardized Bias of Covariates After Propensity Matching

Note. First line of each variable is the comparison group; second line of each variable is the study group.

Covariate	*N*	*M*	*SD*	*t*	*p*	*d*	% bias
Gender	522	0.529	0.500	-0.678	0.498	-0.050	-5.100
	518	0.504	0.500				
Age	522	21.521	6.997	-0.253	0.800	-0.016	-1.300
	518	21.413	6.756				
White	522	0.510	0.500	-0.246	0.805	-0.016	-2.800
	518	0.502	0.500				
Black	522	0.305	0.461	-0.188	0.851	-0.013	1.200
	518	0.299	0.458				
Hispanic	522	0.040	0.210	0.318	0.751	0.047	1.500
	518	0.050	0.219				
Transfer	522	0.460	0.499	-0.010	0.992	-0.002	-0.300
	518	0.459	0.498				
Pell-awarded	522	0.345	0.476	0.155	0.877	0.008	3.000
	518	0.349	0.477				
Full-Time	522	0.653	0.476	-0.091	0.928	-0.004	-1.300
	518	0.651	0.477				
First Dev. Math	522	1.987	0.737	0.253	0.800	0.015	1.900
	518	1.998	0.729				

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 24 (continued)

Covariate	N	M	SD	*t*	*p*	*d*	% bias
Retained to 2nd	522	0.897	0.305	-0.144	0.886	-0.010	-2.100
	518	0.894	0.308				
Dev. Eng	522	0.458	0.499	0.052	0.959	0.002	1.700
	518	0.459	0.499				
Success	522	0.354	0.479	0.930	0.353	0.058	4.900
	518	0.382	0.486				
1st Term GPA	522	2.112	1.464	-0.710	0.478	-0.044	-7.800
	518	2.045	1.552				
% Minority	522	40.996	13.642	-0.391	0.696	-0.024	-3.500
	518	40.664	13.783				
% Pell-awarded	522	33.397	7.359	0.046	0.963	0.003	0.400
	518	33.417	6.970				
% Dev. Eng	522	20.953	5.193	-1.445	0.149	-0.090	-2.800
	518	20.487	5.199				
% Dev. Math	522	36.413	12.217	-1.990	0.047*	-0.123	-5.100
	518	34.927	11.856				
Average Bias							2.800
R-Squared							0.050

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Model IV

To further explore the multilevel analysis in this study, Model IV considered the effect of institutional membership as a covariate. The previous institutional predictors dropped out of the model, due to issues of multi-collinearity. The intent of the model was to capture institutional variation as a construct in explaining average success in college-level math within institutions. Model IV weighted the contribution of each institution and revealed some differences among students within the institutions. Prior to matching, statistically significant differences among students within the institutions were reflected in the t-test analyses. Using College A as the reference group, t-tests yielded differences in College B, $t(2,100)=2.41$, $p< .05$; College F, $t(2,100)= 5.74$, $p< .001$; and College G, $t(2,100)= -4.69$, $p< .001$. The average bias was indicative of the significant differences within the colleges. Using College A as the reference group, the average bias was 12.1 % for College B, 31.8% for College F, and -21.7% for College G.

After matching, the bias was reduced considerably from 13.4% prior to matching to 3.3% after matching, a reduction of 75.4%.. Table 25 and Table 26 show the results from Model IV prior to and after propensity matching, respectively.

Table 25: Model IV— Number in Study and Comparison Groups, Group Means, Standard Deviation, T-tests, *p* values, Cohen's d and Standardized Bias of Covariates Prior to Propensity Matching

Note. First line of each variable is comparison group; second line of each variable is the study group.

Covariate	*N*	*M*	*SD*	*t*	*p*	*d*	% bias
Gender	539	0.516	0.500	2.687	0.007**	0.133	13.400
	1,563	0.582	0.493				
Age	539	21.453	6.909	1.840	0.066	0.093	9.400
	1,563	22.141	7.705				
White	539	0.497	0.500	4.370	0.000***	0.218	21.700
	1,563	0.605	0.489				
Black	539	0.319	0.200	-6.520	0.000	-0.309	-31.100
	1,563	0.186	0.009				
Hispanic	539	0.046	0.211	-0.470	0.636	-0.019	-2.300
	1,563	0.042	0.199				
Transfer	539	0.466	0.499	-1.100	0.270	-0.056	-5.500
	1,563	0.438	0.496				
Pell-awarded	539	0.359	0.480	-3.530	0.000***	-0.172	-17.300
	1,563	0.280	0.449				
Full-Time	539	0.657	0.475	-0.040	0.967	-0.002	-0.200
	1,563	0.656	0.475				
First Dev. Math	539	1.989	0.736	-0.400	0.692	-0.027	-2.000
	1,563	1.974	0.730				
Retained to 2nd	539	0.889	0.314	2.530	0.120	0.120	12.100
	1,563	0.924	0.265				

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p <.01 ***p < .001*

Table 25 (continued)

Covariate	*N*	*M*	*SD*	*t*	*p*	*d*	% bias
Dev. Eng.	539	0.471	0.499	-4.140	0.000***	-0.206	-20.500
	1,563	0.370	0.483				
Success	539	0.365	0.482	-0.460	0.645	-0.023	-2.300
	1,563	0.354	0.478				
1st Term GPA	539	2.07	1.473	7.750	0.000***	0.382	38.300
	1,563	2.62	1.406				
College B	539	0.345	0.475	2.410	0.016*	0.122	12.100
	1,563	0.404	0.490				
College C	539	0.048	0.215	-1.280	0.200	-0.064	-6.200
	1,563	0.035	0.186				
College D	539	0.007	0.086	0.810	0.421	0.051	4.200
	1,563	0.012	0.107				
College E	539	0.041	0.198	-0.510	0.611	0.024	2.600
	1,563	0.046	0.210				
College F	539	0.043	0.202	5.740	0.000***	0.320	31.800
	1,563	0.131	0.338				
College G	539	0.134	0.341	-4.690	0.000***	-0.219	-21.700
	1,563	0.068	0.253				
Average Bias							13.400
R-Squared							0.072

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 26: Model IV—Number in Study and Comparison Groups, Group Means, Standard Deviation, T-tests, *p* values, Cohen's d and Standardized Bias of Covariates After Propensity Matching

Note. First line of each variable is comparison group; second line of each variable is the study group.

Covariate	*N*	*M*	*SD*	*t*	*p*	*d*	% bias
Gender	512	0.525	0.500	-0.527	0.598	-0.032	-3.700
	507	0.509	0.500				
Age	512	21.529	6.992	-0.750	0.453	-0.047	-4.800
	507	21.219	6.189				
White	512	0.510	0.500	-0.217	0.828	-0.014	-3.300
	507	0.503	0.500				
Black	512	0.297	0.457	0.307	0.759	0.020	5.300
	507	0.306	0.461				
Hispanic	512	0.048	0.215	-0.728	0.467	-0.044	-4.400
	507	0.039	0.195				
Transfer	512	0.465	0.499	-0.231	0.817	-0.014	-1.000
	507	0.458	0.499				
Pell-awarded	512	0.346	0.476	0.835	0.404	0.052	7.200
	507	0.371	0.483				
Full-Time	512	0.656	0.475	0.351	0.725	0.023	2.600
	507	0.667	0.472				
First Dev. Math	512	1.984	0.735	0.256	0.798	0.016	1.400
	507	1.996	0.722				

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 26 (continued)

Covariate	N	M	SD	*t*	*p*	*d*	% bias
Retained to 2nd	512	0.901	0.300	-0.156	0.875	-0.013	-1.200
	507	0.897	0.304				
Dev. Eng	512	0.453	0.498	0.206	0.837	-0.012	3.900
	507	0.459	0.499				
Success	512	0.359	0.480	-0.342	0.732	-0.021	-0.900
	507	0.349	0.477				
1st Term GPA	512	2.141	1.460	-0.609	0.543	-0.038	-7.200
	507	2.084	1.556				
College B	512	0.359	0.480	-0.541	0.589	-0.034	-4.200
	507	0.343	0.475				
College C	512	0.051	0.220	-0.556	0.578	-0.038	-4.600
	507	0.043	0.204				
College D	512	0.008	0.088	-0.070	-0.366	-0.024	-2.200
	507	0.006	0.077				
College E	512	0.043	0.203	-0.123	0.902	-0.010	-1.500
	507	0.041	0.199				
College F	512	0.049	0.207	-0.597	0.550	-0.060	-3.200
	507	0.037	0.190				
College G	512	0.121	0.326	0.430	0.667	-0.028	0.400
	507	0.112	0.316				

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 26 (continued)

Covariate	N	M	SD	*t*	*p*	*d*	% bias
Average Bias 3.300							
R-Squared 0.063							

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Five-Year Outcomes

Model I

After matching, five-year outcomes were analyzed and reported for each group and model in order to determine the effect of passing college-level math in terms of credits completed in developmental and college-level courses. The average number of terms retained between the two groups, students completing college-level math and students not completing college-level math, was also examined.

Table 27 (Model I) shows that the outcomes of students in the cohort that completed college-level math were significantly different in developmental credits completed t(1,069)=2.458, $p < .001$ and developmental credits completed with a grade of C or better, t(1,069)=2.850, $p < .001$ with effect sizes of 0.150 and 0.174 respectively.

Since, in theory, all of the students completed developmental math credits, the groups were examined to identify differences between the two groups in this area. Results revealed the study group completed an average of 6.914 credits compared to 5.914 credits in

developmental math. Also, in regards to developmental math credits, the students were significantly different in terms of credits completed, t(1,069)=4.302, $p < .001$, and credits completed with a grade of C or better, t(1,069)= 4.804, $p < .001$.

Furthermore, the students in the study group fared better in terms of college-level credits completed than the comparison group at significantly high levels. The mean of college-level credits was significantly different in all three areas examined: credits attempted, t(1,069)=10.978, $p < .001$; credits completed, t(1,069)=(18.639), $p < .001$; and credits completed with grades of C or better, t(1,069)=20.534, $p < .001$. Effect sizes were considerably high in college-level credit outcomes, 0.679, 1.139, and 1.244, respectively. On average, students in the study group earned 26 more college credits than students in the comparison group.

In terms of college-level math outcomes, attempts, completions, and credits earned with grades of C or better were significantly different between the two groups; the findings also revealed a notable difference in college-level math credits attempted. The study group attempted, on average, 6.075 college-level credits, compared to an average of 4.873 college-level credits for the comparison group. The higher average number of credits attempted in the study group indicates that students in the study group attempted more than one college-level math course. Each course typically awards three credit hours. The other significant differences in college-level math courses were expected and defined the two groups in this study: those completing college-level math, and those not completing at least one college-level math course with grade of C or better.

The average number of terms retained was also significantly different between the two groups, t(1,069) =(10.114), $p < .001$, although the averages between the two groups were similar, 7.775 compared to 6.132. Both student groups were retained at the college at least six terms on average.

Table 28 shows outcomes for the two groups after matching, where applicable to course enrollments. Outcomes in the developmental math courses were significantly different between the two groups after matching in three of the courses examined. Specifically, the analysis showed students in the study group were more likely to earn a grade of C or better in MAT-080, t(663)=2.349, $p < .05$; MAT-070, t((712)=2.420, $p < .05$; and MAT-060, t(254)=2.438, $p < .05$. Compared to the comparison group, students in the study group also completed more college English credits, t(932)=(9.204), $p < .001$, and earned more grades of C or better, t(932)=(10.376), $p < .001$, with effect sizes of 0.672 and 0.679 respectively.

Table 27: Model I—Five-Year Credits and Outcomes of Completers & Non-Completers in College Math after Propensity Matching Number in Comparison and Study Groups, Group Means, Standard Deviation, T-tests, *p* values and Cohen's *d* of Outcomes After Propensity Matching

Note. First line of each variable is comparison group; second line of each variable is the study group.

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Developmental Courses						
Credits Attempted	537	13.891	9.242	-1.124	0.261	-0.069
	534	13.277	8.641			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 27 (continued)

Outcomes	N	M	SD	*t*	*p*	*d*
Credits Completed	537	10.231	7.207	2.458	0.014*	0.150
	534	11.346	7.640			
Credits A-C	537	9.561	6.886	2.850	0.005**	0.174
	534	10.807	7.419			
Developmental Math						
Credits Attempted	537	8.171	4.655	0.301	0.764	0.018
	534	8.254	4.403			
Credits Completed	537	5.914	3.785	4.302	0.000***	0.263
	534	6.914	3.818			
Credits A-C	537	5.415	3.476	4.804	0.000***	0.294
	534	6.464	3.668			
College Courses						
Credits Attempted	537	50.911	27.142	10.978	0.000***	0.671
	534	68.424	25.020			
Credits Completed	537	31.108	23.654	18.639	0.000***	1.139
	534	57.510	22.691			
Credits A-C	537	26.343	21.447	20.354	0.000***	1.244
	534	53.454	22.140			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 27 (continued)

Outcomes	N	M	SD	*t*	*p*	*d*
College Math						
Credits Attempted	537	4.873	2.948	5.407	0.000***	0.330
	534	6.075	4.217			
Credits Completed	537	0.963	1.645	26.426	0.000***	1.614
	534	4.811	2.943			
Credits A-C	537	0.000	0.000	42.033	0.000***	2.565
	534	4.360	2.403			
Retention						
Terms Retained	537	6.132	2.773	10.114	0.000***	0.627
	534	7.775	2.454			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 28: Model I—Additional Five-Year Credits and Outcomes of Completers & Non-Completers in College Math after Propensity Matching Number in Comparison and Study Groups, Group Means, Standard Deviation, t-tests, *p* values, and Cohen's *d* of Enrollment Outcomes After Propensity Matching

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Developmental Math						
Credits Attempted- 080	333	5.273	2.427	-3.214	0.001	-0.249
	332	4.747	1.737			
Credits Completed- 080	333	3.327	1.844	2.480	0.013	0.193
	332	3.651	1.499			
Credits A-C- 080	333	3.012	1.754	2.349	0.019	0.184
	332	3.313	1.511			
Credits Attempted- 070	357	5.177	2.265	-3.654	0.000	-0.274
	357	4.627	1.712			
Credits Completed- 070	357	3.899	1.250	0.675	0.499	0.050
	357	3.955	0.947			
Credits A-C- 070	357	3.585	1.257	2.420	0.016	0.181
	357	3.787	0.948			
Credits Attempted- 060	131	4.366	1.354	-0.479	0.633	-0.060
	125	4.288	1.262			
Credits Completed- 060	131	3.664	1.407	1.636	0.103	0.205
	125	3.936	1.243			
Credits A-C- 060	131	3.389	1.444	2.438	0.016	0.306
	125	3.776	1.054			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p <.001*

Table 28 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits Attempted- 050	43	4.093	0.610	1.827	0.071	0.396
	40	4.600	1.707			
Credits Completed- 050	43	3.810	0.852	0.784	0.435	0.175
	40	4.000	1.281			
Credits A-C- 050	43	3.534	1.297	0.236	0.814	0.053
	40	3.600	1.215			
Developmental English						
Credits Attempted	246	6.707	3.852	-1.215	0.225	-0.110
	240	6.329	2.934			
Credits Completed	246	4.919	2.616	3.729	0.000	0.338
	240	5.758	2.336			
Credits A-C	246	4.756	2.602	3.683	0.000	0.334
	240	5.596	2.418			
Developmental Reading						
Credits Attempted	210	6.019	3.087	0.246	0.806	0.024
	209	6.091	2.888			
Credits Completed	210	4.857	2.615	2.440	0.015	0.238
	209	5.454	2.392			
Credits A-C	210	4.648	2.690	2.749	0.006	0.268
	209	5.339	2.458			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 28 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
College English						
Credits Attempted	467	7.071	3.336	0.774	0.439	0.050
	467	7.223	2.719			
Credits Completed	467	4.441	2.739	9.204	0.000	0.602
	467	5.983	2.367			
Credits A-C	467	3.976	2.651	10.376	0.000	0.679
	467	5.668	2.321			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Model II

Table 29 (Model II) utilized demographics and student progress variables, and shows that the outcomes of students in the cohort who completed college-level math were significantly different in developmental credits completed, t(1,058)=2.443, $p < .05$, and developmental credits completed with a grade of C or better, t(1,058)=2.696 , $p < .01$, with effect sizes of 0.150 and 0.166 respectively.

Since, in theory, all of the students completed developmental math credits, the groups were examined to identify differences between the two groups in this area. Results revealed the study group completed an average of 5.899 credits compared to 5.431 credits in developmental math. Also, in regards to developmental math credits, the students were significantly different in terms of credits completed with a grade of C or better, t(1,058)= 2.210, $p < .05$.

Furthermore, the students in the study group fared better in terms of college-level credits completed than the non-completer group at significantly high levels. The mean of college-level credits was significantly different in all three areas examined: credits attempted, t(1,058)=9.281, $p < .001$; credits completed, t(1,058)=15.731, $p < .001$; and credits completed with grades of C or better, t(1,058)=17.110, $p < .001$. Effect sizes were considerably high in college-level credit outcomes, 0.570, 0.966, and 1.502 respectively. On average, students in the study group earned 25 more college credits than students in the comparison group.

In terms of college-level math outcomes, attempts, completions, and credits earned with grades of C or better were significantly different between the two groups and also revealed a notable difference in college-level math credits attempted. The study group attempted, on average, 6.278 college-level credits compared to an average of 4.876 college-level credits for the comparison group. The higher average number of credits attempted in the study group indicates that students in the study group attempted more than one college-level math course. Each course typically awards three credit hours. The other significant differences in college-level math courses were expected and defined the two groups in this study: those completing college-level math and those not completing at least one college-level math course with grade of C or better.

The average number of terms retained was also significantly different between the two groups, $t(1,058) = 8.825$, $p < .001$, although the averages between the two groups were similar, 7.588 compared to 6.166. Both student groups were retained at the college at least six terms on average.

Table 30 shows outcomes for the two groups after matching, where applicable to course enrollments. Outcomes in the developmental math courses were similar between the two groups after matching in four of the developmental courses examined. However credits earned with a C or better were significantly different in MAT-070, $t(710)=2.472$, $p <.05$. Students in the study group completed more developmental English credits, $t(490)=(2.120)$, $p < .05$ and earned more grades of C or better, $t(490)=(2.173)$, $p < .05$, compared to the comparison group with effect sizes of 0.191, 0.196 respectively. Students in the study group

also completed more college English credits, t(918)=8.199, $p < .001$, and earned more grades of C or better, t(918)=8.814, $p < .001$.

Table 29: Model II— Five-Year Credits and Outcomes of Completers & Non-Completers in College Math after Propensity Matching Number in Comparison and Study Groups, Group Means, Standard Deviation, T-tests, *p* values, and Cohen's *d* of Outcomes After Propensity Matching

Note. First line of each variable is comparison group; second line of each variable is the study group.

Outcome	*N*	*M*	*SD*	*t*	*p*	*d*
Developmental Credits Attempte	531	13.785	9.164	-0.736	0.462	-0.045
	529	13.386	8.514			
Credits Complete	531	10.198	7.127	2.443	0.015*	0.150
	529	11.297	7.515			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 29 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits A-C	531	9.561	6.870	2.696	0.007**	0.166
	529	10.72	7.126			
Developmental Math						
Credits Attempted	531	8.160	4.633	-1.720	0.086	-0.104
	529	7.699	4.182			
Credits Completed	531	5.913	3.737	1.902	0.058	0.117
	529	6.346	3.668			
Credits A-C	531	5.431	3.466	2.210	0.027*	0.136
	529	5.899	3.437			
College Courses						
Credits Attempted	531	51.264	27.088	9.281	0.000***	0.570
	529	66.51	26.389			
Credits Completed	531	31.476	23.669	15.731	0.000***	0.966
	529	54.456	23.889			
Credits A-C	531	16.665	21.494	17.110	0.000***	1.502
	529	49.973	22.842			
College Math						
Credits Attempted	531	4.876	2.955	6.096	0.000***	0.374
	529	6.278	4.397			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 29 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits Completed	531	0.963	1.652	25.833	0.000***	1.540
	529	4.811	3.123			
Credits A-C	531	0.000	0.000	41.347	0.000***	2.494
	529	4.360	2.473			
Retention						
Terms Retained	531	6.156	2.771	8.825	0.000***	0.542
	529	7.588	2.503			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p <.05 **p < .01 ***p < .001*

Table 30: Model II—Additional Five-Year Credits and Outcomes of Completers & Non-Completers in College Math after Propensity Matching Number in Comparison and Study Groups, Group Means, Standard Deviation, T-tests, *p* values, and Cohen's *d* of Enrollment Outcomes After Propensity Matching

Note. First line of each variable is comparison group; second line of each variable is the study group.

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Developmental Math						
Credits Attempted- 080	326	5.240	2.366	-1.299	0.195	-0.102
	324	5.012	2.079			
Credits Completed- 080	326	3.325	1.826	2.504	0.013*	0.197
	324	3.667	1.647			
Credits A-C- 080	326	3.031	1.745	1.746	0.081	0.137
	324	3.259	1.588			
Credits Attempted- 070	355	5.138	2.253	-2.737	0.006**	-0.205
	357	4.717	1.830			
Credits Completed- 070	355	3.899	1.217	1.138	0.256	0.085
	357	4.000	1.161			
Credits A-C- 070	355	3.606	1.231	2.472	0.014*	0.184
	357	3.809	0.953			
Credits Attempted- 060	126	4.381	1.379	-0.826	0.410	-0.104
	125	4.256	0.983			
Credits Completed- 060	126	3.651	1.433	1.748	0.082	0.221
	125	3.936	1.134			
Credits A-C- 060	126	3.365	1.468	1.491	0.137	0.188
	125	3.616	1.183			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 30 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits Attempted- 050	41	4.098	0.625	1.768	0.081	0.391
	40	4.600	1.707			
Credits Completed- 050	41	3.800	0.872	0.363	0.718	0.084
	40	3.900	1.429			
Credits A-C- 050	41	3.510	1.325	-0.041	0.967	-0.008
	40	3.500	1.340			
Developmental English						
Credits Attempted	247	6.632	3.791	1.712	0.088	-0.155
	245	6.106	2.961			
Credits Completed	247	4.943	2.594	2.120	0.035*	0.191
	245	5.416	2.348			
Credits A-C	247	4.806	2.600	2.173	0.030*	0.196
	245	5.282	2.245			
Developmental Reading						
Credits Attempted	209	5.990	3.092	-1.334	0.183	-0.131
	206	5.612	2.673			
Credits Completed	209	4.861	2.592	0.847	0.398	0.083
	206	5.068	2.376			
Credits A-C	209	4.670	2.677	1.313	0.190	0.129
	206	4.990	2.276			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 30 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
College English						
Credits Attempted	461	7.085	3.344	1.504	0.133	0.099
	459	7.401	3.028			
Credits Completed	461	4.473	2.730	8.199	0.000***	0.537
	459	5.890	2.545			
Credits A-C	461	4.015	2.641	8.814	0.000***	0.581
	459	5.495	2.445			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Model III

Table 31 (Model III) utilized demographics, student progress, and institutional level variables. The outcomes of students in the cohort that completed college-level math were significantly different in developmental credits completed, t(1,038)=2.292, $p < .05$, and developmental credits completed with a grade of C or better, t(1,038)=2.322 , $p < .01$ with effect sizes of 0.150 and 0.166, respectively.

Since, in theory, all of the students completed developmental math credits, the groups were examined to identify differences between the two groups in this area. Results revealed the study group completed an average of 5.979 credits compared to 5.464 credits in developmental math. Also, in regards to developmental math credits, the students were

significantly different in terms of credits completed with a grade of C or better, t(1,038)= 2.350, $p < .05$.

Furthermore, the students in the study group fared better in terms of college-level credits completed than the non-completer group at significantly high levels. The mean of college-level credits was significantly different in all three areas examined: credits attempted, t(1,038)=10.068, $p < .001$; credits completed, t(1,038)=16.463, $p < .001$; and credits completed with grades of C or better, t(1,038)=17.998, $p < .001$. Effect sizes were considerably high in college-level credit outcomes, 0.624, 1.021, and 1.116 respectively. On average, students in the study group earned 25 more college credits than students in the comparison group.

In terms of college-level math outcomes, attempts, completions, and credits earned with grades of C or better were significantly different between the two groups, and also revealed a notable difference in college-level math credits attempted. The study group attempted, on average, 6.475 college-level credits compared to an average of 4.851 college-level credits for the comparison group. The higher average number of credits attempted in the study group indicates that students in the study group attempted more than one college-level math course. Each course typically awards three credit hours. The other significant differences in college-level math courses were expected and defined the two groups in this study: those completing college-level math, and those not completing at least one college-level math course with grade of C or better.

The average number of terms retained was also significantly different between the two groups, t(1,038) =7.569, $p < .001$, although the averages between the two groups were similar, 7.569 compared to 6.153. Both student groups were retained at the college at least six terms on average.

Table 32 shows outcomes for the study and comparison groups after matching where applicable to course enrollments. Outcomes in the developmental math courses were similar between the two groups after matching in four of the developmental courses examined. Students in the study group completed more developmental English credits, t(464)=3.370, $p < .001$, and earned more grades of C or better, t(464)=3.382, $p < .001$ compared to the comparison group with effect sizes of 0.313 and 0.314, respectively. Students in the study group also completed more college English credits, t(912)=8.927, $p < .001$ and earned more grades of C or better, t(912)=10.137, $p < .001$.

Table 31: Model III— Five-Year Credits and Outcomes of Completers & Non-Completers in College Math after Propensity Matching Number in Comparison and Study Groups, Group Means, Standard Deviation, t-tests, *p* values, and Cohen's *d* of Outcomes After Propensity Matching

Note. First line of each variable is comparison group; second line of each variable is the study group.

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Developmental Courses						
Credits Attempted	522	13.659	9.104	-0.703	0.483	-0.044
	518	13.278	8.365			
Credits Completed	522	10.174	7.094	2.292	0.022*	0.142
	518	11.208	7.454			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 31 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits A-C	522	9.534	6.774	2.322	0.020*	0.144
	518	10.542	7.219			
Developmental Math						
Credits Attempted	522	8.16	4.663	-0.612	0.541	-0.038
	518	7.986	4.532			
Credits Completed	522	5.962	3.777	2.348	0.019*	0.145
	518	6.511	3.777			
Credits A-C	522	5.464	3.457	2.350	0.019*	0.146
	518	5.979	3.611			
College Courses						
Credits Attempted	522	51.395	27.195	10.068	0.000***	0.624
	518	67.934	25.758			
Credits Completed	522	31.596	23.789	16.463	0.000***	1.021
	518	55.691	23.408			
Credits A-C	522	26.741	21.606	17.998	0.000***	1.116
	518	51.453	22.664			
College Math						
Credits Attempted	522	4.851	2.924	6.732	0.000***	0.417
	518	6.475	4.667			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 31 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits Completed	522	0.978	1.657	25.328	0.000***	1.569
	518	5.025	3.249			
Credits A-C	522	0	0.000	39.763	0.000***	2.461
	518	4.5019	2.587			
Retention						
Terms Retained	522	6.153	2.776	8.577	0.000***	0.532
	518	7.569	2.544			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 32: Model III— Additional Five-Year Credits and Outcomes of Completers & Non-Completers in College Math after Propensity Matching Number in Comparison and Study Groups, Group Means, Standard Deviation, t-tests, *p* values and Cohen's *d* of Enrollment Outcome

Note. First line of each variable is comparison group; second line is the study group.

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Developmental Math						
Credits Attempted-080	324	5.259	2.417	-1.608	0.108	-0.127
	321	4.972	2.110			
Credits Completed-080	324	3.358	1.831	2.303	0.022*	0.181
	321	3.676	1.672			
Credits A-C- 080	324	3.049	1.734	1.848	0.065	0.146
	321	3.290	1.563			
Credits Attempted-070	350	5.120	2.189	-3.733	0.000***	-0.282
	349	4.573	1.645			
Credits Completed-070	350	3.931	1.209	0.926	0.355	0.070
	349	4.011	1.072			
Credits A-C- 070	350	3.611	1.224	1.749	0.081	0.132
	349	3.759	1.000			
Credits Attempted-060	118	4.373	1.382	-0.839	0.403	-0.110
	118	4.237	1.084			
Credits Completed-060	118	3.661	1.439	1.245	0.214	0.162
	118	3.864	1.037			
Credits A-C- 060	118	3.390	1.444	1.336	0.183	0.174
	118	3.627	1.280			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 32 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits Attempted- 050	41	4.098	0.625	1.414	0.161	0.313
	40	4.400	1.215			
Credits Completed- 050	41	4.098	0.625	0.994	0.323	0.002
	40	4.100	1.105			
Credits A-C- 050	41	3.610	1.202	1.355	0.179	0.302
	40	3.900	0.632			
Developmental English						
Credits Attempted	234	6.556	3.779	-0.603	0.547	-0.056
	232	6.367	2.939			
Credits Completed	234	4.901	2.555	3.370	0.001**	0.313
	232	5.681	2.436			
Credits A-C	234	4.756	2.558	3.382	0.001**	0.314
	232	5.539	2.433			
Developmental Reading						
Credits Attempted	198	5.890	2.925	-1.353	0.177	-0.133
	198	5.535	2.398			
Credits Completed	198	4.823	2.589	0.759	0.448	0.078
	198	5.010	2.159			
Credits A-C	198	4.647	2.656	0.987	0.324	0.099
	198	4.889	2.212			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 32 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
College English						
Credits Attempted						
	459	7.102	3.346	1.287	0.198	0.085
	455	7.374	3.013			
Credits Completed						
	459	4.486	2.734	8.927	0.000***	0.590
	455	6.020	2.459			
Credits A-C						
	459	4.012	2.654	10.137	0.000***	0.674
	455	5.705	2.365			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p <.05 **p < .01 ***p < .001*

Model IV

Table 33 (Model IV) utilized demographics, student progress, and institutional membership variables. The outcomes of students in the cohort that completed college-level math differed significantly in developmental credits completed t(1,017)=2.169, $p < .05$, and developmental credits completed with a grade of C or better, t(1,017)= 2.189, $p < .05$, with effect sizes of 0.136 and 0.137, respectively.

Since, in theory, all of the students completed developmental math credits, the groups were examined to identify differences between the two groups in this area. Results revealed the study group completed an average of 5.826 credits compared to 5.319 credits in developmental math. Also, in regards to developmental math credits, the students differed

significantly in terms of credits completed with a grade of C or better: t(1,017)= 2.019, $p < .05$.

Furthermore, the students in the study group fared better in terms of college-level credits completed than the non-completer comparison group at significantly high levels. The mean of college-level credits was significantly different in all three areas examined: credits attempted, t(1,017)=10.083, $p < .001$; credits completed, t(1,017)=16.453, $p < .001$; and credits completed with grades of C or better, t(1,017)=17.905, $p < .001$. Effect sizes were considerably high in college-level credit outcomes, 0.632, 1.031, and 1.122, respectively. On average, students in the study group earned 25 more college credits than students in the comparison group.

In terms of college-level math outcomes, attempts, completions, and credits earned with grades of C or better differed significantly between the two groups and also revealed a notable difference in college-level math credits attempted. The study group attempted, on average, 6.422 college-level credits compared to an average of 4.861 college-level credits for the comparison group. The higher average number of credits attempted in the study group indicates that students in the study group attempted more than one college-level math course. Each course typically awards three credit hours. The other significant differences in college-level math courses were expected and defined the two groups in this study, those completing college-level math and those not completing at least one college-level math course with grade of C or better.

The average number of terms retained was also significantly different between the two groups, t(1,017) =8.245, $p < .001$, although the averages between the two groups were similar, 7.524 compared to 6.139. Both student groups were retained at the college at least six terms on average.

Table 34 shows outcomes for the two groups after matching where applicable to course enrollments. Outcomes in the developmental math courses were similar between the two groups after matching in four of the developmental courses examined. Students in the study group completed more developmental English credits, t(453)=2.272, $p <.05$, and earned more grades of C or better, t(453)=2.187, $p < .05$, compared to the comparison group with effect sizes of 0.213 and 0.215, respectively. Students in the study group also completed more college English credits, t(899)=8.188, $p < .001$, and earned more grades of C or better, t(899)=9.105, $p < .001$.

Table 33: Model IV—Five-Year Credits and Outcomes of Completers & Non-Completers in College Math after Propensity Matching Number in Comparison and Study Groups, Group Means, Standard Deviation, T-tests, *p* values, and Cohen's *d* of Outcomes After Propensity Matching

Note. First line of each variable is comparison group; second line is the study group.

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Developmental Courses						
Credits Attempted	512	13.504	8.894	-0.565	0.572	-0.035
	507	13.191	8.761			
Credits Completed	512	10.080	7.133	2.169	0.030*	0.136
	507	11.087	7.673			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p <.01 ***p < .001*

Table 33 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits A-C	512	9.402	6.834	2.189	0.029*	0.137
	507	10.368	7.258			
Developmental Math						
Credits Attempted	512	8.101	4.603	-0.871	0.384	-0.054
	507	7.854	4.470			
Credits Completed	512	5.906	3.752	2.036	0.042*	0.128
	507	6.379	3.652			
Credits A-C	512	5.398	3.469	2.019	0.044*	0.126
	507	5.826	3.295			
College Courses						
Credits Attempted	512	51.259	27.090	10.083	0.000***	0.632
	507	67.852	25.406			
Credits Completed	512	31.516	23.827	16.453	0.000***	1.031
	507	55.884	23.447			
Credits A-C	512	26.750	21.654	17.905	0.000***	1.122
	507	51.525	22.511			
College Math						
Credits Attempted	512	4.865	2.919	6.304	0.000***	0.395
	507	6.422	4.756			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 33 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits Completed	512	0.978	1.649	23.984	0.000***	1.500
	507	5.045	3.461			
Credits A-C	512	0.000	0.000	37.567	0.000***	2.348
	507	4.507	2.715			
Retention						
Terms Retained	512	6.139	2.785	8.245	0.000***	0.516
	507	7.524	2.576			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 34: Model IV— Additional Five-Year Credits and Outcomes of Completers & Non-Completers in College Math after Propensity Matching Number in Comparison and Study Groups, Group Means, Standard Deviation, t-tests, *p* values, and Cohen's *d* of Enrollment Outcomes After Propensity Matching

Note. First line of each variable is the comparison group; second line is the study group.

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Developmental Math						
Credits Attempted-080	310	5.223	2.428	-1.278	0.202	-0.101
	310	4.994	2.084			
Credits Completed-080	310	3.381	1.800	1.726	0.085	0.139
	310	3.626	1.737			
Credits A-C- 080	310	3.084	1.714	0.765	0.444	0.061
	310	3.187	1.644			
Credits Attempted-070	347	5.141	2.243	-3.399	0.001**	-0.259
	342	4.632	1.642			
Credits Completed-070	347	3.920	1.233	-0.138	0.891	-0.011
	342	3.906	1.222			
Credits A-C- 070	347	3.585	1.259	1.074	0.283	0.082
	342	3.684	1.164			
Credits Attempted-060	118	4.373	1.422	0.417	0.677	-0.093
	115	4.237	1.512			
Credits Completed-060	118	3.661	1.439	1.401	0.163	0.183
	115	3.930	1.497			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 34 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits A-C- 060	118	3.356	1.476	1.062	0.290	0.139
	115	3.548	1.272			
Credits Attempted- 050	38	4.105	0.649	1.021	0.311	0.239
	38	4.320	1.093			
Credits Completed- 050	38	3.895	0.649	0.443	0.659	0.101
	38	4.000	1.315			
Credits A-C- 050	38	3.579	1.244	0.844	0.402	0.193
	38	3.789	0.905			
Developmental English						
Credits Attempted	229	6.611	3.734			-0.095
	226	6.283	3.118	-1.017	0.310	
Credits Completed	229	4.991	2.644	2.272	0.024*	0.213
	226	5.553	2.630			
Credits A-C	229	4.821	2.639	2.187	0.029*	0.205
	226	5.345	2.469			
Developmental Reading						
Credits Attempted	201	5.990	3.076	-1.053	0.293	-0.106
	197	5.685	2.681			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Table 34 (continued)

Outcomes	*N*	*M*	*SD*	*t*	*p*	*d*
Credits Completed	201	4.935	2.623	0.812	0.417	0.082
	197	5.137	2.323			
Credits A-C	201	4.716	2.706	1.013	0.312	0.102
	197	4.975	2.361			
College English						
Credits Attempted	452	7.104	3.356	1.218	0.223	0.084
	449	7.377	3.130			
Credits Completed	452	4.473	2.735	8.188	0.000***	0.546
	449	5.910	2.521			
Credits A-C	452	4.013	2.638	9.105	0.000***	0.607
	449	5.548	2.416			

Cohen's d = Mt – Mc / σpooled; where σpooled = √ σ2t + σ2c / 2
**p < .05 **p < .01 ***p < .001*

Chi-Squared Analysis

As shown in Table 35, Table 36, Table 37, and Table 38, chi-squared analyses of five-year educational outcomes after propensity matching also yielded some differences in expected and observed frequencies in the two groups. The four models were significant at $p < .001$. In terms of expected versus observed frequencies, a higher number of students in the comparison group earned no award compared to students in the study group. Students in

the study group earned significantly more associate degrees than students in the comparison group. However, in the four models, students in the comparison group were twice as likely to earn no outcome after five years compared to the study group. Transfers to four-year and two-year institutions were common in both groups of students, and the study group was twice as likely to transfer out of the institution.

Table 35: Model I— Chi-Square Results for Educational Outcomes in Two Study Groups After Matching

Outcome Categories		Passed College-Level Math Grade of C or Better	Not Passed College-Level Math	Total
No Outcome	Observed	135	336	471
	Expected	234.8	236.2	471.0
Associate	Observed	84	24	108
	Expected	53.8	54.2	108.0
Transfer 2 Year	Observed	49	17	66
	Expected	32.9	33.1	66.0
Transfer 4 Year	Observed	180	79	259
	Expected	129.1	129.9	259.0
Still Enrolled 30 Plus Credits	Observed	86	81	167
	Expected	83.3.	83.7	167.0
Total				1071

Alpha = 0.050
Chi-Squared = 174.154
Degrees of Freedom = 4
p value < .001
Cramer's V = 0.403

Table 36: Model II— Chi-Square Results for Educational Outcomes in Two Study Groups After Matching

Outcome Categories		Passed College-Level Math Grade of C or Better	Not Passed College-Level Math	Total
No Outcome	Observed	152	329	481
	Expected	240.0	241.0	481.0
Associate	Observed	76	24	100
	Expected	49.9	50.1	100.0
Transfer 2 Year	Observed	48	18	66
	Expected	32.9	33.1	66.0
Transfer 4 Year	Observed	159	79	238
	Expected	129.1	129.9	238.0
Still Enrolled 30 Plus Credits	Observed	94	81	175
	Expected	87.3.	87.7	175.0
Total				1060

Alpha = 0.050
Chi-Squared = 133.663
Degrees of Freedom = 4
p value < .001
Cramer's V = 0.355

Table 37: Model III— Chi-Square Results for Educational Outcomes in Two Study Groups After Matching

Outcome Categories		Passed College-Level Math Grade of C or Better	Not Passed College-Level Math	Total
No Outcome	Observed	131	325	456
	Expected	227.1	228.9	456.0
Associate	Observed	84	24	108
	Expected	53.8	54.2	108.0
Transfer 2 Year	Observed	52	18	70
	Expected	34.9	35.1	70.0
Transfer 4 Year	Observed	169	76	245
	Expected	122.0	123.0	245.0
Still Enrolled 30 Plus Credits	Observed	82	79	161
	Expected	80.2	80.8	161.0
Total				1040

Alpha= 0.050
Chi-Squared= 167.728
Degrees of Freedom= 4
p value < .001
Cramer's V= 0.402

Table 38: Model IV— Chi-Square Results for Educational Outcomes in Two Study Groups After Matching

Outcome Categories		Passed College-Level Math Grade of C or Better	Not Passed College-Level Math	Total
No Outcome	Observed	133	322	455
	Expected	226.4	228.6	455.0
Associate	Observed	81	23	104
	Expected	51.7	52.3	104.0
Transfer 2 Year	Observed	44	17	61
	Expected	30.4	30.6	61.0
Transfer 4 Year	Observed	177	74	251
	Expected	124.9	126.1	251.0
Still Enrolled 30 Plus Credits	Observed	72	76	148
	Expected	73.6	74.4	148.0
Total				1019
Alpha= 0.050 Chi-Squared= 165.159 Degrees of Freedom= 4 *p* value < .001 Cramer's V= 0.403				

Summary

Chapter IV presented the results of the study utilizing four iterative models of student retention. The first model**,** Model I, captured student background characteristics. Model II expanded on student background characteristics and added student progress or first-term indicators to the model. Models III and IV added a layer of data to the model and multilevel analyses were examined. Each model was systematically reviewed and analyzed by the

researcher. Logistic regression analyses, both before and after matching results and including covariates, yielded differences in short-term and long-term outcomes.

The final chapter discusses the major findings with an analysis of the research questions as the guiding framework. The chapter concludes with a discussion of study limitations, suggestions of future research studies, and theoretical and practical applications of the current study.

CHAPTER V: DISCUSSION

While there are varying definitions of student success in the community college (Wild & Ebbers, 2002), recent attention has been focused on persistence and retention of students in terms of certificate or degree completion and/or successful transfer to a four-year institution (Pretlow & Wathington, 2006). In a community college context, students who begin work in developmental math courses are a significant proportion of students in the community college (Bailey, 2009; Bailey et al., 2010). The developmental math student's progression and retention from developmental math courses to the college-level gateway math course and subsequent success were the focus of this study.

Research Questions and Hypotheses

The study examined the following research questions and hypotheses:

1) What are the demographics and academic characteristics of the study population?

2) What are the demographics and academic characteristics of the two groups in the study?

3) Is there a difference in demographics and academic characteristics of the two study groups prior to propensity score matching?

Hypothesis 1

Ho: There is no difference between students who first enroll in developmental math and the conditional probability of completing with a grade of C or better one college-level math course.

Ha: There is a statistically significant difference in between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

4) Is there a difference in demographics and academic characteristics of the two study groups after propensity score matching?

Hypothesis 2

Ho: After propensity score matching, there is a statistically significant difference between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

Ha: There is no difference in between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

5) Is there a difference between the two study groups in college outcomes after propensity score matching?

Hypothesis 3

Ho: There is no difference in student outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

Ha: There are statistically significant outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course.

Hypothesis 4

Ho: There is no difference in student outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course between institutions.

Ha: There are statistically significant different outcomes between students who first enroll in developmental math and the conditional probability of completing, with a grade of C or better, one college-level math course between institutions.

Major Findings

Research Question 1

Out of 17,060 students in the database, 5,548 were selected for potential inclusion in the study group. The 5,548 students were first enrolled in developmental math courses and met the first criteria of the study population. The next selection criteria involved selecting students who had progressed from developmental math coursework to enrollment in at least one college-level math course at MAT-101 or higher. The initial group of 5,548 students was reduced to a final sample size of 2,102 (38%) students in seven institutions. Of the 2,102 students in the final sample, 1563 students (74%) passed at least one college-level math course with a grade of C or better. Students passing with a grade of C or better formed the study group. The control comparison group in the study was formed from 539 students (26%) who enrolled in a least one college-level math course but did not pass with a grade of C or better.

Research Question 2

In the final analytical sample, females outnumbered males and females represented 1,188 (56.52%) of the group. The majority of students were 19 years of age and under, although 21% of students entered college later at age 24 years or older. The largest racial ethnic group was White and Black students formed the next largest racial ethnic group. Of the study group, 936 students (44.53%) declared a college transfer program, while the rest of the students were enrolled in certificate, diploma, or terminal associate degree programs. Most of the students were enrolled full-time and the majority of the students returned in the second term.

The students included in the final analytical sample were more likely to have first enrolled in the highest level math courses, MAT-080 and MAT-070, which suggested that students who first enrolled in the lowest levels of developmental math were less represented in the final study group and less likely to progress to the college- level math group.

Research Question 3

Prior to propensity matching, the results indicated that students differed in terms of demographics and academic progress, but not to a large extent. Four iterations of retention models were used to examine differences in the two groups prior to creating the matched groups. In Model I, logistic regression results indicated that four predictors (Gender, Age, Black Race, Pell Recipient) were statistically reliable in predicting membership in the dependent variable. The model was significant, but little variance (4.4%) was explained in Model I. The second model included the independent variables used in Model I and

additional variables to indicate a student's progress. Logistic regression results indicated that the overall model of five predictors (Gender, Black Race, Pell Recipient, First Term GPA) were statistically reliable in predicting membership in the dependent variable. The second model was significant and explained 7.5% of the variance, again indicating the students did not differ to a large extent in terms of demographics or student progression indicators.

In the third model logistic regression was conducted again, entering the institutional level variables to determine which independent variables in Model III were significantly associated with the dependent variable— completing or not completing college-level math with a grade of C or better. Model III included the independent variables used in Model I and Model II along with institutional variables to capture institutional variation. Logistic regression results indicated that the overall model of seven predictors (Gender, Black Race, Pell Recipient, First Term GPA, Percent minority, Percent Pell, Percent developmental English) were statistically reliable in predicting membership in the dependent variable. Students in institutions with higher percentages of minorities were slightly (5%) more likely to pass college-level math with a grade of C or better. Minority concentration yielded a small positive effect and supported racial concentration effects examined in previous research (Bahr, 2010b). Students in institutions with higher percentages of Pell recipients and developmental English enrollments were slightly less likely to pass college-level math, 5% and 10%, respectively. Martorell & McFarlin Jr. (2007) found that high remedial institutions were less effective in developmental education for students in Texas, and this study supported the previous findings in regards to developmental English instruction.

Logistic regression was conducted for the final iteration, entering the institutional level variables as membership variables to determine which independent variables in Model II were significantly associated with the dependent variable—completing or not completing college-level math with a grade of C or better. Model IV includes the independent variables used in Model I and II, along with institutional membership variables to capture institutional variation. Logistic regression results indicated that the overall model of seven predictors (Gender, Black Race, Pell Recipient, First Term GPA, College B, College F, College G) were statistically reliable in predicting membership in the dependent variable. With College A as the reference group, students in College B were 32% more likely to pass college-level math than students in the reference group, and students in College F were 197% more likely to pass college-level math than students in the reference group. However, students in College G were 54% less likely to pass college-level math than students in the reference group. Overall, moving from a single-level to multilevel model explained more variance in the two groups, from 4.4% to 11.6%, and suggested within institutional variance was evident in the student data.

To further analyze whether there was a difference in demographics and academic characteristics of the two study groups prior to propensity score matching, t-tests were provided and analyzed. Since the study examined retention of developmental math students and subsequent success in college-level math, four iterations of analyses were constructed to conceptualize the retention model. Each time, logistic regression was utilized to create a

propensity score, t-test analyses examined the results of before and after matching on the covariates.

The first model examined the student demographics of gender, age, race, transfer goals, enrollment status, Pell recipient, and the first developmental math course in which a student enrolled. The initial t-tests yielded significant differences in gender, $t(2,100)=2.69$, $p < .01$, and Pell recipients, $t(2,100)= -3.53$, $p < .01$. The results also indicated that students passing college-level math with a C or better differed significantly for White $t(2,100)= 4.37$, $p < .01$, and Black students, $t(2,100)=-6.52$, $p < .01$, compared to the other race category. Before matching, the absolute values in bias ranged from 3% to 31%. The average bias before matching was 10.4%. After matching, the absolute value in bias ranged from 3.2% to 23%, and average bias remained high at 10.3%. In short, the model did not achieve good balance.

Model II was the second model examined and included the remaining individual student predictors. The model retained the predictors from Model I, and added academic integration variables that in theory predict student retention. The academic progress variables that were added were first-term grade point average and enrollment in developmental English developmental math, student success coursework and retained to second major term. T-tests analysis indicated that gender remained significant in Model II and before matching the bias among the two groups of students was 13%. Black and White race also remained significant. Prior to matching, the bias for White students was 21% and for Black students the bias was -31.1%. A negative bias indicated that the mean for the comparison group was higher than the

mean for the study group or that Black students were less likely to complete college- level math with a grade of C or better. Pell recipients in the first term were also negatively biased at -17.3%, indicating that Pell recipients were less likely to complete college-level math compared to non-Pell recipients. The mean difference in first-term grade point average was also significant and the bias prior to matching was positive and high, at 38.3%. Students in the study group earned an average first-term grade point average of 2.62 compared to 2.07 for students in the comparison group. The average bias was 12.5% prior to matching and six covariates yielded bias estimators greater than the recommended threshold of less than or equal to 5%.

After matching, the reduction in bias was 66.4%, from an initial average bias of 12.5% to 4.2% after matching. Since the goal of propensity matching was to create two equivalent groups on average, Model II yielded a stronger matched group of students for analyses. Gender, Black, and Hispanic race were biased predictors that failed to meet the threshold of less than 5% after matching. However, there was a large reduction in bias in Black race, from -31% prior to matching to 5.9% after matching. After matching, the comparison group had a slightly higher mean than the study group, but the two groups were balanced and not significantly different in terms of retention to the second major term.

Model III expanded the retention model of student-level predictors to include institutional-level predictors. College indicators included percent minority and percent Pell awarded students as a percentage of fall 2007 official enrollments. Remedial costs were

examined by comparing the percent of institutional budget FTE as an indicator of costs associated with developmental math and English courses at the various institutions.

T-test analyses indicated that, prior to matching, percent minority was significantly different among the two groups, t=(2.85), $p < .001$ and a positive biased predictor of success in college-level math of 13.7%. In addition, percent developmental math was also a positive predictor with a high level of bias of 20.4%. T-test analyses indicated that, prior to matching, the two groups differed significantly, t=(4.15), $p < .001$. In addition, prior to matching, the average amount of bias was 13.0%.

After matching, the average bias was 3.2%. However, t-tests revealed significant differences in four student-level predictors: Age, White race, Pell-awarded students, and first-term grade point average. Overall, effect sizes were small and the reduction in bias was 78.5%, from 13.0% prior to matching to 2.8% after matching.

To further explore the multilevel analysis in this study, a fourth model considered the effect of institutional membership as a covariate in the model. The previous institutional predictors dropped out of the model due to issues of multi-collinearity. The intent of the model was to capture institutional variation as a construct in explaining average success in college-level math within institutions. The fourth model weighted the contribution of each institution and revealed some differences among students within the institutions. Prior to matching, statistically significant differences among students within the institutions were reflected in the t-test analyses. The average bias was indicative of the significant differences

within the colleges. Using College A as the reference group, the average bias was College B (12.1%), College F (31.8%), and College G (-21.7%).

After matching, the bias was reduced considerably (75%), from 13.4% prior to matching to 3.3% after matching.

Research Question 4

After matching, five-year outcomes were analyzed and reported for each group and for each model in order to determine the effect of passing college-level math in terms of credits completed in developmental and college-level courses. The average number of terms retained between the two groups, students completing college-level math and students not completing college-level math, was also examined.

In the first model, the outcomes of students in the cohort that completed college-level math were significantly different in developmental credits completed. Since in theory, all of the students completed developmental math credits, the groups were examined to identify differences between the two groups in this area. Results revealed the study group completed an average of 6.914 credits compared to 5.914 credits in developmental math. Also, in regards to developmental math credits, the students were significantly different in terms of credits completed and credits completed with a grade of C or better. Furthermore, the students in the study group fared better in terms of college-level credits completed than the non-completer group at significantly high levels. On average, students in the study group earned 26 more college credits than students in the comparison group.

In terms of college-level math outcomes, attempts, completions, and credits earned with grades of C or better were significantly different between the two groups and also revealed a notable difference in college-level math credits attempted. The average number of terms retained was also significantly different between the two groups, t(1,069) =(10.114), $p < .001$, although the averages between the two groups were similar, 7.775 compared to 6.132. Both student groups were retained at the college at least six terms on average.

Model II utilized demographics and student progress variables. As in Model I, the outcomes of students in the cohort that completed college-level math were significantly different in developmental credits completed, t(1,058)=2.443, $p < .05$, and developmental credits completed with a grade of C or better, t(1,058)=2.696 , $p < .01$, with effect sizes of 0.150 and 0.166, respectively. On average, students in the study group earned 25 more college credits than students in the comparison group, similar to findings in Model I.

Model III utilized demographics, student progress, and institutional-level variables. The multilevel model yielded similar outcomes after propensity matching estimations. The outcomes of students in the cohort that completed college-level math were significantly different in developmental credits completed, t(1,038)=2.292, $p < .05$ and developmental credits completed with a grade of C or better, t(1,038)=2.322 , $p < .01$, with effect sizes of 0.150 and 0.166, respectively.

In terms of college-level math outcomes, attempts, completions, and credits earned with grades of C or better were significantly different between the two groups and also revealed a notable difference in college-level math credits attempted. The study group

attempted, on average, 6.475 college-level credits compared to 4.851 college-level credits for the comparison group. The average number of terms retained was also significantly different between the two groups, t(1,038) =7.569, $p < .001$, although the averages between the two groups were similar, 7.569 compared to 6.153. Both student groups were retained at the college at least six terms on average, similar to the student-level model.

Model IV utilized demographics, student progress, and institutional-level variables. The outcomes of students in the cohort that completed college-level math were significantly different in developmental credits completed, t(1,038)=2.292, $p < .05$, and developmental credits completed with a grade of C or better, t(1,038)=2.322 , $p < .01$, with effect sizes of 0.150 and 0.166, respectively. Model I, II, and III supported this finding.

On average, students in the study group earned 25 more college credits than students in the comparison group. The average number of terms retained was also significantly different between the two groups, t(1,038) =7.569, $p < .001$, although the averages between the two groups were similar, 7.569 compared to 6.153. Both student groups were retained at the college at least six terms on average. In summary, both groups of students were retained at the college, but outcomes were significantly different for students completing college-level math with a C or better, in both student and multilevel models.

Chi-squared analyses of five-year educational outcomes after propensity matching also yielded some differences in expected and observed frequencies in the two groups. The four models were significant at $p < .001$. In terms of expected versus observed frequencies, a higher number of students in the comparison group earned no award compared to students in

the study group. Completers of college level math earned significantly more associate degrees than non-completers of college level math. However, in the four models, non-completers of college level math with C or better were twice as likely to earn no outcome after five years when compared to the completers of college level math group with C or better. Transfers to four-year and two-year institutions were common in both groups of students, and the study group (completers of college level math) was twice as likely to transfer out of the institution.

Limitations and Areas for Future Research

There were four limitations to the study and areas for future research. First, no research methodology is free of bias (Bostian, 2008). One objective of propensity matching is to reduce the selection bias in intervention research when researchers compare two different groups and do not control for initial and intermediate differences between the students. There are unobserved differences that are not explained in the model. In other words, the models do not explain or account for all of the variance in the two groups (Titus, 2007). Some important variables such as marital status, number of children and high-school grade point average, variables found to be significant indicators of student success, are not available in the dataset and thus the missing data introduce the potential for unobserved differences.

Multi-level analyses introduced the potential for missing data at the aggregate level and subsequently within each institution. Missing data were sensitive to propensity score estimations and yielded, in effect, missing propensity scores and dropped cases. Care must be

taken to ensure that missing data are monitored and controlled during each use of the propensity matching technique. Future research is needed to examine sensitivity to missing data among the various techniques utilized in multilevel propensity score research (Kim & Seltzer, 2007).

Second, the propensity score method reduces the number of cases through the matching procedure. Concern has been raised that the propensity score method eliminates a good number of students who are not part of the analysis. The elimination of cases occurs during the matching process when a match is not found based on the propensity score (Caliendo & Kopeinig, 2008). Moreover, smaller institutions, with less than 300 cases were a limitation in the multilevel analyses (Caleindo & Kopeinig, 2008). In addition, propensity score matching was not able to reduce the bias in some of the covariates. While the objective of propensity matching is to reduce the bias in the two study groups, the ignorability assumption implies that bias exists in simulated experimental designs (Kim & Seltzer, 2007). Thus, future research should provide guidance regarding sample sizes and appropriate covariates that the researcher can comfortably disregard in multilevel analyses.

The researcher encountered multilevel data with a larger number of students in the study group than the comparison group. The norm in propensity matching research is for the researcher to select a small purposeful sample of cases from a large sample of potential controls (Kim & Seltzer, 2007). During the matching procedure, the researcher attempted to use different algorithms available in STATA software to match students yielding poor matches and biased results. The tradeoffs between variance and bias inherent in propensity

matching designs were best approached using a without replacement function in STATA, yielding a smaller matched study group. As such, future research should examine the extent to which the larger number of students in the study group required additional adjustments to the methodology during the propensity score estimation.

Third, timing of developmental courses and timing of the subsequent college-level course were not part of this study. The timing of each sequence of developmental math and the subsequent college math success rates were likely impacted by timing of the courses and the student's ability to schedule courses in sequence and without an interruption of the math series of courses. Frye et al. (2013) found that minority students that attempted developmental math in the first year were more likely to complete college-level math with a grade of C or better compared to students who delayed developmental math courses. Hence, future research should explore the timing of developmental math courses and college-level math in order to provide policy guidance to community colleges currently revising advising models. The advising models are under revision to ensure students are appropriately informed of the implications of course sequencing, course prerequisites, and program of study requirements.

Fourth, the research study examined initial student differences that were not explored in the research. Since one of the major objectives of the study is to create equivalent study and comparison groups using propensity matching, there was less focus on the exploration of the initial differences between the groups. Although the intent of the study was to simulate a scientific experimental design, student and academic factors associated with initial

membership in the two groups are important to explore in future research. Qualitative research examining institutional variation is needed to help explain the contextual variation that yielded different outcomes for students between institutions. Performance measures in North Carolina are based on measures of student performance among all institutions. Institutions performing at high rates in some measures could be used as peer-mentoring institutions, but research is needed to determine why some institutions are performing at higher rates than others.

Theoretical Implications and Discussion

Most of the research in remediation is converging towards the finding that remediation proficiency levels matter in terms of remediation and college success (Roksa et al., 2009). Bahr's (2009) initial findings that successful remediators, even those students faced with serious deficiencies, can succeed at comparable levels to non-remedial students contrasts with other studies including Bahr's (2010b) own subsequent study examining racial disparities the following year. For example, Roksa et al. (2009) found that students in the need of the most remediation do not succeed at comparable rates to those students with fewer remedial needs, especially in math. Bailey (2009) found, as expected, that developmental completion rates were negatively related to remediation levels in the subject areas examined, math and reading.

The research presented here indicated that students persisting through the developmental math sequence and subsequently passing college-level math succeeded in their college-level work, as reflected in grade point averages and credit hours earned with a

grade of C or better. The finding held despite differences in initial enrollment levels in the developmental math sequence. Recent research supported this view, indicating that community college students who successfully complete their developmental sequence go on to graduate or transfer to a four-year institution at comparable rates to students who began at college level (Bahr, 2010). This finding holds regardless of either the depth, the number of developmental course(s) required in one subject, or the developmental breadth, the number of subject areas. However, the majority of the study population did not start the remedial coursework at the lowest levels of developmental math.

This study defined successful remediators as students that passed the developmental sequence of courses and passed at least one college math course with a grade of C or better. The measure is similar to the North Carolina Community College System Performance Measure (NCCCS, 2013b) that examined completion rates in math as a percentage of attempters who complete with a grade of C or better. The successful group is the group of students that completed the college-level math course with a grade of C or better. The second group attempted college-level math but did not pass the course with a grade of C or better. After the two groups were identified, the student must also have been enrolled in a developmental math course in the current or previous academic year. The assumption is that the student passed the required sequence of developmental math courses and was subsequently eligible to enroll in college-level math. Initial publications and release of the performance measure data reflect variation between institutions and performance rates, ranging from 31% to 91% (NCCCS, 2013b).

This study confirmed variation between colleges and implies that administrators and policy makers need to take seriously the call to student success that is tantamount to the community college student success movement. Colleges should strive to increase the number of students that complete college-level math with a grade of C or better. The results of this study indicated that getting more students to complete college-level math with a grade of C or better can significantly increase the number of credit hours earned by 25 credit hours on average. The difference between the average number of credit hours attained by a student in the study group and the average number of credit hours attained by a student in the comparison group is large enough that enormous efforts to focus on improving this measure is likely to increase completion rates substantially.

Moreover, the classification mechanisms most commonly used by community colleges known as placement tests are under scrutiny. There is an ongoing examination of the validity of placement testing, and recent studies have revealed high school grade point averages were better predictors than placement tests in predicting success in college-level work (Belfield & Crosta, 2012; Hughes & Scott-Clayton, 2011; Scott-Clayton, 2011). As a result, policies involving multiple measures associated with placement taking into account high school grade point averages were adopted in North Carolina in 2013. The move toward multiple measures is likely to challenge the entire placement referral system and how placement operates in community colleges. In addition, multiple measure policies enacted in North Carolina should shorten the time a student spends in developmental education and also enhance the potential for increasing the rate of performance among community college

students. Regression discontinuity studies have also demonstrated that marginal students would probably do just as well as if they did not take remedial courses. Some North Carolina colleges enacted policies to simultaneously enroll marginal students in a math lab along with the college-level math course, in order to mainstream them into college-level work at a faster rate.

Taken together, researchers have shaped and influenced developmental education programming in community colleges across the country and the pieces of the remediation puzzle are coalescing. The shift to a focus on student success in community colleges is paramount to understanding the educational landscape in developmental programming. Specifically, researchers argue that far too many students enter the developmental education “gauntlet” without ever progressing to college-level work or completing a program of study. Given that some research indicates those students on the margin gain little or no benefit from remedial instruction, and that developmental education yields inconclusive benefits at best and makes no difference at worst, developmental education is under major revision to accelerate and shorten the length of time students spend in developmental education programming.

However, a group of students in need of the most remediation is likely to continue to populate the enrollments of the community college rosters. Community colleges must continue to serve these students. As Attewell et al. (2006) noted, "If children of poor and minority families disproportionately leave high school with poor academic skills, should social policy encourage colleges to redress those skill problems, or should failure at the high

school level be irreversible?" (p. 916). Researchers can continue to inform the field regarding the group of students least likely to succeed in terms of what policies, practices, or programs are working for students in need of the most remediation.

In addition, Bahr's (2010b) finding that persistence in college does not mitigate some of the factors associated with success for the students with the highest remedial need is critical to understanding the debate in developmental education that is at the crux of who will populate the community college rosters in developmental education. In general, Martorell and McFarlin's (2007) research revealed remediation rates were higher for two-year than four- year college students, remediation for math was more common, remedial students were older or late entry, economically disadvantaged, more likely to be non-White, and have much lower placement test scores than non-remedial students. This research found negative effects for Pell recipients, a proxy for economic disadvantage and racial disparities. Bahr (2010b) concluded that racial disparities reflect poorer resources in elementary school and its effect on student college readiness.

While the developmental education literature is increasing, an important question to address should regard educational research methodology as a topic of significance. Indeed, Levin and Calcagno (2008) critiqued evaluation studies in the community college and argued that ideal studies utilize an experimental design. The second choice among these researchers would be quasi-experimental designs. The researchers argue that community college research offices are under-staffed and evaluation studies are not appropriate in providing effective information. Institutional researchers need to provide information that institutions can use to

improve services to developmental students but spend most of their time generating reports for state and federal agencies. The size of the institutional research office plays a role in the type of research conducted. In most cases, staff are assigned numerous roles in an institutional research office and experimental designs are not practical (Levin & Calcagno, 2008). However, in developmental education, given the extent of costs in time and resources taxed on students, faculty, and the institutions, community colleges need to research developmental education costs and impact.

It is common to encounter student unit record data in the community college and to analyze the impact of educational interventions using two groups of students, those exposed to the intervention and those not exposed. Yet results are limited in that the students are not typically randomly selected into experimental and comparison groups. Non-random selection implies that the two groups of students may be very different on key factors that affect the results of analyses through self-selection bias and other differences. In addition, Bettinger & Long (2005b) argued that comparing remedial students with non-remedial students introduced the potential for bias.

Propensity matching is a technique designed to simulate an experimental design, controlling for selection bias and creating almost equivalent experimental and comparison groups on key indicators. Comparisons of student outcomes using propensity matching, has been used to yield less biased results than are derived using simple comparisons (Rojewski et al., 2010). It is critical that researchers apply methodologies that control for selection bias.

Resources are scarce and monies need to be allocated to educational programs that are making a difference in students' success and in students' long-term outcomes.

Multilevel propensity matching—that is, matching across multiple institutions in order to account for variation in community colleges— is relevant to the study of developmental education. Multilevel studies can be used to exploit the contextual variation in developmental education across multiple institutions and to determine what institutional factors impact developmental programs. This researcher pointed to institutional variables that need further examination, such as the percent of racial concentration and percent of remedial needs as institutional constructs. In addition, institutional disparities in income level as measured by Pell recipients or students who are first generation are important constructs to examine.

Conclusion

Multilevel propensity matching was used in order to illuminate the methodology's use in the evaluative literature and to connect the study to one of the purposes for evaluation in educational settings: to strive for continuous improvement of educational programs. As Grubb (2001) pointed out in his review of the developmental education literature, "colleges must evaluate remedial education using sophisticated techniques such as treatment and control groups; program evaluation and improvement is central to improving remedial student's outcomes" (p. 10). Notably, as Attewell et al. (2011) wrote:

> As policymakers [foundations, researchers] intensify their calls for increasing graduation rates, it becomes important to consider the relative effect sizes of various predictors, to determine which factors have direct effects and which are mediating variables, and to determine whether mechanisms vary across types of college—so that policies can be well targeted and potential interventions be appropriately prioritized. (p. 537).

There is much intentional and focused work ahead of community college practitioners, researchers and policy makers. This study demonstrated the use of a powerful tool to examine the propensity to complete - college math and to facilitate the future policy, institutional and national discussions surrounding student success frameworks.

REFERENCES

Achieving the Dream. (2010). *Process for Institutional Change in Achieving the Dream.* Retrieved from http://www.achievingthedream.org

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and Bachelor's degree attainment.* Washington, DC: U.S. Department of Education. Retrieved from http://www2.ed.gov/pubs/Toolbox/index.html

Adelman, C. (2005). *Moving into town— and moving on: The community college in the lives of traditional-age students.* Washington, DC: U.S. Department of Education. Retrieved from http://www2.ed.gov/rschstat/research/pubs/comcollege/index.html

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences.* Upper Saddle River, NJ: Prentice Hall.

Alfred, R.L., & Lum, G.D. (1988). Remedial program policies, student demographic characteristics, and academic achievement in community colleges. *Community College Journal of Research and Practice, 12*(2), 107-120.

American Association of Community Colleges. (2010). *Community College Statistics.* Washington, D.C. Retrieved from http://www.aacc.nche.edu

Ashar, H., & Skenes, R. (1993). Can Tinto's student departure model be applied to nontraditional students? *Adult Education Quarterly, 43*(2), 90-100. doi: 10.1177/0741713693043002003

Attewell, P., Heil, S., & Reisel, L. (2011). Competing explanations of undergraduate noncompletion. *American Educational Research Journal, 48*(3), 536-559.

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

Bahr, P.R. (2008). Does mathematics remediation work?: A comparative analysis of academic attainment among community college students. *Research in Higher Education, 49*(5), 420-450.

Bahr, P.R. (2009). Revisiting the efficacy of postsecondary remediation: The moderating effects of depth/breadth of deficiency. *Review of Higher Education, 33*(2), 177.

Bahr, P.R. (2010a). Making sense of disparities in mathematics remediation: What is the role of student retention? *Journal of College Student Retention: Research, Theory, and Practice, 12*(1), 25-49.

Bahr, P.R. (2010b). Preparing the underprepared: An analysis of racial disparities in postsecondary mathematics remediation. *Journal of Higher Education, 81*(2), 209-237.

Bahr, P.R. (2013). The deconstructive approach to understanding community college students' pathways and outcomes. *Community College Review, 41*(2), 137-153. doi: 10.1177/0091552113486341

Bailey, T. (2009). Challenge and opportunity: Rethinking the role and function of developmental education in community college. *New Directions for Community Colleges,* (145), 11-30.

Bailey, T., Jenkins, D., & Leinbach, T. (2005). *Community college low-income and minority student completion study: Descriptive statistics from the 1992 high school cohort.* New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/publications/low-income-minority-completion-1992.html

Bailey, T.R., Jeong, D.W., & Cho, S.W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review, 29*(2), 255-270.

Bartlett, J.E., Kotrlik, J.W., & Higgins, C.C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology Learning and Performance Journal, 19*, 43-50.

Bean, J.P., & Metzner, B.S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research, 55*(4), 485-540.

Belfield, C., & Crosta, P.M. (2012). *Predicting success in college: The importance of placement tests and high school transcripts.* (CCRC Working Paper No. 42). New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/publications/predicting-success-placement-tests-transcripts.html

Bettinger, E., & Long, B.T. (2003). *The role and effect of remedial education in two-year colleges*. Retrieved from http://www.ilr.cornell.edu/cheri/conferences/upload/2003oct/chericonf2003_05.pdf

Bettinger, E.P., & Long, B.T. (2005a). *Addressing the needs of under-prepared students in higher education: Does college remediation work?* (Working Paper No. 11325). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w11325.pdf

Bettinger, E.P., & Long, B.T. (2005b). Remediation at the community college: Student participation and outcomes. *New Directions for Community Colleges, 2005*(129), 17-26. doi: 10.1002/cc.182

Bostian, B.E. (2008). *Avoiding remedial education: Academic effects on college transfer students*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. AAT 3320959)

Boylan, H., Bliss, L., & Bonham, B. (1997). Program components and their relationship to student success. *Journal of Developmental Education, 20*(3), 2-8.

Boylan, H.R., & Saxon, D.P. (1999). What works in remediation: Lessons from 30 years of research. *The League for Innovation in the Community College.* Retrieved from http://www.losmedanos.edu/deved/documents/WhatWorksinRemediation.pdf

Breneman, D.W., & Harlow, W.N. (1998). Remedial education: Costs and consequences. Remediation in Higher Education Symposium. *Thomas B. Fordham Foundation Report, 2*(9), 1-57.

Burdman, P. (2009). Sustaining changes that support student success in community college. *New Directions for Community Colleges, 2009*(145), 31-41.

Calcagno, J.C., Crosta, P., Bailey, T., & Jenkins, D. (2006). *Stepping stones to a degree: The impact of enrollment pathways and milestones on community college student outcomes.* (CCRC Working Paper No. 4). New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/publications/impact-enrollment-pathways-milestones.html

Calcagno, J.C., & Long, B.T. (2008). The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance. Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14194

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31-72.

Cho, S.W., & Karp, M.M. (2013). Student success courses in the community college: Early enrollment and educational outcomes. *Community College Review, 41*(1), 86-103. doi: 10.1177/0091552112472227

Clery, S., & Topper, A. (2008). Keeping informed about achieving the dream data. *Data Notes*, *3*(4), 1-4. Retrieved from http://files.eric.ed.gov/fulltext/ED521310.pdf

Completion by Design (CBD). (2011). *What is completion by design's approach to this work?* Retrieved from http://completionbydesign.org/faq/what-completion-designs-approach-work

Creswell, J.W. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.

Deil-Amen, R., & Rosenbaum, J.E. (2002). The unintended consequences of stigma-free remediation. *Sociology of Education, 75*(3), 249-268.

Feldman, M.J. (1993). Factors associated with one-year retention in a community college. *Research in Higher Education, 34*(4), 503.

Fike, D.S., & Fike, R. (2008). Predictors of first-year student retention in the community college. *Community College Review, 36*(2), 68.

Frye, B.E., Bartett, J.E., & Smith, K.D. (2013). The propensity to avoid developmental math in community college: A focus on minority students. In R. T. Palmer & J. L. Wood (Eds.), *Community Colleges and STEM: Examining Underrespresented Racial and Ethinic Minorities* (pp. 101-122). New York, NY: Routledge.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York, NY: Cambridge University Press.

Green, S.B., & Salkind, N.J. (2011). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data* (International Edition). Upper Saddle River, NJ: Pearson-Prentice Hall.

Grubb, W.N. (2001). *From black box to Pandora's Box: Evaluating remedial/developmental education.* New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/publications/black-box-evaluating-remedial-education.html

Grubb, W.N., & Cox, R.D. (2005). Pedagogical alignment and curricular consistency: The challenges for developmental education. *New Directions for Community Colleges, 129*), 93-103.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Hofmann, D.A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23*(6), 723-744.

Hoyt, J.E. (1999). Remedial education and student attrition. *Community College Review, 27*(2), 51.

Hughes, K L., & Scott-Clayton, J. (2011). Assessing developmental assessment in community colleges. *Community College Review, 39*(4), 327-351.

Huneycutt, K. (2010). *Using institutional context and student characteristics to predict a student passing a developmental course at a community college: A multilevel approach.* (Doctoral Dissertation). Retrieved from North Carolina State University Institutional Repository: http://repository.lib.ncsu.edu/ir/handle/1840.16/6115

Jenkins, D., Zeidenberg, M., & Kienzl, G.S. (2009). *Educational outcomes of I-BEST, Washington State Community and Technical College System's Integrated Basic Education and Skills Training Program: Findings from a multivariate analysis.* (CCRC Working Paper No.16). New York, NY: Columbia University, Teachers College, Community College Research Center.

Kasworm, C.E, Sandmann, L.R., & Sissel, P.A. (Eds.). (2000). *Adult learners in higher education*. San Francisco: Jossey-Bass.

Kelcey, B. (2011). Multilevel propensity score matching within and across schools. *Journal of the American Statistical Association, 99*, 609-618.

Kim, J., & Seltzer, M. (2007). Causal Inference in Multilevel Settings in Which Selection Processes Vary across Schools. (CSE Technical Report 708). *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*. Retrieved from http://www.cse.ucla.edu/products/reports/r708.pdf

Kolenovic, Z., Linderman, D., & Karp, M. M. (2013). Improving Student Outcomes via Comprehensive Supports Three-Year Outcomes From CUNY, Accelerated Study in Associate Programs (ASAP). *Community College Review, 41*(4), 271-291.

Leinbach, D.T., & Jenkins, D. (2008). Using longitudinal data to increase community college student success: A guide to measuring milestone and momentum point attainment. *CCRC Research Tools No. 2*. New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/publications/research-tools-2.html

Leuven, E. and Sianesi, B. (2003). *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.* Retrieved from http://ideas.repec.org/c/boc/bocode/s432001.html

Levin, H.M., & Calcagno, J.C. (2008). Remediation in the community college: An evaluator's perspective. *Community College Review, 35*(3), 181.

Martorell, P., & McFarlin Jr., I. (2007). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *The Review of Economics and Statistics, 93*(2), 436-454.

Mertler, C.A., & Vannatta, R.A. (2010). *Advanced and multivariate statistical methods: Practical application and interpretation* (4th ed.). Los Angeles, CA: Pyrczak Publishing.

North Carolina Community College System (NCCCS). (2013a). *Combined Course Library.* Retrieved from http://www.nccommunitycolleges.edu/ccl.html

North Carolina Community College System (NCCCS). (2013b). *Performance Measures for Student Success.* Retrieved from http://www.nccommunitycolleges.edu/Publications/docs/Publications/2013%20Performance%20Measures%20Report.pdf

North Carolina Community College System (NCCCS). (2009). *2009 critical success factors: For the North Carolina community college system.* (20th ed.). Retrieved from www.nccommunitycolleges.edu/Publications/archivedCsfReports.html

Oakes, J.M., & Johnson, P.J. (2006). Propensity score matching for social epidemiology. In J.M. Oakes & J.M. Kaufman (Eds.), *Methods in Social Epidemiology* (pp. 364-386). San Francisco, CA: Jossey-Bass.

O'Banion, T. (2013, October). *Keynote Address, 2013 Fall Forum*. Speech presented at Central Piedmont Community College, Charlotte, NC.

O'Hear, M.F., & MacDonald, R.B. (1995). A critical review of research in developmental education: Part I. *Journal of Developmental Education, 19*(2), 2-6.

Oudenhoven, B. (2002). Remediation at the community college: Pressing issues, uncertain solutions. *New Directions for Community Colleges,* 117, 35-44.

Padgett, R.D., Salisbury, M.H., An, B.P., & Pascarella, E.T. (2010). Required, practical, or unnecessary? An examination and demonstration of propensity score matching using longitudinal secondary data. *New Directions for Institutional Research*, (S2), 29-42.

Pascarella, E.T., & Chapman, D.W. (1983). A multi-institutional, path analytic validation of Tinto's model of college withdrawal. *American Educational Research Journal, 20*(1), 87-102.

Perin, D. (2005). Institutional decision making for increasing academic preparedness in community colleges. *New Directions for Community Colleges, 2005*(129), 27-38.

Pretlow III, J., & Wathington, H.D. (2011). Cost of developmental education: An update of Breneman and Haarlow. *Journal of Developmental Education, 35*(1), 2-12.

Reynolds, C.L., & DesJardins, S.L. (2009). The use of matching methods in higher education research: Answering whether attendance at a 2-year institution results in differences in educational attainment. In J.C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (pp. 47-97). New York, NY: Springer.

Rojewski, J.W., Gemici, S., & Lee, I.H. (2009). *Propensity score matching with observational data.* Unpublished manuscript, University of Georgia, Athens, GA.

Rojewski, J.W., Lee, I.H., & Gemici, S. (2010). Using propensity score matching to determine the efficacy of secondary career academies in raising educational aspirations. *Career and Technical Education Research, 35*(1), 3-27.

Roksa, J., Jenkins, D., Jaggars, S.S., Zeidenberg, M., & Cho, S. (2009). *Strategies for promoting gatekeeper success among students needing remediation: Research report for the Virginia Community College System.* New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://files.eric.ed.gov/fulltext/ED507824.pdf

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Rosenbaum, P.R., & Rubin, D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516-524.

Rosenbaum, P.R., & Rubin, D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*(1), 33-38.

Roueche, J.E., & Hurlburt, A.S. (1968). The open-door college: The problem of the low achiever. *The Journal of Higher Education, 39*(8), 453-456.

Scott-Clayton, J.E. (2011). *Do high-stakes placement exams predict college success?* (CCRC Working Paper No. 41). New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/publications/high-stakes-placement-exams-predict.html

Strong American Schools. (2008). *Diploma to Nowhere.* Retrieved from http://www.deltacostproject.org/resources/pdf/DiplomaToNowhere.pdf

St. Pierre, E.A. (2006). Scientifically-based research in education: Epistemology and ethics. *Adult Education Quarterly, 56*(4), 239-266. doi: 10.1177/0741713606289025

Thoemmes, F.J., & Kim, E.S. (2011). A systematic review of propensity score methods in the social sciences. Multivariate Behavioral Research, 46(1), 90-118.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (Vol. 2nd). Chicago: The University of Chicago Press.

Tinto, V., & Pusser, B. (2006). Moving from theory to action: Building a model of institutional action for student success. *National Postsecondary Education Cooperative.* Retrieved from http://web.ewu.edu/groups/academicaffairs/IR/NPEC_5_Tinto_Pusser_Report.pdf

Titus, M.A. (2007). Detecting selection bias, using propensity score matching, and estimating treatment effects: An application to the private returns to a master's degree. *Research in Higher Education, 48*(4), 487-521.

Twombly, S., & Townsend, B.K. (2008). Community college faculty: What we know and need to know. *Community College Review, 36*(1), 5-24.

U. S. Department of Education. (2007). *National Center for Education Statistics: IPEDS Fall,2007 Enrollment Survey.* Retrieved from http://nces.ed.gov/ipeds/datacenter/InstitutionByName.aspx

Umoh, U.J., Eddy, J. & Spaulding, D.J. (1994). Factors related to student retention in community college developmental education mathematics. *Community College Review, 22*(2), 37-47. doi: 10.1177/009155219402200205

Wild, L., & Ebbers, L. (2002). Rethinking student retention in community colleges. *Community College Journal of Research and Practice, 26*(6), 503-519.

Xu, D., & Jaggars, S.S. (2011). *The effectiveness of distance education across Virginia's Community Colleges: Evidence from introductory college-level math and English courses. Educational Evaluation and Policy Analysis, 33*(3), 360-377.

APPENDIX

Appendix A: Letter to Chair of College Institutional Review Board

Dear Chair of College Institutional Review Board,

My name is Bobbie Frye, Director of Institutional Research at Central Piedmont Community College and doctoral student in the Leadership, Policy, Adult and Higher Educational Research department (LPAHE) at North Carolina State University (NCSU).

My dissertation, with Dr. James Bartlett, at NCSU, investigates the student and academic factors associated with successful outcomes of developmental math community college students in North Carolina. As such, I hope you are willing to help me by providing the secondary data or a raw data extract provided to Completion by Design (CBD) colleges as part of their participation in the CBD initiative.

The data will be treated with anonymity and only aggregate results will be presented. If you wish to receive results upon request they will be shared with your college.

Your participation is imperative to the success of this study, which will analyze data from cooperating Completion by Design NC community colleges.

Thank you in advance for your support.

Bobbie Frye
Doctoral Student
Department of Education
North Carolina State University
Raleigh, North Carolina