2018

The Relationship between Neighborhood Characteristics and College Academic Outcomes Among an NCAA Division I Student-Athlete Population: A Multilevel Approach

Ann Kearns Davoren
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

THE RELATIONSHIP BETWEEN NEIGHBORHOOD CHARACTERISTICS AND COLLEGE ACADEMIC OUTCOMES AMONG AN NCAA DIVISION I STUDENT-ATHLETE POPULATION: A MULTILEVEL APPROACH

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

PROGRAM IN RESEARCH METHODOLOGY

BY

ANN KEARNS DAVOREN

CHICAGO, IL

MAY 2018

ACKNOWLEDGEMENTS

There are countless people to thank. I would like to start with Dr. Meng-Jia Wu for her guidance as chair of this committee, for her excellence in teaching during my coursework, for her persistent belief in me, and for her enduring kind manner. I also need to thank her for reminding me to believe in myself. I'll remember fondly our discussion of the anxiety bell curve – I know to not let it get the best of me. To Drs. Fujimoto and Boker, thank you for your thoughtful review of my work. Your presence on this committee pushed me to do better and more advanced work, and your recommendations improved this final product immensely. Dr. Boker, thank you, too, for a dataset that was a doctoral student's dream.

I have been blessed with many wonderful teachers and mentors who have inspired and challenged me. Specifically, Dr. Vicent Berdayes at Saint Mary's College and Dr. Richard Rubin at Indiana University, thank you for helping to show me my potential and for introducing me to the field of research.

A very special thank you needs to go to Dr. John J. (Jack) McArdle, who was the impetus for this entire endeavor. One fateful drive to the Indianapolis airport resulted in the pursuit of this degree. Without his counsel and a little nudge, I would not have done this. I am equally grateful for Jack's mentorship over the years, for his patience in working with me on NCAA initiatives, and for his friendship and humor.

I owe a special thank you to the NCAA research department for the data used in this study, for the time to devote to my education, and for years of collaboration and friendship.

Specifically, thank you to Mr. Gregg Summers for his patience in working with me on this dataset. I know I tried it on more than one occasion! And, a special thank you to Mr. Todd Petr. I have been honored to work on your team and count myself as truly fortunate that I have been able to continue my work with the research staff throughout all these years. If I can be half the leader you are, I will be a success.

To my siblings, nieces, friends, and all of my extended family who have endured countless conversations about methodology, statistics, and "neighborhood effects," I thank you for your interest in my work (even when it wasn't very interesting), for your support, and for helping me to celebrate the milestones along the way.

To my parents, Judy and Jerry, without whom none of this would have been possible, thank you. Thank you for my education. Thank you for believing in me and teaching me that there are no limits. Thank you for allowing me to lean on you when I needed some extra support. I hope you know how much my efforts are because of you.

And finally, to Tom. Who I am and what I do today is largely due to your presence in my life and to your consistent optimism. I see the world in a much brighter way because of you. I thank you from the bottom of my heart for everything you did to ensure I felt supported through this process, for your enthusiastic encouragement, and for reminding me that every once in a while, I needed to sharpen the saw. I am so grateful I get to do this life with you.

I dedicate this to Tom.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APP	Academic Performance Program
APR	Academic Progress Rate
GPA	Grade point average
GSR	Graduation Success Rate
EC	Eligibility Center
HBCU	Historically Black College and University
HGLM	Hierarchical generalized linear modeling
HLM	Hierarchical linear modeling
HSCGPA	High school core grade point average
IEC	Initial Eligibility Clearinghouse
I-E-O	Input-Output-Environment (Astin's theory)
IPEDS	Integrated Postsecondary Education Data System
ICC	Intraclass correlation coefficient
MBA	Men's Baseball
MFB	Men's Football
M/WBB	Men's or Women's Basketball
MSI	Minority Serving Institution
MTO	Moving to Opportunity
NCAA	National Collegiate Athletic Association

PTD	Progress toward degree
SES	Socioeconomic status
SF1	Summary File 1
SF3	Summary File 3

ABSTRACT

Over 170,000 students participate annually in National Collegiate Athletic Association (NCAA) Division I sports. Approximately one-third of these student-athletes fail to graduate from their initial school of enrollment within six years. While some will go on to graduate from a transfer institution, roughly 15% will fail to earn a degree while competing for an NCAA Division I school. Using U.S. census block group data, this study adds the neighborhood characteristics of education, employment, income, and racial composition to prediction models of first-year GPA and six-year baccalaureate degree attainment among an NCAA Division I student-athlete sample. The use of multilevel modeling accounted for nesting of student-athletes within institutions and controlled for the potential of group effects on individual outcomes. Improving the statistical models used to predict academic outcomes among NCAA student-athletes will help to ensure that they are properly evaluated for the potential to be academically successful and enable early identification of those with heightened academic risk. Early identification can help institutions direct relevant academic support services to those most in need, and in some cases, consider the potential of an academic redshirt year.

The findings indicate that while consideration of the educational attainment of the neighborhood adds to the predictive ability of first-year GPA, the meaningful impact is quite small. Cost-benefit analyses may reveal that the added burden of data collection and reduction in transparency is not worth the minimal addition of explained variance in the outcome, particularly in light of the lack of a significant relationship with the other outcomes.

CHAPTER ONE

IN WHICH THE TOPIC IS INTRODUCED

In a world that is increasingly reliant on technology and decreasingly reliant on the service industry, a college undergraduate education is becoming a necessary but not sufficient step in ensuring employability and job and financial security (Kuh, Kinzie, Buckley, Bridges & Hayek, 2007). And for some, it is a necessary but not sufficient step in realizing dreams of a fulfilling career, gaining freedom, and striving for equality. College, however, is not for everyone. Federal six-year bachelor's degree attainment rates hover around 60% (U.S. Department of Education, 2016a). And while federal rates do not account for students who transfer between schools or stop out and eventually graduate, it is safe to assume that a sizeable proportion of the U.S. population enters college without completing a degree.

Studies have shown that academic struggles are a primary cause of those who leave prior to earning a bachelor's degree (Stinebrickner & Stinebrickner, 2013). For some, academic difficulties are merely the symptom of another problem – adjustment difficulties, homesickness, interpersonal or relationship problems, health problems, financial concerns – the list goes on. For others, they arrive on campus academically ill-prepared for the rigors of a college education. Early identification of these students with the proper interventions in place to assist them academically could be the difference between a student who leaves an institution with a diploma in hand and a student who leaves empty-handed.

A special population of college students found only in the United States is the varsity student-athlete. There are currently over 485,000 student-athletes attending NCAA member institutions (NCAA, 2016a). Some of these student-athletes compete merely for the love of the game, receive no athletics-based financial aid, and attend institutions that are not permitted to consider their athletics talent when evaluating their applications for admission. Others compete on a much more visible scale. There are over 170,000 student-athletes annually who attend one of the 347 NCAA Division I member institutions. These student-athletes compete at the highest collegiate level. The majority help to finance their educations through athletics-based financial aid. They spend upwards of 42 hours each week on athletics activities (NCAA, 2016b), and they often are in the spotlight of American sport culture with their games appearing on national television. These Division I student-athletes represent between 1-37% of the total student population on their campuses (E. Irick, personal communication, May 22, 2017).

A great deal of controversy surrounds Division I student-athletes, particularly as it relates to the *student* in student-athlete. From popular media to academic journals, debate has ensued as to whether Division I student-athletes are participating in their academic lives in a manner that is comparable to their non-athlete peers (Shulman & Bowen, 2000; Tracy, 2017; Umbach, Palmer, Kuh, & Hannah, 2006). High profile academic scandals have increased the skepticism many have that student-athletes are legitimately fulfilling their academic obligations (Bauer-Wolf, 2017; Brutlag Hosick, 2016; Petr & McArdle, 2012). Data, however, show that student-athletes graduate at the same rate as their non-athlete peers, and when comparing subgroups based on gender and ethnicity, student-athletes graduate at greater rates in every category except white males (NCAA, 2016c). Student-athletes themselves also report great satisfaction with their collegiate academic careers (NCAA, 2016b).

Background of the Study

The NCAA was founded in 1906 in response to safety concerns for football players (Oriard, 2012). In its earliest years, the primary concern of the association was matters of amateurism and commercialism. While academics were informally addressed at annual conventions, the NCAA did not become involved in academic matters in any formal capacity until the 1950s when progress-toward-degree criteria was established (Oriard, 2012). Then, in 1965, the NCAA passed its first initial eligibility criteria, outlining the standards prospective student-athletes must meet to participate in NCAA Division I athletics. Prospective student-athletes would be eligible for athletics-based aid only if they entered college with a high school rank and standardized test score that predicted a 1.6 college grade point average (GPA) (Oriard, 2012). The so-called 1.600 rule was eventually abolished in 1973, and the NCAA would not establish initial academic eligibility criteria again until 1983. Since then, the national association has become not only more involved in establishing legislation member institutions must follow when determining the academic eligibility of their student-athletes for athletics competition, but it also has become much more data-driven in its policy setting (Petr & McArdle, 2012).

Since 1994, the NCAA has required that all prospective student-athletes be certified by the NCAA before competing for an NCAA institution. The academic certification states that the prospective student-athlete has met the minimum, initial academic criteria, and s/he is granted academic eligibility for athletics participation. The student-athletes must then be admitted to the individual institutions based on institutional admission standards. The NCAA's initial academic eligibility criteria has gone through a series of changes over the years. The changes were driven primarily by the desire to ensure student-athletes who were being admitted had the minimum qualifications to succeed in postsecondary education while also trying to ensure that certain subpopulations were not differentially affected by the criteria, most notably, low-income and racially underrepresented groups. The most recent change to the initial academic eligibility standards took effect in the 2016-17 academic year. The GPA used by the NCAA in initial academic eligibility decisions uses grades in core academic courses only, including English, math, natural/physical sciences, social science, foreign language, and comparative religion/philosophy (NCAA, 2016d). The current standards require a 2.30 high school grade point average in 16 core courses, otherwise known as the high school core GPA (HSCGPA), along with a 900 SAT or 75 sum score on the ACT to be fully eligible for competition. Since 2003, the NCAA has used a sliding scale that pinpoints the required standardized test score (using either the SAT or ACT) based on their HSCGPA, with 2.30 as the minimum for full qualification (Petr & McArdle, 2012). The NCAA currently has three levels of initial eligibility based on the student's high school academic record: 1) ineligible or nonqualifier; 2) partial qualifier, also referred to as an academic redshirt, and 3) qualifier. Student-athletes in this last category are fully eligible to receive athletics-based financial aid, practice, and compete with the team. Student-athletes who are partial qualifiers, or academic redshirts, have an academic record that places them at-risk for succeeding in college, but they show promise. These students are permitted to receive athletics-based financial aid and can practice with the team, but they are not eligible for competition in their first year. Walter Harrison, President of the University of Hartford and Chair of the former NCAA Committee on Academic Performance, said about the new initial eligibility rules,

> It is the hope…that this approach will continue to allow access to student-athletes who have reasonable chances of succeeding but will assist those who are borderline students to get themselves appropriately integrated academically before adding the rigors of athletics competition to their lives at the college or university level. (Harrison, 2012, p.76)

Statement of the Problem

The current NCAA models that rely on HSCGPA and standardized test can predict reasonably well how a student-athlete will perform in college (Petr & McArdle, 2012). Any opportunity, however, to improve upon these models aids the student-athletes in potentially granting access to those who may otherwise have been left out, identifying students who may be good candidates for an academic redshirt year[1], and denying eligibility to those who, based on the data, may not have the capabilities at this time to succeed in a postsecondary curriculum with the added pressure of NCAA Division I athletics. The problem this study sought to address was whether the use of neighborhood characteristics can improve the predictive validity of college academic outcome models when coupled with the high school academic information already used.

Research Questions

The purpose of this research study was to discover if there is added value in including neighborhood factors in individual-level college outcome prediction models of first-year GPA, first-year retention, and six-year degree attainment. The study's research questions focused on

---

[1] Ideally, student-athletes who are selected for an academic redshirt year will be afforded targeted academic interventions while there is still time to make the necessary adjustments or impart the necessary skills so that the students may be successful at their chosen institutions.

discovering if the addition of U.S. census block group data improved the predicative validity of college outcome models that include student-level high school academic and individual demographic characteristics, as well as institutional characteristics. This study also explored what the ideal combination of the variables was and whether the neighborhood data relates differently to the outcomes for non-white students compared to white students or student-athletes who participated in academically at-risk sports compared to their counterparts in sports not deemed academically at-risk.

For the purposes of this study, the student population of interest was NCAA Division I student-athletes. As such, they are traditional-aged, full-time college students who were attending four-year baccalaureate-granting institutions. Using a combination of hierarchical linear modeling and hierarchical generalized linear modeling, the study addressed the following research questions:

1. Are U.S. census block group data significantly related to NCAA Division I student-athlete first-year college cumulative grade point average, first-year retention and eventual six-year degree attainment after controlling for student-athlete demographics and pre-college academic characteristics and college-level institutional characteristics?
2. Do U.S. census block group data relate to NCAA Division I student-athlete first-year college cumulative grade point average, first-year retention, and eventual six-year degree attainment differently for student-athletes who participate in academically at-risk sports and their counterparts in sports not deemed academically at-risk?

3. Do U.S. census block group data relate to NCAA Division I student-athlete first-year college cumulative grade point average, first-year retention, and eventual six-year degree attainment differently for minority student-athletes and white student-athletes?

Study Significance

The study provided information regarding if and how prediction models of first-year GPA, first-year retention, and six-year degree attainment, using strictly precollege characteristics, improved with the addition of neighborhood factors. Strengthening the models used to predict student-athlete success will most directly and immediately inform NCAA initial academic eligibility policies that affect hundreds of thousands of students annually. It also has the potential to shape the admissions criteria used by the nearly 2,500 four-year higher education institutions across the country that serve more than 10 million students.

Ensuring that students are properly and thoroughly evaluated for the potential to be successful either as an NCAA student-athlete or as a student-at-large at a given institution helps to potentially broaden access for students, ensure students and their families are investing in an education that has a realistic probability of resulting in a degree, and helps institutions to direct their academic support services to those most in need. In turn, this will reduce the number of students who transfer institutions in search of finding a better fit and reduce the number of students who drop out entirely before completing a degree. The act of transferring may delay graduation, which also delays time to employment and adds to the financial burden of added tuition payments beyond the traditional four years (Petr & Paskus, 2009). For students who drop out and do not return to complete their degrees, a host of difficulties and challenges follow, including higher unemployment, lower earnings, and greater defaults on student loans (Nguyen, 2012).

Beyond the potential practical implications of the research, this study contributes to the field of neighborhood effects in education research, which is still very much a growing field. Most of the research to date between neighborhood and education has looked at outcomes along the educational pipeline up to and including high school graduation (Harding, 2003; Leventhal & Brooks-Gunn, 2000). Little has been done with the relationship between the census data and college academic outcomes. Other uses of U.S. census data have focused on how neighborhoods affect a variety of social outcomes that occur across the lifespan, including infant mortality (Wilson & Daly, 1997), early child development (Brooks-Gunn, Duncan, Klebanov, & Sealand, 1993), adolescent risky behavior (Harding, 2003; Leventhal & Brooks-Gunn, 2000), and adult employment (Lynn & McGeary, 1990), among others.

While very few have explored the relationship between U.S. neighborhood characteristics and college outcomes, those that have, have found promising results for a significant link between the two (Aaronson, 1998; Chetty, Hendren, & Katz, 2015; Duncan, 1994; Leventhal & Brooks-Gunn, 2000; Nieuwenhuis & Hooimeijer, 2016). The focus of the research, however, has been primarily on initial college enrollment or years of attainment, omitting in-college outcomes, such as first-year persistence and grades, which this research addresses. Much of the research also has relied on small samples (Leventhal & Brooks-Gunn, 2000), restricted geographic areas (Chetty et al., 2015) or has not accounted for the individuals' academic backgrounds (Aarsonson, 1998; Duncan, 1994), which this research also addresses. Finally, the bulk of neighborhood effects research has had to rely on census tracts, which contain, on average, 4,000 individuals and were “designed to be relatively homogenous with respect to population characteristics, economic status, and living conditions” (Iceland & Steinmetz, 2003). This research, however, used block groups, which are, on average, comprised of 600 to 3,000 individuals and are contained wholly within census tracts (Iceland & Steinmetz, 2003). The use of block group data better ensures the homogeneity of the neighborhood characteristics used in the modeling.

One study that is worth mentioning specifically and, in more detail, given the nature of the research design, is the Moving to Opportunity (MTO) study. Moving to Opportunity was an experiment administrated by the U.S. Department of Housing and Urban Development between 1994-1998. Low-income families living in Baltimore, Boston, Chicago, Los Angeles or New York were randomly assigned to one of three groups: a group that provided the families with housing vouchers restricted to low-poverty neighborhoods; a second group received a Section 8 housing voucher, and a control group. The main focus of the project was on the effect of the vouchers on long-term outcomes for the children in these families, including educational outcomes (Chetty et al., 2015). The experimental component of the MTO makes it truly unique in its ability to tease out the effects of neighborhood on educational attainment. The experiment, however, is able to evaluate only the role of poverty – or lack thereof – on educational attainment. Other factors, such as racial composition, average education of the neighborhood, employment, etc. were not considered. The use of census data in the present study accounts for these in an effort to determine not only if the neighborhood in which a student lives has a relationship with educational attainment, but what aspects of the neighborhood are most relevant.

The following chapters provide an overview of what has been done historically on college outcome prediction modeling, the variables known to be related to college outcomes, and a discussion of some of the research that has been done using neighborhood factors and U.S. census information. Following is a description of the data, including the sample and variables, and a description of the analytic methods that were used. After the presentation of the findings, a discussion of the potential implications of those findings as well as limitations of the research and suggested future research is presented.

CHAPTER TWO

IN WHICH THE LITERATURE IS EXPLORED

This chapter begins with a high-level discussion of outcome measures in higher education. This sets the framework for the subsequent and more nuanced discussions of variables known to be related to higher education outcomes and how these variables have been used in higher education prediction models. Specifically, the discussion of known related variables focuses on three overarching categories: individual factors, neighborhood factors, and institutional factors. The chapter concludes with a brief discussion of methodological issues when modeling with neighborhood characteristics and nested data.

Outcome Measures in Higher Education

Operationalizing student outcomes in higher education can take many forms. Historically, higher education outcomes have been defined by objective academic measures – among the most common are enrollment, grades, persistence, and eventual degree attainment (Burton & Ramist, 2001; Venezia, Callan, Finney, Kirst, & Usdan, 2005). Outcomes of interest, however, have evolved over the years to include subjective measures like student satisfaction (Kuh, Kinzie, Buckley, Bridges & Hayek, 2007) and sense of belonging or campus connection (Osterman, 2000), which are more difficult to quantify. Most recently, there has been growing attention paid to even more abstract measures, including things such as citizenship, critical thinking, and inclusiveness (Kuh et al., 2007). Not only are there diverse ways to measure outcomes, but outcomes that are meaningful will vary depending on student and institutional characteristics.

Degree attainment and persistence are dichotomous and easily labeled as success or failure. Degree attainment, particularly at a four-year baccalaureate institution, is among the most common college outcomes studied and a strong indicator of student success. Persisting to the next grade is considered a successful outcome, while leaving the institution is considered an institutional failure even if the student successfully transfers to another four-year institution. Persistence between years one and two, in particular, has been shown to be significantly correlated with eventual degree attainment (Kanarek, 1989, in Burton & Ramist, 2001). The simple act of transferring institutions is linked to lower graduation rates (Camara, 2003).

Grade point average, while measured on a continuum, still can be used to discriminate between successful outcomes versus others. First-year GPA, especially, is an early indicator of whether a student will persist at an institution through degree completion (Adelman, 2006; Kuh et al., 2007; Tinto, 1975). Pascarella and Terenzini said that first-year grades “are probably the single most revealing indicator…of successful adjustment to the intellectual demands of a particular college’s course of study” (1991, p. 388). They also are typically a measure that is relatively comparable across institutions because of the more or less homogenous courses taken across institutions in the first year (Burton & Ramist, 2001).

Using first-year GPA, first-year retention, and eventual degree attainment as outcomes present straightforward evidence of the success or failure of a student. The role of higher education admission offices is to evaluate applicants on their ability to be successful at their institutions. Another way to say this is that admission officers evaluate risk of student failure. These outcomes, while all correlated (Stinebrickner & Stinebrickner, 2013; Tinto, 1975), will produce prediction models that differ in the strength of the covariates to the outcome of interest (Burton & Ramist, 2001).

Factors Related to Higher Education Outcomes

Admission officers have a limited amount of data available to evaluate a student's likelihood of successfully completing a degree of study at their institutions. Individual factors, particularly past academic characteristics, have long been the primary variables considered when determining a student's potential for success. Included in these precollege academic characteristics often are the student's grades and test scores, the strength of the high school the student attended, and the strength of the student's high school curriculum (Adelman, 2006). Where there has been little attention is the effect of the environment in which the student lived prior to enrolling in college. Theoretical models hypothesize that the human and physical resources of the student's neighborhood help to not only directly shape the student's academic outcomes (Jencks & Mayer, 1990) but also indirectly shape the outcomes by influencing the student's academic expectations and motivations (Jencks & Mayer, 1990). Little, however, has been done to examine just how neighborhood effects are related to student-level, objective, college academic outcomes.

Individual Factors

Student-level college outcome models are driven by individual factors. Most notable are demographics and academic preparedness. Both gender and ethnicity have been shown to be related to each of the outcomes of interest. When assessing individual factors, however, it is the pre-college academic characteristics that carry the most weight, most notably the students' academic curriculum in high school.

Demographics. "It is sometimes said that when predicting future events that demographics is destiny" (Kuh, et al., 2007, p. 21). While demographics have been shown to consistently add to the ability to predict how a student will perform in college, when used alone, they leave a great deal of variability unaccounted for (Adelman, 2006). The primary demographics for the purposes of this study are gender and race/ethnicity. When comparing within demographic area, data show that there are vast differences in academic preparedness prior to college, college attendance, performance in college, and graduation rates. All leads to a need to account for these demographics when predicting college outcomes.

Gender. There is an abundance of data that demonstrates the widening gender gap in college-going rates and college completion rates. Between 1994 and 2012, the female-male gap of high school graduates who enrolled in college in the fall following completion of their high school curriculum grew from 2% to 10% (Lopez & Gonzalez-Barrera, 2014). When looking at the college-going gender gap by race, the differences for many are even wider. Among Hispanics and blacks, for instance, the female-male gap was 13% and 12% respectively in 2012 (Lopez & Gonzalez-Barrera, 2014). Across racial and ethnic groups, females earn baccalaureate degrees at greater rates than males as well. In the 2012-13 academic year, the narrowest gender gap in degrees conferred was among Asian students with 54% of females and 46% of males earning a degree. The widest gap was among black students with 65% of females and 35% of males earning a degree (U.S. Department of Education, 2016b). And, related to this, women outperform men while in college earning, on average, higher grades (Kuh et al., 2007).

There also is evidence that women are not only pursuing and earning baccalaureate degrees in greater numbers than their male counterparts, but that they are better prepared to do so (Kuh et al., 2007). Using decomposition analysis, Cho (2007) found that the gains women made in high school, including increases in math and science units and math and reading scores, help explain more than one-half of the change in the gender attendance gap since 1974.

Race and ethnicity. While the growing trend of females attending and graduating from four-year colleges at greater rates than males has been consistent and steady over the last three decades, the changes in attendance and graduation among various racial and ethnic groups is not as clear-cut. The overall gap of those enrolling in a two or four-year college the fall after completing high school has narrowed between whites and blacks and whites and Hispanics since 2015 while the gap between Asians and the three other racial groups has grown steadily since 2003 (U.S. Department of Commerce, 2016). Currently, the percent of 18 to 24-year-olds enrolled in college by racial group largely reflects the percent of public high school graduates in that group. For instance, in 2012, Hispanics made up 18% of the public high school graduates in the 2011-12 academic year, and 19% of the 18 to 24-year-old college population enrolled in 2012 (Krogstad & Fry, 2014). Graduation statistics, however, are not as representative. Hispanics, for example, made up just 9.5% of baccalaureate degrees conferred in the 2011-12 academic year (U.S. Department of Education, 2014).

College readiness and performance once in college, measured by GPA, also varies greatly by race. Taking remedial course work can fatally stall time to completion of a degree. Only 35% of students who take remedial courses earn a baccalaureate degree within 6 years. This is more than 20% lower than those who do not need remedial course work (Casselman, 2014). And, non-whites disproportionately need remedial coursework when compared with white students. Among black students, for example, 45% took remedial coursework as first-year students during the 2007-08 academic year compared with 31% of white students (Casselman, 2014). Once in college, blacks were nearly three times as likely as whites to have a cumulative GPA at time of graduation of less than 2.50. Whites also were more likely to earn a cumulative GPA at the time of graduation of at least a 3.0 when compared with Hispanics (75% vs. 64% respectively). There was no difference between whites and Asians (U.S. Department of Education, 2012).

Coming from a low-income family seems to compound the relationships between educational outcomes and gender or race/ethnicity alone. Among white high school graduates, for instance, 35% of low-income females enrolled in college immediately compared with 25% of males. The female-male gap among African-Americans was 19% (Kuh et al, 2007).

Academic preparedness. While data show that demographics certainly matter when predicting success in higher education, it is the academic preparation of the student that has the greatest relationship with college success – what the student takes in high school, the grades they earn, and their performance on standardized aptitude tests (Adelman, 2006).

Credits. When considering the use of credits or core courses in a model, it is less informative to consider just the number of credits or units that were taken because there is no accounting for the rigor of the class (Adelman, 2006). It is more beneficial to consider the number of units that were taken in certain academic disciplines and, to the degree possible, the level of the course. In a study of more than 30,000 college students enrolled in a postsecondary institution in Florida, the number of rigorous courses a student took in high school explained roughly one-third of the variance in enrollment in a four-year college (Long, Conger, & Iatarola, 2012). Once enrolled, Camara (2003) found that students who had not completed a standard high school core curriculum were 25% less likely to be on track to earn their bachelor's degree compared to those who had taken rigorous classes in high school. Camara also found that the negative effects of being a first-generation college student in persisting to a degree were dramatically reduced if the student had completed more than a core high school curriculum.

Using a sample of over 8,000 first-year students, Pike and Saupe (2002) found that completing a high school's core course contributed significantly and positively to a student's first-year college GPA and that the combination of class rank, standardized test and completion of the high school core course accounted for a little over one-third of the variance in first-year GPA.

High school GPA. While high school curriculum has shown consistently to have a stronger relationship with eventual degree attainment than other precollege academic characteristics (Adelman, 2006; Pike & Saupe, 2002), differences in record keeping and course naming strategies across high schools often make using curriculum in large-scale studies overly burdensome for researchers. In many cases, precollege characteristics used in modeling outcomes are restricted to GPA and standardized tests. In the majority of these cases, high school cumulative GPA is the stronger predictor of first-year grades in college. This finding holds when looking at smaller, institution-level studies, as well as larger, national studies (Adelman, 2006; Geiser & Santelices, 2007; Geiser & Studley, 2003).

High school GPA also has been shown to be a strong predictor of longer-term outcomes, including fourth-year college GPA. Cumulative high school GPA coupled just with SES accounts for 20% of the variance in fourth-year GPA in one study (Geiser & Santelices, 2007), and when paired with demographics and other precollege characteristics, high school GPA remains the best predictor of both cumulative fourth-year GPA and eventual graduation (Geiser & Santelices, 2007).

Standardized tests. While high school GPA is generally a stronger predictor of college outcomes when compared with standardized tests, test scores do consistently add to the prediction models (Geiser & Santelices, 2007; Geiser & Studley, 2003), and the use of both is better than either alone (Burton & Ramist, 2001; Geiser & Studley, 2003). They remain, however, one of the most controversial elements of college admission, largely due to evidence of differential scores by race (Geiser & Studley, 2003).

In spite of the debate over the appropriateness of using tests in high-stakes decision-making like admissions, their continued use and usefulness in prediction models of college outcomes, most notably GPA and retention, is well-documented. Bettinger, Evans, and Pope (2013) did an extensive study with a sample of over 20,000 college students on the ability of the ACT to predict outcomes at various stages of a college student's career. They found not only that the composite ACT score was highly correlated with first and second-year GPA and first and third-year retention, but that these findings were robust once individual demographics, high school GPA, and college and major fixed effects were added to the models. They also, interestingly, found that the English and Mathematics ACT subscores drove the models.

Studies exploring the added value of SAT scores in college outcome prediction models found similar results. Bridgeman, Pollack, and Burton (2004) compared like students on a variety of pre-college measures, including high school curriculum rigor, high school grades, and college selectivity and found that only 14% of students who had achieved a maximum of a 1000 on the SAT had earned a 3.5 or greater first year college GPA. This was contrasted against students who had an SAT in the range of 1010-1200, 1210-1400, and 1410+. Within those groups, 30%, 51%, and 77% respectively had earned a 3.5 or greater first year college GPA.

Neighborhood Factors

Geodemography, or neighborhood effects research, historically has been used in the public sector to study such issues as adolescent development, delinquency, and crime. The bulk of the research has focused on the relationship between where a person lives and primarily disadvantaged individual outcomes. Over three decades ago, William Julius Wilson did extensive research into the relationship between neighborhood, race and poverty, and his works, *The Declining Significance of Race* and *The Truly Disadvantaged*, are heavily referenced still today. While the exact statistics may have changed between Wilson's time and the present, the underlying findings have not. Much of Wilson's work focused on inner-city neighborhoods and housing developments and the extreme concentration of poverty, racial segregation, single-headed households, and crime in these areas (Wilson, 1987). More recent research indicates that these trends continue today. According to Sampson, Morenoff, and Gannon-Rowley (2002), neighborhood effects research has focused on the "geographic isolation of the poor, African-American, and single parent families with children" (p. 445-6) and the associated outcomes, including among others, low birthweight, teenage pregnancy, dropping out of high school, and child delinquency. After an extensive review of the literature, they found that there are a few consistent findings across studies, including: poverty and racial segregation are related, particularly poverty and high concentrations of African-Americans; neighborhoods do experience outcomes disproportionately, and these are largely related to "concentrated poverty, racial isolation, single-parent families, rates of home ownership, and length of tenure" (p. 446), and these findings are generally consistent regardless of the neighborhood unit of analysis (Sampson et al., 2002). More recently, the private sector has benefitted from geodemographic modeling and its enhanced marketing tools (Flowerdew & Goldstein, 1989; Singleton & Spielman, 2014). And while a fair amount of geodemographic modeling has been done on educational issues in primary and secondary education, very little has been done using geodemography in the study of higher education, particularly student-level higher education outcomes (Ainsworth, 2002).

Jencks and Mayer (1990) enumerate four different theories on how the neighborhood in which a child or adolescent lives may affect their individual outcomes. Collective socialization theory focuses on the inevitable responsibility adults in the neighborhood take on as role models to the children and adolescents. As children age, they likely will mirror the behaviors of adults they see every day in their neighborhood. Not only are children likely to mirror what they see adults doing, but they also are able to deduce some cause and effect when observing the behaviors of members of the community. Ainsworth (2002) explains,

> …with fewer positive role models in their neighborhood, children may be less likely to learn important behaviors and attitudes that lead to success in school…, both because of a lack of exposure to them and because they have no direct evidence that these attitudes and behaviors are useful or desirable." (p. 120).

The second theory also examines the role of adults but focuses on how adults from outside the neighborhood can influence adolescent behavior. Referred to as the institutional model (Jencks & Mayer, 1990), these adults work in the schools and other neighborhood institutions with which the adolescents interact on a regular basis. This means that whether the neighbors themselves influence behavior, it is possible for the institutions and the outsiders who are closely affiliated to influence behavior within a neighborhood. Peers arguably have a greater influence on adolescent behavior than do the adults. Epidemic models argue that behavior among peers literally is contagious. Engaging the theory from a negative viewpoint, Jencks and Mayer (1990) state,

> The critical feature of the model is that among individuals of any given susceptibility, the likelihood of antisocial or self-destructive behavior increases with exposure to others who engage in similar behavior. If children from low-SES families are more susceptible to

> such influences, increases in the proportion of low-SES families in a neighborhood will lead to exponential increases in bad behavior." (p. 114)

There is evidence, however, that the inevitable heterogeneity in neighborhoods will exacerbate challenges felt by those already at a disadvantage either because of income or race. The relative deprivation model argues that, in the case of educational attainment, students will compare themselves to those around them. If they see others out-performing them, they may lower their expectations or aspirations out of frustration of essentially not being able to keep up with the Jones'.

It is important to note that collective socialization theory, the institutional model, and the contagion model all can result in positive reinforcement for children in high-SES neighborhoods, negative reinforcement for children in low-SES neighborhoods, or each could affect a positive or negative change for adolescents whose familial SES differs from that of the neighborhood. The relative deprivation model, however, highlights the negative influence affluence in the neighborhood may have on lower SES families. It is conceivable that the effects of the relative deprivation could cancel out positive change that occurs from the influence of positive indigenous adults, positive outside adults within the neighborhood institutions or positive peers (Jencks & Mayer, 1990).

One other view point says that it is not the neighbors that exert influence, it is the neighborhoods that do. When comparing neighborhoods strictly based on affluence, more affluent neighborhoods have better schools, lower crime, and generally more opportunities for enriching experiences, such as museums, parks, and good libraries (Mayer & Jencks, 1989). Likely, it is both the physical and social space of neighborhoods that matter.

There is an important note to mention regarding the literature in neighborhood effects and student-level college attainment studies. In the American education literature, school effects and neighborhood effects are at times used interchangeably, particularly when discussing the research that has been done on the high school to college pipeline (Jencks & Mayer, 1990). With the rise in private education (U.S. Department of Education, 2017) and the use of vouchers (Cierniak, Stewart, & Ruddy, 2015), using school SES as a proxy for neighborhood SES is not feasible like it was decades ago when these studies were done. Moreover, there are roughly 1 to 58 census tracts in a zip code (U.S. Department of Housing and Urban Development, 2010), and because zip codes help to define the neighborhood school district lines, using school effects as a proxy for neighborhood effects will substantially increase the heterogeneity of the area being studied.

The research on neighborhood effects using national census data and college outcomes is much more established in the international literature. Brattbakk (2014) found that after controlling for individual and familial factors, the average education of the neighborhood had the greatest relationship with university degree attainment by the age of 29 in Norway. Where most of the work, however, has been done in higher education access is in England. The Higher Education Funding Council in England in 2005 issued a report stating "'the full extent of participation inequalities is revealed by using neighborhood level geographies such as census wards.'" (Adelman, 2009, p. 42). Scholars have devoted a great deal of time and creativity in studying the attendance rates of English students in higher education and maximizing the capabilities of their geodemographic models. Singleton, Wilson, and O'Brien (2012) used geodemographics to inform findings from a spatial interaction model exploring participation in higher education in England, specifically the flow of students from their local authorities to universities. Geodemographic analysis helped to explain some over and under prediction results from the model, notably related to income and race/ethnicity. Another study in England found that a measure of community wealth that placed the neighborhood in the 9^{th} decile was related to a likelihood of participation in higher education that was four time greater than communities in the 1^{st} decile (Harris, Grose, Longley, Singleton, & Brunsdon, 2010).

While the literature on neighborhood effects and college outcomes using a U.S. population is sparse, there are some consistent findings as to what neighborhood characteristics will be related most strongly to the outcomes of interest. The three dimensions that have been examined most frequently are neighborhood SES, residential stability, and the demographic composition of the neighborhood.

Neighborhood SES. By far, the most common neighborhood factor used in geodemographic studies is the poverty or income level of the neighborhood. Some studies will incorporate additional measures to form a composite SES. The way SES is operationalized will differ across studies, including both how it is defined and how it is used – as a continuous variable or as a categorical variable that has been dichotomized into affluence and poverty or high and low SES. Aspects of SES may include one or various combinations of income, employment, heads of households, proportion of professionals in the neighborhood, and level of educational attainment in the neighborhood (Aaronson, 1998; Ensminger, Lamkin & Jacobson, 1996; Leventhal & Brooks-Gunn, 2004). While much of the literature focuses on the perceived effects of poverty on academic outcomes and attainment (Duncan, 1994; Harding, 2003; Wilson, 1987), there are a few scholars who believe that is the presence of affluence that is most relevant (Brooks-Gunn et al, 1993, Sampson et al, 2002)

A few studies have found an interaction effect between SES and individual characteristics and academic outcomes. Ensminger, et al. (1996) found an interaction effect between SES measures and gender and high school graduation. Interestingly, in their study, the average poverty level of the neighborhood did not have a significant relationship with high school graduation for either females or males, but the proportion of residents with a white-collar occupation was significantly and strongly related to male high school degree attainment. Males who lived in a census tract with less than 40% of its residents in white-collar occupations were 3.5 times more likely to drop out of high school (Ensminger, et al., 1996). Maternal education also was significantly and positively related to both female and male high school graduation. (Ensminger, et al., 1996). These findings support a model that parcels out components of SES to explore the direct effects of each on academic outcomes in lieu of creating one composite variable.

A study from Wodtke, Harding, and Elwert (2011) found an interaction between SES, race, and high school graduation. The probability of graduating from high school was 20 points lower for black children in the most disadvantaged quintile compared to those in the most advantaged. The difference within a nonblack group was 10 points. Disadvantage was defined using a composite SES measure including poverty, unemployment, welfare, heads of household, educational attainment, and proportion in a managerial position.

Several studies have found that neighborhood SES has a relationship with academic attainment (Leventhal & Brooks-Gunn, 2000; Mayer & Jencks, 1989; Vartanian & Gleason, 1999). Leventhal and Brooks-Gunn (2000) found that neighborhood SES has a more consistent relationship with academic attainment than does race/ethnicity or residential stability. In a meta-analysis, five of the reviewed studies examined the role of neighborhood in academic attainment among young adults (aged 18+). In most, affluence had a positive relationship with high school completion and college attendance. College graduation was not an outcome of interest in the reviewed studies, nor was college persistence. Significance of results, and in some cases, the direction of results varied as a function of race and gender. When controlling for not only familial factors but also family processes, such as parental involvement, Dornbusch, Ritter, and Steinberg (1991) also found that neighborhood SES had a significant and positive relationship with self-reported high school grades and had a significantly stronger relationship among African-American students when compared with whites.

Vartanian & Gleason (1999) studied the relationship between neighborhood SES and college graduation among a population of students who were in high school between 1968 and 1981. They found a differential impact of neighborhood SES, comprised of the neighborhood's poverty rate, mean income, proportion of single-female head of household, and the percentage of managerial workers, by race. After controlling for familial characteristics, neighborhood SES did not have a significant relationship with college graduation for the black population. It did, however, have a significant relationship for white students. The greater the SES, the more likely the white students were to graduate from college. This finding only applied, however, to white students from affluent families.

Residential stability. Sharkey and Faber (2014) put forth a persuasive argument against the notion that neighborhood effects should be treated as static indicators. Any effect of the neighborhood on individual outcomes happens contextually, including the length of time the child has lived in the neighborhood. There are two ways of theorizing the relationship between residential stability and academic outcomes. The first is that intuitively, the longer the length of the exposure to a phenomenon, the greater the impact it will have (Sharkey & Faber, 2012).

Largely, this has been studied in the context of disadvantaged neighborhoods and has been shown to affect academic outcomes. Crowder and South (2011) and Wodtke et al. (2011), for example, show that the fraction of childhood spent in high-poverty areas is negatively correlated with outcomes such as high-school completion. Chetty and colleague's (2015) quasi-experimental study of more than five million families that moved across areas in the U.S. found that neighborhoods have causal exposure effects on children's outcomes. In particular, every year spent in a better area during childhood increases college attendance rates and earnings in adulthood, so the gains from moving to a better area are larger for children who are younger at the time of the move. Likewise, Crowder and South (2011) found that exposure to disadvantaged neighborhoods reduced the likelihood of high school graduation. In their study, however, they found that the effects of prolonged exposure had a greater negative relationship among a white sample than a black sample.

Another way of conceptualizing residential stability is as the average tenure of the residents in the neighborhood as opposed to the tenure of the individual in the neighborhood. Ainsworth (2002) focuses on the processes, including collective socialization, that cause neighborhoods to influence academic outcomes. In order for adults to have an effect as a role model – either in a positive or negative manner – children must have regular and extended exposure to them. Sharkey and Faber's (2012) assertion regarding the relationship between length of exposure and impact applies here as well. The longer a child is exposed to the same neighbors, the greater the impact those neighbors will have. Residential stability defined in this way, however, was found to not have a significant relationship with 10th grade standardized test scores. When studying the relationship between student-level tests scores and neighborhood employment, residential stability, economic deprivation, and racial diversity, only the proportion of what Ainsworth referred to as “high-status residents,” which was a composite of college graduates and individuals holding professional occupations, was significant in predicting the test scores. Residential stability, which was defined as the proportion who have lived in the same residence for at least 5 years, was not significant (Ainsworth, 2002). Although there is some conflicting evidence on the impact of residential stability in neighborhood effects models, scholars have consistently trumpeted the importance of including it, even if theory was the driving motivation (Sharkey & Faber, 2012).

Racial/ethnic composition of neighborhood. The demographic makeup of a neighborhood is the third common construct included in empirical research using neighborhood effects and is related to both SES and residential stability. While Ainsworth (2002) did not find a significant relationship with 10th grade standardized test scores and neighborhood racial/ethnic diversity after controlling for poverty and neighborhood stability, Sharkey and Faber (2014) theorize that the lack of racial variation within neighborhood is at least partially to blame for racial and ethnic composition having no effect or conflicting effects within the research.

> It is not possible to analyze the differential effects of exposure to highly disadvantaged residential settings for black children compared to other groups because in cities such as Chicago children from different racial and ethnic backgrounds occupy entirely different types of communities.” (Sharkey & Faber, 2014, p. 570)

Many of these neighborhood effects studies that found no evidence of a racial composition effect were conducted using 1990 or earlier census tract data. New data shows that while there still is a gap in the proportion of non-whites who lived in concentrated poverty compared to whites (Jargowsky, 2016), that gap is closing and is largely due to changes in the black-white neighborhood concentrations (Austin, 2013; Firebaugh & Acciai, 2016).

Although the data is somewhat inconclusive regarding the racial composition of the neighborhood and its relationship with higher education outcomes, the data regarding individual race and educational attainment is strong (Casselman, 2014; U.S. Department of Commerce, 2016; U.S. Department of Education, 2016b). There also is an abundance of data regarding the relationship between neighborhood racial composition and SES (Jargowsky, 2016) and neighborhood racial composition and residential stability (Crowder and South, 2011; Quillian, 2003; Wodtke, et al., 2011). For these reasons, and because of the changing tides in the concentration of minorities in impoverished neighborhoods, the racial composition of the neighborhood is an important covariate.

Institutional Factors

Similar to the role that the neighborhood plays in individual academic achievement, the institutions that students attend – the institutional environment, characteristics, and connection to the students – also will be related to individual outcomes (Kuh, et al., 2007; Ryan, 2004; Titus, 2004). Colleges and universities vary in resources, enrollment, and student characteristics, all of which can influence how a student performs at the given institution. What works for one student may not work for another. There are, however, measurable, institutional factors that have been shown to be related to the performance and retention of their students, including institutional demographics, campus finances, and the typical academic success of the student body. The findings, however, are not uniform across studies. The impact of certain covariates seems to differ greatly depending on the methodology, the population, the outcome, and the study's controls.

Institutional demographics. Among the more common institutional demographics studied are so-called structural-demographic features, including enrollment, racial composition of the student body, and the control of the institution (private/public) (Berger & Milem, 2000). Titus (2004) found that student body size was significant and positive in predicting student retention after controlling for several precollege and in-college covariates. The racial makeup of the student body, however, was not significant. Titus also found that the control of the institution was not significant after including the other individual and institution-level characteristics. An important point of departure between Titus' study and the current study is that Titus included in-college student behavior.

Ryan's (2004) findings contradicted much of what Titus reported; however, their methods varied. While Titus used individual outcomes and hierarchical modeling, Ryan focused on institutional outcomes. Using College Board and Integrated Postsecondary Education Data System (IPEDS) data for over 350 institutions, Ryan found that institutional size and private status both were positively related to institutional degree attainment. Ryan also found that the proportion of minority students on campus was negatively related to degree attainment while an Historically Black College or University classification was positively related. Kuh et al. (2007) also found that minority students who attend an Historically Black College or University do better academically than their minority counterparts at predominantly white institutions.

Campus finances. The study of institutional expenditures on student-level outcomes has produced varying results (see Ryan, 2004 for an overview of the literature). In his own study, Ryan found that both expenditures per full-time equivalent student on instruction and expenditures per full-time equivalent student on academic support have a strong and positive relationship with eventual institutional graduation rate (Ryan, 2004). In this particular study, the only variable with a stronger relationship with graduation rate than expenditure on instruction was the SAT scores of the incoming class (Ryan, 2004). He also looked for a significant finding between expenditures contributing to students' well-being and graduation and expenditures for administrative functions and graduation and did come up with significant results.

Using a sample of private, baccalaureate institutions, Gansemer-Topf and Schuh's (2006) findings support those of Ryan. They found that several measures of institutional expenses, including per capita expenses for instruction, academic support, student services, institutional support, and grants were predictive of both retention and graduation rates after controlling for the selectivity of the school. In fact, the expenditures account for over 60% of the variance in both outcomes.

Another interesting financial piece that is related to the current study is the athletics expenditures of the institution. Low-resource institutions that are members of the NCAA are provided financial assistance and leeway in meeting some of the academic thresholds because they historically struggle to meet the NCAA academic guidelines (Johnson, 2014; Paskus, 2012). Cunningham (2012) speculates that the disadvantage that many of these low-resource institutions face will only worsen over time because of the athletics department reliance on general university funding and the extent to which that funding is increasingly being stretched.

Aggregate institutional academic performance. Kamens (1971) put forth a hypothesis for the cause of the relationship between aggregate institutional performance and individual outcomes. He proposed that individual commitment to the institution increases the greater the prestige of the institution and the better its students and alumni do. Using a student-athlete population, institutional graduation rate has shown to have mixed results. McArdle, Paskus, and Boker (2013), for example, found when using a collegiate student-athlete population that the institutional graduation rate had a significant but negative relationship with first-year grades. One potential explanation for this was that those schools are more highly selective and therefore have more demanding first-year classes or grading structures (McArdle, et al., 2013). In another study, however, McArdle and Hamagami (1994) found that the addition of institutional graduation rate to multilevel models of student-athlete graduation was significant and positive after controlling for precollege academic characteristics.

Predicting Higher Education Outcomes

While past academic performance is the best predictor of future academic performance, there is a substantial amount of variation in college outcomes that these pre-college measures do not capture (Camara & Echternacht, 2000). Previous models that used simply the static incoming academic characteristics coupled with individual or institutional characteristics were limited in their ability to explain their outcomes.

Scholars like Alexander Astin, Vincent Tinto, Ernest Pascarella and Patrick Terenzini, and George Kuh have devoted much of their careers to uncovering factors that make up the remaining unexplained variance. Their collective works have explored extensively how in-college behaviors contribute to college outcomes (Astin, 1993; Burton & Ramist, 2001; Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008; Pascarella & Terenzini, 1991; Tinto, 1975). Their findings show that when in-college behaviors are added to models predicting college outcomes like first-year GPA or graduation, the in-college behaviors dramatically reduce the strength of the relationship of high school academic characteristics.

Tinto's model of student departure (1975) is among one of the most well-known and replicated. Kuh et al. (2007) even referred to it as enjoying "'near paradigmatic status'" (p. 13).

Tinto's model relies upon a student's behavioral expression of their commitment to the institution to predict whether that student will persist. This expression of commitment is informed by the student's background, including his or her past academic characteristics, and by the student's academic goals (Tinto, 1975). Pascarella and Terenzini (1980) used discriminant analysis to test Tinto's model and found that when measures of institutional integration and goal commitments were added to traditional high school academic and demographic variables, correct classification of 1st year persisters increased by more than 20%, and correct classification of 1st year voluntary dropouts increased by more than 40%. Astin's input-environment-outcome (I-E-O) model was a significant addition to the literature in predicting college outcomes, which he defined as the student characteristics that the institution attempted to influence (Astin, 1993). Astin's I-E-O model considers not only the behaviors of the student but the environment in which the student lived and operated as well.

Braxton (2004) expanded upon Astin's theory of involvement to also include psychosocial engagement, which is engagement in social interactions within the institution (Kuh, et al., 2008). Kuh, one of the foremost experts in the field of student engagement, used Braxton's work when he was developing his own definition of engagement, which he ultimately defined as the "time and effort students devote to activities that are empirically linked to desired outcomes of college and what institutions do to induce students to participate in these activities" (Kuh, 2009, p. 683). Student engagement and educationally purposeful activities, according to Kuh et al. (2008), include typical academic-centric activities like actively participating in class or participating in a study abroad program as well as more abstract or difficult to quantify behaviors like working hard or interacting with individuals different from yourself. The addition of measures of engagement to models already containing high school academic behavior and individual demographics improved prediction models of first-year GPA. One study saw an improvement in the explained variance go from 29% to 42%, and the correct classification of students who persisted to a second year increased from 47% to 72% (Kuh, et al, 2008).

While many of these seminal works were developed and written decades ago and still are used to better understand the covariates related to success in college, it is important to develop modern models that can successfully anticipate how a student will fare before having the in-college data. This not only aids in delivering appropriate college admissions decisions, but it also can help institutions identify students who have the potential to be successful but would benefit from interventions that would increase their likelihood of success. Burton and Ramist (2001) conducted a meta-analysis of studies that predicted cumulative college GPA and graduation of classes who were enrolled in college between 1980 and the mid-1990s. These studies used only the SAT verbal and math scores and high school record to predict college outcomes. Burton and Ramist (2001) found that a combination of SAT verbal, math, and high school record was the best predictor of college GPA when compared with any the three alone. They found a weighted average correlation with college GPA using a combination of high school record and SAT scores of 0.52. Using high school record alone produced a weighted correlation of 0.42, and test scores alone produced a weighted correlation of 0.36 (Burton & Ramist, 2001). The weighted correlations with degree attainment were not as strong. The best combination of SAT scores and high school record had a weighted correlation with degree attainment of 0.29. Most of these studies, however, use relatively small samples or special population samples (e.g., white students only or students without a disability only).

Astin, Tsui, and Avalos (1996) were able to examine the relationship between pre-admission characteristics and eventual degree attainment for 53,000 students attending over 300 institutions in 1985. They found that while just 10% of those with the lowest high school grades and test combination had earned a four-year degree by 1989, 80% of those with the highest combination of high school grades and tests had. While this provides useful information from a large sample, the work is now nearly 30 years old. A more recent study, referenced earlier, by Geiser and Santelices (2007) examined the predictive validity of high school GPA and SAT on college graduation. Both high school GPA and the SAT alone were significant in predicting 4-year graduation after controlling for parental education, family income, and school rank, but the combination of GPA and test was better than either alone.

NCAA Student-Athletes: A Special Subpopulation of College Students

Within the collegiate athlete population, there is a great deal of variability in gender, race, and academic preparation. Just as it is important to account for each of these in modeling the academic outcomes of a general college student population, it also is important to account for them when modeling the outcomes of student-athletes. Varsity athletes vary in one other important way – their sport group, particularly high-profile athletes versus athletes in non-high-profile sports. For the purposes of the discussion here, high-profile athletes are student-athletes who participate in the sports of baseball, men's and women's basketball and football.

The appropriateness of including NCAA varsity athletics into the fabric of American higher education, the role of the student-athlete on college campuses, and the academic success of student-athletes[1] have long been topics for discussion and debate (Comeaux & Harrison, 2011; Hiltzik, M., 2015; LaForge & Hodge, 2011; Levine, Etchison, & Oppenheimer, 2014; Umbach, et al., 2006). Student-athletes spend upwards of 42 hours each week on athletics activities

[1] The discussion herein will be constrained to NCAA Division I student-athletes. For simplicity, this group will be referred to as student-athletes.

(NCAA, 2016b), which has been a source of concern among some in academia. Having opportunities to explore other extracurricular activities, secure gainful employment, or participate in enriching academic activities like independent research or studying abroad often are sacrificed for their sport (Comeaux & Harrison, 2011; Fields, 2012). In student-athletes' defense, some question whether students who participate in student government, have internships, or take advantage of any of the countless opportunities on college campuses perhaps devote similar time to activities outside of their academics. Moreover, data has shown that most student-athletes *want* to be devoting this kind of time to their sport. In a survey of over 20,000 student-athletes, over one-third reported they would like to spend even more time on athletics (NCAA, 2016b). There certainly are other areas of student-athletes' experiences on campus and their role in higher education that have caused controversy and conversation. Most recently, court cases focusing on academic integrity (Bauer-Wolf, 2017), amateurism (Edelman, 2015), and student-athletes use of their own likeness (Keilman & Hopkins, 2015) have been covered extensively by popular media outlets.

Coinciding with this ongoing debate has been increasing activity by the NCAA in establishing bylaws that dictate academic benchmarks student-athletes must meet to be eligible for athletics participation. Both the debate over the suitability of universities and colleges sponsoring varsity athletics and the increased presence of the NCAA in the academic lives of their student-athletes has resulted in a good deal of research into student-athlete academic trajectories, including their pre-college academic backgrounds, in-college academic behavior, and their college outcomes. The following sections detail some of this research, including results from decades of modeling student-athlete academic outcomes.

Pre-college Academic Background

While NCAA member institutions have autonomy over student-athlete admission, the NCAA does establish minimum academic requirements, known as initial eligibility, every Division I student-athlete must meet to be immediately eligible for athletics participation. Over the years, the initial eligibility requirements have become more stringent as NCAA researchers and policy-makers have learned more about the characteristics of academically successful student-athletes. The changes made to the requirements were intended to better reflect the linkage between high school academic characteristics and college academic success. The data show, for example, that a student-athlete with a 2.5 high school core GPA and an 820 SAT "is predicted to have a roughly 38% chance of eventual graduation" (Petr & Paskus, 2009, p.85). Initial eligibility changes are made also while trying to ensure certain demographic groups are not disproportionately negatively affected.

For the majority of the membership, the NCAA admission standards do not align with the individual member institutions' requirements – for some, the NCAA standards are more stringent; although, for most, the NCAA standards are not as rigorous as their own. At these institutions, student-athletes may be admitted under special admit criteria (Espenshade, Chung & Walling, 2004; Go, 2008), and data do show that, for many, student-athletes are entering college academically less prepared than their non-athlete peers (Espenshade, Chung & Walling, 2004). Male athletes, for example, who attend an institution in the NCAA's Pac-12 athletic conference[2] have an average SAT score 172 points lower than male non-athletes. The within institution differential ranges from 92 to 309 – all favoring the nonathletes (Taylor, 2012).

[2] NCAA member institutions organize into athletics conferences which serve as the basis for most of their athletics competitions. The Pac-12 conference is made up of 12 colleges and universities located in Arizona, California, Colorado, Oregon, Utah, and Washington.

When establishing and evaluating the initial eligibility requirements, the NCAA closely examines what pre-college characteristics student-athletes who have earned their bachelor's degree had. They have found, similar to the studies reported on earlier that used a general college population, that high school coursework, high school GPA and standardized test all are predictive of eventual degree attainment. Specifically, NCAA research has found that a student-athlete who takes less than 16 non-remedial core courses in high school is at a significantly increased risk of not persisting to graduation (NCAA, 2009); although, in at least one NCAA study, the number of core courses did not significantly add to prediction models of first-year grades when controlling for high school GPA and standardized tests (McArdle, et al., 2013). Analyses also consistently have shown a small but significant added impact of standardized tests in models of college outcomes that also include high school grades (McArdle, et al., 2013; Petr & McArdle, 2012). McArdle (2008) found that using both high school GPA and ACT or SAT individually contributed to the prediction of freshman grades, freshman dropout, and eventual graduation. High school GPA was the stronger of the two predictors when college grades were the outcome, and standardized test was the stronger of the two predictors when the binary outcome of dropout or graduation was the outcome. Overall, those two predictors accounted for roughly one-quarter to one-third of the variance in the outcomes (McArdle, 2008). NCAA research also has found that using a GPA made up of grades earned in core courses, including English, math, physical science, and social science, improves the accuracy of models compared to a cumulative GPA that is calculated using all available grades (Petr & McArdle, 2012).

In-college Academic Behavior

The NCAA also establishes legislation that specifies minimum academic requirements for continued eligibility once the student-athlete is enrolled in a Division I member institution.

Known as the progress-toward-degree (PTD) requirements, student-athletes must meet GPA, credit hour, and percentage of degree completion requirements per term and per academic year. The current rules, for example, will put student-athletes on a path to graduate within 5 years of their initial enrollment. The rules state that student-athletes must earn a minimum of six semester credits each term and a minimum of 18 for the academic year. By the end of their third year of enrollment, they must have completed 60% of the requirements for a degree, and they must have 100% of the minimum institutional GPA needed for graduation (NCAA, nd). These PTD requirements have increased over the years and have been established from the results of modeling of the “in-college academic profiles of eventual graduates” (Petr & McArdle, 2012, p. 35). The current thresholds were established based on roughly 95% of current graduates meeting them.

Another measure of student-athlete in-college academic behavior is the Academic Progress Rate (APR). The APR is a team academic metric. Student-athletes who are on athletics aid are assigned term-by-term points for maintaining academic eligibility per the PTD requirements and any additional institutional standards and for retaining to the next term (See LaForge and Hodge, 2011 or http://www.ncaa.org/aboutresources/research/academic-progress-rate-explained for a detailed explanation). Failure to meet APR benchmarks results in a hierarchy of team penalties including a post-season competition ban. As noted by LaForge and Hodge (2011), the APR is a measure of academic progress, not performance. The relevance of the APR to the study here is the awareness that not only are Division I student-athletes held to rigid standards with the PTD to maintain their athletic status and their ability to compete, but that there are added incentives to meet these standards with team members and coaches relying on the individual players to do their part to avoid team sanctions.

Academic Outcomes

Division I student-athletes' GPAs, persistence, and eventual graduation all are monitored by the NCAA national office. Both PTD and APR have GPA requirements, and as described earlier, persistence is a key component of the APR. Graduation rates are the most discussed and debated student-athlete academic outcome. Part of the controversy comes from how to best define graduation, or perhaps more accurately stated, how best to define the cohort used when calculating graduation rates. Federal graduation rates are calculated using all first-time, full-time students. A student is classified as a student-athlete if they receive athletics-based financial aid. Students who transfer into an institution are not included in that institution's federal graduation rate denominator, and students who transfer out of an institution are not removed from the institution's denominator. The latter situation then renders those students as non-graduates, reducing the institution's graduation rate. In order to compare the graduation rates of student-athletes with their non-athlete peers, the NCAA continues to collect and report federal graduation rates. The NCAA also, however, reports a graduation rate that accounts for transfers in and out of the institution known as the Graduation Success Rate (GSR). The GSR produces graduation figures roughly 10-15% higher than the federal rate (Brown, 2015) signaling the proportion of student-athletes who transfer and who do so successfully.

The latest six-year federal graduation rates data show that 68% of student-athletes graduated from their initial school of enrollment compared with 66% of all students (NCAA, 2017a). Much of the work that has been done modeling student-athlete graduation has been presented earlier. To summarize those findings, the research has shown that white student-athletes are graduating at greater rates than minority student-athletes (Paskus, 2012; Petr &

McArdle, 2012); in every racial and gender subcategory, student-athletes are graduating at greater rates than their non-athlete peers with the exception of white males (NCAA, 2016c); incoming academic characteristics have a stronger relationship with first-year outcomes (GPA and retention) than they do eventual graduation (Petr & McArdle, 2012), and one study showed that standardized test was a better predictor of six-year graduation than was high school GPA (McArdle, 2008). Another important finding applicable to this study is the role of the institution in predicting student-athlete academic outcomes. A study from McArdle, et al. (2013) found that the nesting of students within colleges does account for some of the variance in first-year GPA, and one important institutional variable is overall graduation rate.

Methodological Issues Relevant to this Study

The design of this study and its research questions poses two methodological issues. The first is the use of census data and the inability to control for selection bias of families choosing neighborhoods, and the second concerns the nesting of student-athletes within institutions.

One of the greatest critiques of using geodemography to better understand variations in educational attainment is the issue of selection bias in where a person lives. One of the largest studies of neighborhood effects was the MTO experiment, which bypassed this concern by randomly assigning families to neighborhoods with varying levels of poverty. The experiment included over 4,600 families living in the cities of Baltimore, Boston, Chicago, New York, and Los Angeles. To be eligible, families with children needed to live in public housing in a neighborhood with a poverty rate of 40% or greater (Chetty et al., 2015). Families were randomly assigned to one of three groups: one group received a housing voucher that could be used to move to a low-poverty neighborhood; a second group received a Section 8 housing voucher, and the third was a control group (Chetty, et al., 2015). The families were followed over a 15-year period to assess the long-term effects of housing on several outcomes, including educational attainment. The effects on educational attainment are mixed and may be heavily reliant on the time of the move. Chetty and colleagues (2015) found a significant relationship between the randomized groups and both college attendance and the quality of the college in which the students enrolled. The findings, however, were age dependent. If the child was 12 or younger at the time of the move and was part of the low-poverty voucher group, they were significantly more likely to attend college and were more likely to attend a high-quality college than were those in the control group. Children, however, who were 13 or older at the time of the move were actually less likely to attend college and less likely to attend a high-quality college than their peers in the control group. The relative deprivation theory could help explain this if the children felt they were too far behind academically compared to their peers. When not assessing effects by the age group of the students, the final impact evaluation of the MTO experiment concluded that there was no effect for the children in either of the voucher groups in educational attainment.

Barring the opportunity to conduct experimental research, census data serves as explanatory variables intended to capture the effects of living in a certain neighborhood with its particular neighborhood characteristics. This census data, however, presents methodological challenges. Ensuring that the effects of the neighborhood are not overestimated requires controlling for family characteristics, including a full accounting of family SES and family composition (Jencks & Mayer, 1990), which the data for this study does not include. More will be discussed in Chapter 5 in the presentation of the study’s limitation.

As mentioned in Chapter 1, student-athletes represent between 1-37% of the population on Division I campuses. Multilevel modeling is a statistical method that accounts for the nesting of individuals within higher order groups. Because students who come from the same school will experience the same potential effects of their environment, these observations will not be independent. Prior to the introduction of multilevel modeling, ordinary least squares was the preferred methodology for predictive validity studies in education. Single-level models like multiple linear regression or generalized linear models, however, carry with them an assumption of independence of observations. When this assumption is violated, it potentially increases the probability of rejecting the null hypothesis even absent a statistically significant finding (Huta, 2014; Osborne, 2000). Multilevel modeling produces conservative estimates of both within- and between-group effects (Raudenbush & Bryk, 2002). The following chapter will provide an overview of the methodology, which will include a broader discussion of multilevel modeling.

CHAPTER THREE

IN WHICH THE METHODOLOGY IS PRESENTED

At least one-third of student-athletes are not graduating from their initial institution of enrollment within six years, and roughly 15% fail to graduate after transferring to a different Division I school (NCAA, 2016e). Academic difficulties are often the cause of transferring to a different institution or dropping out altogether (Stinebrickner & Stinebrickner, 2013). Identifying academically at-risk students-athletes at the time of enrollment and directing applicable academic support services to them may help some persist to graduation. While high school academic performance paired with individual demographics help to predict a student's initial academic adjustment to college and whether a student will eventually graduate, there is a good deal of unexplained variance left in these prediction models (Burton & Ramist, 2001; Geiser & Santelices, 2007; McArdle, Paskus & Boker, 2013; Petr & Paskus, 2009; Pike & Saupe, 2002). One piece of data that may help to reduce this unexplained variance is the characteristics of the neighborhood in which the student lives prior to enrolling in college. While the literature on the relationship between neighborhood characteristics and academic outcomes is much more established among high school academic outcomes than college; the limited work that has been done with college outcomes shows promise (Leventhal & Brooks-Gunn, 2000).

The goal of this study was to investigate the added value in using census data in modeling collegiate academic outcomes among a group of NCAA Division I student-athletes. While this study was informed by the theories put forth by Jencks and Mayer (1990) including the collective socialization theory, the institutional model, and the epidemic model, it did not seek to prove any of the theories related to neighborhood effects. Instead, it was an exploratory study examining if and how much more of the variability in student-athlete's first-year GPA, first-year retention and six-year degree attainment is explained with the addition of census information. In addition, this study focused on examining the added benefit in using census data in modeling collegiate academic outcomes among student-athlete subgroups, including minority student-athletes and student-athletes in sports considered to be at a greater academic risk.

The research questions and hypotheses that guided this analysis include:

1. Are U.S. census block group data significantly related to NCAA Division I student-athlete first-year college cumulative grade point average, first-year retention and eventual six-year degree attainment after controlling for student-athlete demographics and pre-college academic characteristics and college-level institutional characteristics?

It is hypothesized that after controlling for student-athlete demographics and pre-college characteristics that the neighborhood characteristics of SES, racial composition and residential stability will significantly contribute to first-year GPA, first-year retention and six-year graduation. It is further hypothesized that these neighborhood characteristics will remain significant after the inclusion of college-level institutional information.

2. Do U.S. census block group data relate to NCAA Division I student-athlete first-year college cumulative grade point average, first-year retention and eventual six-year degree attainment differently for student-athletes who participate in academically at-risk sports and their counterparts in sports not deemed academically at-risk?

The college athletics literature has not explored whether there is predictive bias in academic outcomes models by the academic risk status of the student-athlete's sport. It is hypothesized that the neighborhood characteristics of SES, racial composition and residential stability will be related to student-athletes' first-year GPA, first-year retention, and six-year graduation comparably for student-athletes participating in academically at-risk sports and their counterparts in sports not deemed academically at-risk.

3. Do U.S. census block group data relate to student-athlete first-year college cumulative grade point average, first-year retention and eventual six-year degree attainment differently for minority student-athletes and white student-athletes?

There is very limited information on predictive bias by race within the neighborhood effects literature. Vartanian and Gleason (1999) did find that neighborhood characteristics have a stronger relationship with college graduation for whites than they do for blacks; however, this finding was conditional on familial characteristics, which are not accounted for in this study. In another study, Crowder and South (2011) found that prolonged exposure to a disadvantaged neighborhood had a greater negative relationship with high school graduation among a white sample than a black sample. In the college athletics literature, there has been no evidence of predictive bias by race in academic outcome models of eventual degree attainment (McArdle, 2008). With this limited information, it is hypothesized the neighborhood characteristics of SES, racial composition and residential stability will be related to student-athletes' first-year GPA, first-year retention, and six-year graduation comparably for white and non-white student-athletes.

Population and Sample

The population of interest in this study is all first-year student-athletes whose initial year of enrollment was the 2009-10 academic year and who enrolled in and competed for an NCAA Division I institution. Two of the research questions focus on minority and academically at-risk subpopulations of student-athletes. For the purposes of this research, minority student-athletes are defined as any student-athlete who is non-white including student-athletes of two or more races, one of which may be white. Student-athletes in the sports of football, men's and women's basketball, and baseball are considered to be at greater academic risk than their counterparts. According to NCAA research on academic risk among entering student-athletes, there are four categories that can be used to assess risk at the time the student-athlete first enrolls in a postsecondary school, including pre-college academic characteristics, the role of academics in the student-athlete's life, personal history (e.g., first generation college student, financial hardship, etc.), and sport characteristics (e.g., high profile sport, team culture does not emphasize academics, time demands of sport, etc.) (NCAA, 2009). The sports of football, men's and women's basketball, and baseball are considered high profile sports, and all have historically lagged behind their counterparts in team academic metrics (NCAA, 2017b). Additionally, baseball and football student-athletes have among the highest athletic time demands of any Division I sport (NCAA, 2016 January); baseball and men's basketball student-athletes have comparatively the strongest athletic identities (NCAA, 2011), and football student-athletes are most likely to be a first-generation student (NCAA, 2016 January).

Secondary data containing over 1.8 million student-athletes who graduated high school between 1986 and 2010 were made available to the author. The 2009-10 academic year was utilized in the current study because it represents the most recent collection year with the most complete data that allows for a six-year graduation timeline to be examined. The initial sample of prospective student-athletes who completed high school in the 2008-09 academic year totaled 74,373.

The dataset was subset to students who had both high school and college academic data (37%) and excludes students who attended a Division II or III institution but participated in a Division I sport (4%), transferred into their initial institutions (<1%), and student-athletes who attended an institution that provides athletics-based aid but did not receive aid their first year (16%). To be included in the data reporting, student-athletes must have received athletics-based financial aid[1]. NCAA reporting requirements state that Division I institutions must report term-by-term academic data for all student-athletes who receive athletics-based financial aid. Therefore, student-athletes who attended an institution that grants athletics-based financial aid and did not receive athletics aid in their first year were not part of the initial reporting cohort and will not have first-year outcomes. It is important to note that it is possible for a student-athlete to lose his or her athletics aid and therefore be removed from the cohort.

College-level information also is included in this study. After accounting for records with missing data, the final sample consisted of 18,417 student-athletes from 327 institutions, which is a near census of the 333 institutions in the division during 2009-10 academic year. On average, there were 56 student-athletes per institution; however, the range was 8 to 170. In 2009, 98

[1] Military academies and institutions in the Ivy League do not offer any scholarships based on athletics. Student-athletes who play on teams that do not offer athletics-based aid are included in the cohort if they were a member of the team on or after the first date of competition (NCAA, 2017c).

shools within Division I did not sponsor football, which carries with it a large roster size, and contributes to the wide range in aided student-athletes entering in the 2009-10 academic year.

Table 1 provides more information regarding the institutional characteristics of the sample. The average federal graduation rate was 66%, and as the standard deviation figures indicate, there is a wide range with regards to the enrollment and financial variables.

Table 1. Institutional Characteristics

	Statistic	SD
Percent Minority Serving Institution	5.7%	
Percent Private Institution	31.0%	
Mean Undergraduate Enrollment	13,307	9080
Mean Out of State Total Cost	$36,226	9847
Mean Out of State Grant in Aid	$32,151	10978
Mean Grand Total Athletics Expenses	$32.0 million	27.0 million
Six Year Federal Graduation Rate for 2009 Freshman Class (Full Student Body)	66.1%	18.4
NOTE: Data represents institutional statistics for those institutions (N=327) included in the sample.		

The student-level sample demographic characteristics are presented in Table 2. The final sample had a slight overrepresentation of female, and white student-athletes and an underrepresentation of non-resident, international students and student-athletes with a race/ethnicity of unknown. This underrepresentation likely is due to restricting the sample to student-athletes with U.S. census data.

Table 2. Sample Demographics of NCAA Division I First-Year Students in Fall 2009

		Sample Statistics % (N)	Division I Student-Athlete First Year Population % (N)
Race	American Indian/Alaskan Native	0.5% (98)	0.5% (135)
	Asian	1.5% (276)	1.5% (409)

	Native Hawaiian/Pacific Islander	0.7% (121)	0.6% (161)
	Black	23.5% (4,319)	20.8% (5,667)
	Hispanic/Latino	4.2% (779)	4.2% (1,134)
	White/Non-Hispanic	67.2% (12,379)	63.5% (17,289)
	Non-Resident International	0.3% (55)	2.8% (772)
	Two or More Races	2.1% (390)	1.9% (523)
	Unknown	--	4.2% (1,132)
Gender	Male	50.9% (9,373)	52.5% (14,287)
	Female	49.1% (9,044)	47.5% (12,935)
NCAA Sport Group	High Profile (Men's/Women's Basketball, Baseball, Football)	35.0% (6,442)	33.8% (9,190)
	All other sport groups	65.0% (11,975)	66.2% (18,032)
NOTE: Data represents all Division I student-athletes who were first-time, first-year students in the 2009-10 academic year and received athletics based financial aid.			

Regarding average incoming academic characteristics, the sample had a slightly lower incoming test when compared with the population (see Table 3). The sample's mean HSCGPA was just under a 3.3, and the average test score in SAT units was a 1063.

Table 3. High School Academic Characteristics of NCAA Division I First-Year Students in Fall 2009

	Sample Statistics		Division I Student-Athlete First Year Population	
	Mean	SD	Mean	SD
Total Core Course Units	17.64	2.19	17.66	2.31
High School GPA in Core Courses	3.29	0.55	3.31	0.56
Best Standardized Test (SAT or ACT) on SAT scale	1063	164	1070	167
NOTE: Data represents all Division I student-athletes who were first-time, first-year students in the 2009-10 academic year and received athletics based financial aid.				

The sample neighborhood statistics, including variables that make-up neighborhood SES, as well as the racial/ethnic composition of the neighborhood and residential stability, were comparable to the population statistics as seen in Table 4.

Table 4. Neighborhood Characteristics of NCAA Division I First-Year Students in Fall 2009

Overarching Category	Subcategories	Variables	Sample Statistics		Division I Student-Athlete First Year Population	
			Mean	SD	Mean	SD
Neighborhood SES	Neighborhood Education	% 25yr old + Male without HS Diploma	13.8%	10.8	13.7%	10.8
		% 25yr old + Male HS Graduate	23.4%	11.1	23.3%	11.4
		% 25yr old + Male with Some College	27.0%	7.8	26.8%	7.9
		% 25yr old + Male Bachelor's Degree	21.7%	10.9	21.8%	11.0
		% 25yr old + Male Advanced Degree	14.1%	10.1	14.4%	10.8
		% 25yr old + Female without HS Diploma	13.6%	9.9	13.5%	10.0
		% 25yr old + Female HS Graduate	26.5%	10.0	26.5%	10.2
		% 25yr old + Female with Some College	29.7%	7.3	29.6%	7.6
		% 25yr old + Female Bachelor's Degree	19.9%	10.0	20.0%	10.1
		% 25yr old + Female Advanced Degree	10.2%	7.0	10.4%	7.1
	Employment	% 16yr old + Male Employed Full-Time	69.6%	8.8	69.6%	8.9

		% 16yr old + Female Employed Full-Time	46.2%	8.5	46.2%	8.7
	Head of Household	% Married Couple Family	80.8%	10.8	81.0%	10.8
	Income	Average Median Household Income	$57,787	23086	$57,858	23,536
Racial/Ethnic Composition of Neighborhood		% Non-White	20.0%	20.7	19.6%	20.4
Residential Stability		Average Median # Years Unit Occupied by Householder	6.2	3.0	6.26	3.1
NOTE: Data represents all Division I student-athletes who were first-time, first-year students in the 2009-10 academic year and received athletics based financial aid.						

Finally, the outcome variables within the sample and the population were comparable for each of the three outcomes: mean cumulative first-year GPA, first-year retention, and six-year bachelor's degree attainment (see Table 5).

Table 5. Outcome Variables for NCAA Division I First-Year Students in Fall 2009

	Sample Statistics		Division I Student-Athlete First Year Population	
	Statistic	SD	Mean	SD
Mean First Year Cumulative GPA	2.89	0.63	2.89	0.64
% Retained to 2^{nd} Year	81.4%	38.9	82.0%	38.4
% Earned Bachelor's Degree Within 6 Years	56.6%	49.6	57.5%	49.4
NOTE: Data represents all Division I student-athletes who were first-time, first-year students in the 2009-10 academic year and received athletics based financial aid.				

Data Collection and Instrumentation

The final dataset that was used for this research was compiled by the author from 7 different sources. Three of the datasets offered information on the NCAA colleges and universities, including a graduation rates file that contained both student body and student-athlete federal graduation rates for the institution, a financial data file that contained detailed institutional financial data, and an institutional characteristics file that contained demographic information such as public/private status. The NCAA research department provided each of the datasets. NCAA research also provided the student-athlete high school academic data from the NCAA Eligibility Center (EC) and college academic data from the Academic Performance Program (APP). The U.S. census datafile was provided by Dr. Steve Boker, an NCAA independent research consultant. One additional variable was accessed by the author. Minority Serving Institution (MSI) status was pulled from the U.S. Department of Education website and is generated from the Integrated Postsecondary Education Data System (IPEDS).

The development of the master dataset started with receipt of 2000 Census Summary File 1 (SF1) data and 2000 Census Summary File 3 (SF3) data. The Census 2000 SF1 and SF3 files each contained data for 1,808,240 prospective student-athletes who applied to the NCAA EC between the years of 1986 to 2012. The prospective student-athletes were identified by an NCAA-generated student identifier, NCAA_ID, that follows the student-athletes across NCAA Division I institutions should they transfer or stop-out and resume enrollment and athletics competition on scholarship at another NCAA Division I institution. The SF1 and SF3 files were merged on NCAA_ID to form one wide census file.

Dr. Boker also provided a partial NCAA EC datafile for the same population of students in the census files. The student-level identifier, NCAA_ID, was used to merge this data with the census data. The EC data provided the PSA's high school graduation date. The majority of the population had a high school graduation year between 1994 and 2009 (99.6%). Selecting on the high school graduation year of 2009 reduced the census file to 74,373 individuals. The prospective student-athletes' standardized test scores (SAT or ACT) were missing from the EC file used for this initial merge with the census data. This variable was merged into the file using the NCAA research department's EC file.

NCAA research staff also provided a six-year longitudinal NCAA APP datafile (N=32,959) for NCAA Division I student-athletes who were first time college enrollees in the 2009-10 academic year. There was an 83% match rate to the census and EC file. The file also used NCAA_ID as a student identifier, which facilitated the merge between the APP data and the census and EC data.

The next step brought in the college-level institutional characteristics and financial data as well as graduation rates data. The NCAA assigns a unique institutional ID, INSTID, that can be used to merge these institutional-level variables. The INSTID is included in the APP dataset and was used to merge in the institutional characteristics, financial data, and graduation rates. Both the institutional characteristics data and graduation rates files had data for the entire Division I membership. The financial data was missing information for four schools.

Finally, the MSI designation was hand-entered by the author based on institution name.

Table 6 provides the total number of cases in each file and the match rates for each merge. The final sample was generated after taking missing data into account, which is reflected in the table.

Table 6. Datafiles that Comprise the Master Dataset

Datafile	N	Merge	Resulting N
Census SF1	1,808,240	--	--
Census SF3	1,808,240	100%	1,808,240

EC (provided by Dr. Boker)	1,808,240	100%	1,808,240
Subset data to just the entering first-year class of 2009 cohort			74,373
Standardized Test from EC file provided by NCAA	137,288	99.3%	73,887
APP	32,959	82.6%	27,222
NCAA institutional characteristics	Institutional N = 333	100%	Data was subset to just those who attended a Division I institution N=26,082
Graduation Rates Data	Institutional N = 333	100%	26,082
IPEDS	Institutional N = 333	100%	26,082
Financial Data	Institutional N = 329	98.4%	25,652
Subset data to just non-transfers into initial institution			25,560
Subset data to just those who were aided as first year students			21,496
Excluded cases with missing data. Final sample.			18,417

The final dataset contained 1,662 variables. The data collection began at their freshman year of high school and concluded when either the student-athlete graduated, retired from Division I competitive athletics, lost his/her athletics scholarship, transferred to an institution that is not an NCAA Division I member, stopped out and did not return to an NCAA Division I institution as a scholarship student-athlete, or dropped out of college entirely. Not only are these data the most comprehensive accounting of the academic lives of student-athletes, but they also represent among the most comprehensive datasets of college students at-large in the United States (Petr & Paskus, 2009). It should be noted that the author is employed full-time by the NCAA research department and has access to these data as part of her normal work responsibilities.

Using information from the literature, the final dataset was widdled down to 27 variables deemed the most salient to the research questions. More detailed information regarding each of the data sources follows.

2000 Census Data

Data made available for the purposes of this research included both SF1 and SF3 2000 Census of Population and Housing data. The SF1 data come from the census questions asked of all people. It is referred to as “the 100-percent data” (U.S. Department of Commerce, 2001a, p. 1-1). The SF3 data are sample data, which represent information asked of a sample of the population. The SF3 data was used for this study. The census data merge with the individual student-athlete was done by an NCAA consultant, Dr. Steve Boker. For the purposes of the merge, Dr. Boker used the address the student-athlete submitted on his/her application to the NCAA Eligibility Center, which is required for any prospective student-athlete interested in competing in Division I or II athletics.

Census data typically is analyzed at one of three levels. The smallest is the block level, which, in order of size, is followed by block groups, and then census tracts. The block levels are formed by “streets, roads, railroads, streams and other bodies of water, other visible physical and cultural features, and the legal boundaries shown on Census Bureau maps” (U.S. Department of Commerce, 1994, p. 11-1). For the 2000 census, there were 8,269,131 block levels including U.S. territories (U.S. Census Bureau, nd). Block groups are clusters of block level areas. Block levels are not split among block groups but are wholly contained within a group. The 2000 census contained 211,827 block groups (U.S. Census Bureau, nd), and each was comprised of between 600 and 3,000 individuals (Iceland & Steinmetz, 2003). Census tracts, which contain, on average, 4,000 individuals are comprised of several block groups. They were established to be fairly stable over time to allow for trend analyses across census collection periods. They also were “designed to be relatively homogenous with respect to population characteristics, economic status, and living conditions” (Iceland & Steinmetz, 2003). For the purposes of this research, block group data from the student-athletes’ home neighborhoods was used. This is the smallest unit of data collection that includes neighborhood SES information as well as demographic and housing information.

NCAA Eligibility Center Data

The NCAA EC is an arm of the NCAA that is responsible for determining whether a prospective student-athlete meets the athletic and academic guidelines to participate in NCAA sanctioned Division I or Division II sports. This process is referred to as certifying the prospective student-athletes. The EC has been in operation since 2006. From 1994-2006, the Initial Eligibility Clearinghouse (IEC) was run by the ACT and performed the same duties that now are completed by the EC.

Both high schools and prospective student-athletes submit data to the EC, which includes individual-level demographic information, a comprehensive accounting of the prospective student-athlete’s high school educational records, and a comprehensive accounting of their athletics participation. Data made available to the NCAA research department includes the number of credits earned in core academic courses, which include non-remedial coursework in English, math, physical and social sciences, and other areas such as comparative religions and foreign languages. It also includes the HSCGPA, the prospective student-athlete’s best SAT and ACT scores and all applicable subscores, final EC eligibility determination, and all reasons for an ineligible finding. Since its inception, the initial eligibility process has generated over 2 million records of high school performance that is available for NCAA research purposes (Petr & Paskus, 2009). For the purposes of this study, 119 EC variables for the 74,373 students who enrolled in the 2009-10 academic year were made available to the author.

NCAA Academic Performance Program Data

The NCAA APP is a data collection program that began in 2003 and requires all NCAA Division I institutions to submit term-by-term academic data to the NCAA for all student-athletes who receive athletics-based financial aid. The APP data include cumulative and term-by-term (including summer session) GPA, credits attempted, credits earned, retention or graduation information, degree of study, sport(s), and any reasons a student-athlete may have been deemed ineligible to participate in athletics at the conclusion of the term. Student-level demographic information and institutional characteristics also are included. Student-athletes have unique identifiers that follow them as long as they maintain enrollment and remain on a roster and receive an athletics scholarship from an NCAA Division I institution. Should a student-athlete stop out or transfer and resume their education at a Division I institution and receive athletics-based aid, their unique identifier will reappear in the data. This allows for the creation of a longitudinal dataset that can model longer-term outcomes across institutions. Should a student-athlete transfer into an NCAA Division I institution from outside the division, high-level transfer data is included in the APP data, including the total number of credits transferred as well as credits in each major subject, transfer GPA, and any remediation that may be needed.

Data are collected in the fall following the academic year of interest. The data can be delivered manually, via a text file import or an import from Compliance Assistant, which is an NCAA collection system.

NCAA Division I Financial Data

As part of their Division I membership requirements, institutions submit annually detailed information on both their general university (not restricted to athletics) and athletics-specific expenses and revenues. Total revenue and expenses are included in the data as well as sport-level revenue and expenses, including detailed information regarding coaches' salaries, scholarships awarded, and the sources of the revenue and targets of the expenses. In addition to the financial data, information on university personnel, enrollment, and sports participation also is included. In all, there are over 3,500 reported and derived variables in this dataset. The institutions are identified in the datafile with their unique NCAA institutional identifier.

NCAA Graduation Rates Data

NCAA Division I institutions are required to report to the NCAA their federal graduation rates for both their student-body and their student-athletes. Also included in this file is the institution's graduation success rate, which is an NCAA calculation of graduation that removes from the denominator student-athletes who transfer out of the institution and includes those who transfer into the institution.

NCAA Institutional Characteristics Data

NCAA institutional demographic data, including public/private status, NCAA division and subdivision, institutional name and contact information are housed in a central database and kept by the NCAA research department.

IPEDS

The Integrated Postsecondary Education Data System is a data warehouse of institutional information for every postsecondary school that receives Title IV funding. Information included in IPEDS includes enrollment, graduation, personnel, cost and financial aid data. This information is made available to the public and can be downloaded for free from the IPEDS website. The institutions are identified with a U.S. Department of Education identifier, UNITID. The NCAA has a crosswalk file that ties the UNITID to the NCAA_ID to permit merges of IPEDS data into the NCAA institutional datafiles.

Variables for this Study

The variables for this study occur at two levels: student-level and institution-level. The variables are grouped into one of the four categories: student-level demographic information, student-level pre-college academic data, student-level neighborhood data and institution-level information.

Student-level Demographic Data

Three student-level demographic variables are included in the modeling. All were reported by the student-athlete’s institution as part of the APP reporting. Race is a multinomial variable with the following categories: American Indian/Alaskan Native, Asian, Native Hawaiian/Pacific Islander, Black, Hispanic/Latino, White/Non-Hispanic, Non-Resident Alien, two or more races, and unknown. The variable was then dichotomized into white and non-white with white as the referent. Gender is a dichotomous variable with female as the referent. Finally, the sport groups are presented in the data by sport and gender. For example, women’s tennis or men’s ice hockey. These sport groups were dichotomoized into high profile and other sports.

High profile consists of baseball, football, and men's and women's basketball. All other sports are classified as other, and this group served as the referent so that the relationship between the outcomes and high-profile sports participation could be assessed directly.

Pre-College Academic Data

The EC data served as the source for the three pre-college academic variables. The total core course units includes the number of non-repeated, non-remedial courses taken in the following subjects: English, math, physical science, social science, comparative religion, and foreign language. To be eligible for NCAA athletics, prospective student-athletes must earn credits in 16 core courses. The calculation of the HSCGPA includes the 16 minimum core and then any additional core courses that will aid the student-athlete's HSCGPA. The HSCGPA for the sample ranges from 1.88 – 5.00 with a mean of 3.29 and a standard deviation of 0.55. Finally, the NCAA accepts both the SAT and the ACT for eligibility decisions, and within the membership, both the SAT and the ACT can be used for admission purposes. NCAA research has used a concordance table that assigns an SAT value for all ACT scores. Academic modeling done by the NCAA uses the best test on the SAT scale. The modeling here will follow suit. The best test range for the sample is 540 – 1600 with a mean of 1063 and a standard deviation of 164.

Neighborhood Data

All neighborhood variables were measured at the block group level. Based on a review of the literature, the neighborhood characteristics included are socioeconomic status (SES) of the neighborhood, residential stability in the neighborhood (Aaronson, 1998; Ensminger, Lamkin & Jacobson, 1996; Leventhal & Brooks-Gunn, 2004; Sharkey and Faber, 2014), and racial composition (see chapter two for more information).

Socioeconomic status was measured with nine variables that quantify the educational attainment of the neighborhood, employment, head of household, and income. The educational attainment of the neighborhood was represented with four different variables: 1) the proportion of males 25 years or older without a high school diploma; 2) the proportion of males 25 years or older with a bachelor or advanced degree; 3) the proportion of females 25 years or older without a high school dimploma, and 4) the proportion of females 25 years or older with a bachelor or advanced degree. The proportion of individuals 25 years or older without a high school diploma is a variable in the census data. The mean for males within the sample is 13.8% (SD=10.8), and for females, it is 13.6% (SD=9.9). The proportion of individuals 25 years or older with a bachelor or advanced degree are derived variables captured by summing the proportion of individuals 25 years or older (males and females separately) who earned a bachelor degree and the proportion of individuals 25 years or older who earned an advanced degree. The mean for males within the sample is 35.8% (SD=19.8), and for females, it is 30.1% (SD=15.5). Scholars are divided regarding whether it is the presence of poverty and lack of positive role models or the presence of affluence and the abundance of positive role models that is most relevent in individual academic outcomes (Brooks-Gunn et al, 1993; Duncan, 1994; Harding, 2003; Wilson, 1987). Because of this, measures of both deprivation and abundance will be included.

The remaining SES variables all come directly from the census file with no manipulation needed. Neighborhood employment is measured with the proportion of individuals who are 16 years of age or older who work full-time. This measure is separated by gender. The mean proportion within the sample of males and females who are 16 years or older and are employed full time is 69.6% with a standard deviation of 8.8 and 46.2% with a standard deviation of 8.5

respectively. Head of household is measured using the proportion of households who have a married head of household. In the sample data, the mean proportion for this is 80.8% (SD=10.8). Finally, income is measured using median household income. The average of the median household incomes in the sample is $57,787 (SD=23086).

The two remaining areas captured by the census data, residential stability and neighborhood racial composition, are measured with one variable each. The median number of years a dwelling is occupied by a householder captured residential stability. For the sample, the average is 6.2 years with a standard deviation of 3.0. The proportion of the neighborhood that is non-white captured the neighborhood residential composition. This is a derived variable summed from the proportion of all non-white and multi-racial figures. The sample mean is 20.0% with a standard deviation of 20.7.

Institutional Data

Institutional demographic characteristics, financial data and federal graduation rates came from an NCAA institutional characteristics datafile, NCAA revenues and expenses data, and the NCAA graduation rates data. One additional variable, MSI, came from the IPEDS. Minority-serving institution is a federal designation that is assigned either based on the percentage of minority student enrollment or if the institution is a legislated Historically Black College or University or a Tribal College and University. If an institution is neither but has an enrollment of at least 25% of Black, Hispanic, Asian/Pacific Islander, or American Indian/Alaskan Native students “while students of all other individual minority groups each constitute less than 25 percent of the total undergraduate enrollment,” they can be classified as an MSI (U.S. Department of Education, 2007, p. v). Furthermore, an institution that does “not fit any of the above categories but in which minority students as a whole constitute at least 50 percent of the total undergraduate enrollment" are considered a MSI (U.S. Department of Education, 2007, p. v). The variable is dichotomous with not being a MSI as the referent. Other institutional characteristic data include the public/private status of an institution, which is a dichotomous variable with public serving as the referent, total undergraduate enrollment, which is comprised of the total number of full-time undergraduate students who enrolled in the fall term.

Three additional institutional variables come from the NCAA financial data. Total out-of-state cost reflects the total cost to attend the institution for out-of-state students, including tuition, room-and-board, books and supplies, and miscellaneous expenses, which are calculated by the institution and often include a travel allowance. This is a continuous variable with a range for the sample of $1 to 57,861 with an institutional mean of $36,226. The out-of-state grant-in-aid is a continuous measure of mean grant aid gifted to out-of-state students. This measure includes all institutional monies as well as Federal dollars. It does not include loan dollars or work-study dollars even if part of a financial aid package. The range for the sample is $0 – 56,681 with an institutional mean of $32,151. Finally, total athletics expense is the total amount the institution spent on the athletics program. Included in this figure is all salaries for administration and coaches, facilities expenses, travel expenses, the budget for athletics scholarships and student-athlete support services, etc. The range for the sample is $3.17 million to 130 million with a mean of roughly $32.0 million.

Outcomes

The study examined three outcomes of interest: first-year cumulative GPA, first-year retention, and six-year degree attainment. First-year cumulative GPA is the student's cumulative

GPA at the conclusion of his or her first year in college. It is measured on a 4.0 scale with a sample range of 0.0 – 4.0 and a sample mean of 2.89. First-year retention is a derived variable based on whether a student was enrolled full-time and receiving athletics aid in the 2010-2011 academic year. It was coded as a dichotomous variable with not retained as the referent. Finally, bachelor's degree attainment within six years is a derived dichotomous variable based on term-by-term graduation data for the student-athlete. If a student-athlete was coded as a graduate or a graduate student in any term within the six years that is included in this dataset, they were classified as having earned his/her bachelor's degree within six years. Type of measurement and a short description of each variable are presented in Table 7.

Table 7. Summary of Model Variables with Corresponding HLM Level

<table>
<tr><th colspan="2">Variable</th><th>Type of Measurement</th><th>Description</th></tr>
<tr><td colspan="4">Student-Level Demographics</td></tr>
<tr><td colspan="2">Individual Race</td><td>Dichotomous</td><td>White=0; Non-white=1</td></tr>
<tr><td colspan="2">Individual Gender</td><td>Dichotomous</td><td>Female=0; Male=1</td></tr>
<tr><td colspan="2">Individual Sport group</td><td>Dichotomous</td><td>Non high profile=0; High profile (M/WBB, MBA, MFB)=1</td></tr>
<tr><td colspan="4">Pre-College Academic Data</td></tr>
<tr><td colspan="2">Individual Total Core Course Units</td><td>Continuous</td><td># of units in core high school courses
Range: 6.50 – 30.00</td></tr>
<tr><td colspan="2">Individual High School GPA in Core Courses</td><td>Continuous</td><td>Range: 1.88 – 5.00</td></tr>
<tr><td colspan="2">Individual Best Standardized Test (SAT or ACT) on SAT scale</td><td>Continuous</td><td>Range: 540 – 1600</td></tr>
<tr><td colspan="4">Neighborhood Data</td></tr>
<tr><td rowspan="2">Neighborhood Education</td><td>% 25yr old + Male without HS Diploma</td><td>Continuous</td><td>Range: 0 – 100</td></tr>
<tr><td>% 25yr old + Male with Bachelor or Advanced Degree</td><td>Continuous</td><td>Range: 0 – 100</td></tr>
</table>

	% 25yr old + Female without HS Diploma	Continuous	Range: 0 – 100
	% 25yr old + Female with Bachelor or Advanced Degree	Continuous	Range: 0 – 100
Neighborhood Employment	% 16yr old + Male Employed Full-Time	Continuous	Range: 0 – 100
	% 16yr old + Female Employed Full-Time	Continuous	Range: 0 – 100
Neighborhood Head of Household		Continuous	Proportion of households with a married couple head. Range: 0 – 100
Neighborhood Income		Continuous	Range: 3,804 – 200,001
Neighborhood Racial Composition		Continuous	Proportion non-white Range: 0 – 100
Residential Stability		Continuous	Median # Years Unit Occupied by Householder Range: 0 – 31
Institutional-Level Characteristics			
Minority Serving Institution		Dichotomous	1=MSI; 0=Not MSI
Private Institution		Dichotomous	1=Private; 0=Public
Total Undergraduate Enrollment		Continuous	Range: 1,448 – 45,490
Federal Graduation Rate		Continuous	Range: 11 – 97%
Out of State Total Cost		Continuous	Range: 1 – 57,681
Out of State Grant in Aid		Continuous	Range: 0 – 56,681
Total Athletics Expenses		Continuous	Range: 3.2 million – 130 million
Outcome Variables			
Mean First Year Cumulative GPA		Continuous	Range: 0 – 4.0
Retained to 2nd Year		Dichotomous	0=Not retained to 2nd year; 1=retained to 2nd year
Bachelor's Degree Attainment Within 6 Years		Dichotomous	0=Did not earn Bachelor's within 6 years; 1=Earned Bachelor's within 6 years

Ethical Considerations

The ethical considerations surrounding this research focus on two areas. The first is the impact on the human participants, and the second is data security. The involvement of the student-athletes poses less than minimal risk to them. The student-athletes initially contact the NCAA EC. They provide their demographic information and sport involvement history. They also consent to their high schools supplying the EC with their detailed academic records. In their application to the EC, the student-athletes sign a consent, part of which stipulates the following: “I further understand and agree that the information provided to the NCAA Eligibility Center and the NCAA may be used for NCAA Eligibility Center and NCAA research concerning athletics eligibility, the academic preparation and performance of student-athletes, and other related research purposes. I also understand and agree that such research may be published or distributed to third parties, but that I will not be identified in any such published or distributed data.” The risk for the student-athletes in applying to the EC and releasing this demographic, academic, and sport history information is that they will be deemed ineligible for NCAA Division I athletics participation. Once this information has been released to NCAA research, no additional participation from the student-athletes is needed.

The master dataset contains sensitive student-level academic data as well as detailed institutional financial data, including salary information. The student-athletes have their own unique NCAA-generated number that enables merging between datasets. There is no other student-level identifier in the data, however. Similarly, institutions are identified by an NCAA-generated institutional number. The master dataset resided on an encrypted, password protected external hard drive. The working files were reduced to just those key variables needed for the analyses, and these also resided on an encrypted password protected external hard drive.

Upon approval of the dissertation proposal, an application to the Loyola IRB was submitted with a request for an expedited review, per the recommendation of the Loyola IRB coordinator. A copy of the consent that the student-athletes sign in their application to the EC was submitted with the IRB application. The IRB approved the application on January 10, 2018.

Data Analysis

There were three primary stages to the analysis. The first was to evaluate the covariates to ensure that multicollinearity would not impact the results and to ensure that the variables intended to be used in the models were not unduly skewed. If either was an issue, data reduction and data transformations could be employed prior to analyses. The second was to ensure, through individual inferential analyses, that the covariates were significantly related to the outcomes of interest. If no relationship was found at this stage, the covariates would not be included in the modeling. Finally, multilevel modeling was used to assess the relationship between the covariates and the outcomes of interest to determine if, after controlling for individual and institutional characteristics, the characteristics of the neighborhood had a relationship with the academic outcomes. Because of the large sample size, the more conservative type I error rate of .01 was used to determine statistical significance and to help mitigate the risk of making a type I error, and effect sizes were reported.

To assess the relationship between the student-level demographic variables and first-year GPA, independent sample t-tests were used. To assess the relationship between the student-level demographic variables and the two dichotomous outcomes, first-year retention and six-year graduation, chi-square tests were done. The relationship between the student-athletes' high school academic characteristics, including total core courses taken, HSCGPA, and best standardized test and the outcome first-year GPA were evaluated using bivariate correlations, and independent sample t-tests evaluated these incoming academic characteristics' relationship with the dichotomous outcomes. Finally, all the neighborhood characteristics are measured on a continuous scale. The relationship between these and first-year GPA were assessed using bivariate correlation, while independent sample t-tests again were used to examine the relationship between neighborhood characteristics and first-year retention and six-year graduation. If these analyses yielded statistically significant results, the independent variables were included in the preliminary multilevel models.

As discussed in chapter two, colleges and universities have unique missions, enroll student bodies with distinct academic characteristics and have different financial resources at their disposal to aid students both financially and in terms of personnel who assist in their academic pursuits. The 18,417 student-athletes who comprise the sample for this research are nested within just 327 higher education institutions. Not accounting for this nesting or grouping of students violates the key assumption of independence of error terms required of both single-level multiple linear regression and generalized linear models. Prior to the introduction of hierarchical linear modeling, nested data structures typically were either disaggregated to analyze both level 1 and level 2 variables at level 1 or aggregated to analyze level 1 and level 2 variables at level 2. Disaggregation of data does not account for group differences and violates the assumption of independence of error while aggregation ignores the individual characteristics' relationship with the outcomes and can result in significantly distorted findings regarding the relationship among variables (Woltman, Feldstain, MacKay & Rocchi, 2012).

Multilevel modeling that considers both within and between group variance was adopted for this study. In this study, the level one variables consist of all student-level predictors, including demographics, incoming academic characteristics, and neighborhood characteristics. The level two variables are the institution-level variables (see Table 7). Hierarchical linear modeling (HLM) is appropriate for the continuous outcome, first-year GPA, while hierarchical generalized linear modeling (HGLM) is appropriate for the binary outcomes of first-year retention and six-year degree attainment. A multilevel model for a continuous outcome is represented in the equations below (Raudenbush & Bryk, 2002):

$$Y_{ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} + r_{ij} \tag{1}$$

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + u_{qj} \tag{2}$$

Equation 1 is for level 1 while equation 2 is for level 2. In the top equation Y_{ij} is the outcome for individual *i* in group *j* (for the purposes of the study, group is defined as the college or university); β_{0j} is the intercept for school *j*; $\sum_{q=1}^{Q} \beta_{qj} X_{qij}$ is the sum of the independent variables *X* and their corresponding slopes β for individual *i* in group *j*, and r_{ij} is the error term for individual *i* in group *j*. In the level 2 equation, β_{qj} represents the calculation of the intercept and slope at level 2; γ_{q0} represents the intercept and γ_{qs} represents the slope accounting for the set of level 2 predictors W_{sj}; $\sum_{s=1}^{S_q} \gamma_{qs} W_{sj}$ represents the sum of the independent variables *W* and their corresponding slopes γ, and u_{qj} represents the error term at the school level. There is an assumption that the errors are normally distributed with a mean of 0 and a variance of $\sigma^2_{u.}$ (Raudenbush & Bryk, 2002).

A multilevel model for a binary outcome is represented in the equations below (Raudenbush & Bryk, 2002). Because a linear structural model cannot be applied to a binary outcome, a link function is needed. For the purposes of this study, a logit link was used.

$$\eta_{ij} = \log(\frac{\varphi_{ij}}{1 - \varphi_{ij}}) \tag{3}$$

Here, η_{ij} is the log odds of success, or more specifically, the log odds of first-year retention and six-year graduation. The log odd of success transforms the linear structural model.

$$\eta_{ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} \tag{4}$$

And the final HGLM level one model can be written as:

$$\varphi_{ij} = \frac{1}{1 + \exp\{-\eta_{ij}\}} \tag{5}$$

The level two models then follow the same equation presented for the HLM above.

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + u_{qj} \tag{6}$$

The first step in conducting a multilevel analysis is to determine if a multilevel analysis actually is needed. In a two-level model, the intraclass correlation coefficient (ICC) is an indicator of the amount of variance in the level 1 outcome that is accounted for by the level 2 units (Raudenbush & Bryk, 2002). If the ICC for the multilevel models is not significantly different from zero, the level 2 units, in this case the institutions, do not explain much variation observed in student-athlete's first-year GPA, first-year retention or eventual graduation. The software, HLM 7.03, was used to run an intercept-only model, which allowed for determining whether multilevel modeling was needed for this study.

The preliminary exploration showed that there was a sufficient proportion of the individual-level variance that can be explained by the student-athletes' institutions needed for multilevel analysis (see Table 8). Institution accounts for 5.6% of the variance in first-year GPA, 5.7% of the variance in first-year retention, and 6.3% of the variance in six-year degree attainment.

Table 8. Results from Intercept-only Multilevel Models

OUTCOME	VARIANCE COMPONENT	SE	ICC
1st Year GPA	0.02260	0.00235	5.6%
Retained to 2nd year *	0.20060	0.02480	5.7%
Graduate *	0.22160	0.02431	6.3%
NOTE: Retained to 2nd year and Graduate were run via the Bernoulli method. The ICC was calculated using the simulation method (Merlo et al., 2006)			

The multilevel analyses included a series of increasingly complex models to evaluate how each grouping of independent variables contribute to the prediction of the outcomes. The first models that were run were the null models. As previously mentioned, the calculated ICC provides evidence that there is a cluster effect and that multilevel modeling is appropriate for these analyses. Because the assumptions and processes are slightly different for HLM and HGLM models, the discussion of RQ1 that follows addresses them separately.

Research Question 1

The first research question is: Are U.S. census block group data significantly related to student-athlete first-year college cumulative grade point average, first-year retention and eventual six-year degree attainment after controlling for student-athlete demographics and pre-college academic characteristics and college-level institutional characteristics?

First-year GPA/HLM model. Multilevel modeling assumes normality and linearity, which can be assessed by examining the distribution of the residuals. The intent of this study was to compare increasingly complex models. Per the recommendation of Hox (2010), the assumptions were assessed initially in the null model and again with the final model. Following the review of the null models and the initial assumptions, multilevel modeling using full maximum likelihood estimation was done.

The model-building process began with the development of a level 1 fixed effects model. Student-athlete demographics, pre-college academic characteristics, and the neighborhood covariates were added to a fixed effects model in three phases so that the improvement to the model could be assessed with each additional block of variables. The overall improvement to the models was assessed using the likelihood ratio test. The continuous pre-college academic and neighborhood characteristics were grand mean centered. The significance of the individual parameters was assessed and reported, and the improvement to the explained variance at level 1 also was evaluated and reported.

The development of the level 1 model continued with a random coefficient model that allowed the slopes of each of the covariates to vary across institution. Hox (2010) recommends forcing the slopes to vary one-by-one, assessing the significance of each, and then including each of the significant random slopes simultaneously in a model. That is what was done for this present study. The significance of each of the random slopes again was assessed and those that were not significant were then fixed in a subsequent model. The significance of the fixed effect also was evaluated and those that were not significant and did not vary across institution were removed from the model. This process continued iteratively until each of the level 1 parameters had a significant fixed effect and/or a significant random effect.

The relationship between the institutional characteristics at level 2 and the student-athlete outcomes was evaluated by adding a cross-level interaction of each of the level 2 covariates to each level 1 explanatory variable included in the final random coefficient model. This was done separately for each level 1 covariate. Those level 2 characteristics found to have a significant relationship with a level 1 covariate then were all included simultaneously in a model. Iteratively, level 2 characteristics were removed until all included were significant. Using a chi-square test of the deviances, the improvement between the random slope and intercept model and the model with addition of the level 2 characteristics was assessed.

Throughout, all model fit statistics were reported. To determine the relative importance of the neighborhood characteristic in predicting first-year GPA, the final model was run again using standardized coefficients.

Six-year degree attainment/HGLM models. This multilevel model-building process was then repeated for six-year degree attainment, using HGLM and Laplace estimation. Because Laplace estimation can sometimes overestimate the standard errors, the standard errors from several exploratory models using restricted maximum likelihood were compared against the standard errors attained using Laplace estimation. It was found that the Laplace estimation did not drastically overinflate the standard errors with Laplace estimation standard errors no more than a few hundredths of a point higher. In addition to the significance of the individual parameters, odds ratios also were reported.

The above analyses allowed for an examination of whether U.S. census data has a significant relationship with student-level first-year GPA and six-year degree attainment after controlling for student-level demographics, pre-college academic characteristics, and college-level institutional characteristics. The model building process also enabled an assessment of the added value of neighborhood effects to the level 1 models after accounting for the student-athlete's demographics and incoming academic characteristics.

Research Question 2

The second research question focuses on whether the U.S. census data contributes to prediction models differently for student-athletes who participate in academically at-risk sports compared to their counterparts in sports not deemed academically at-risk. To evaluate this question, the final models from RQ1 were run with the addition of an interaction term between student-athlete high-profile sport status, which is the equivalent of an academically at-risk sport for the purposes of this study, and the significant neighborhood characteristic. The significance of the fixed effect, its random slope, and any cross-level interactions with the level 2 covariates were evaluated and reported as evidence of any predictive bias in the neighborhood characteristic variables when predicting outcomes for the two separate groups. Also, a purely exploratory model was run that included an interaction term between high-profile sport and each of the neighborhood characteristics, regardless of their significance in the final RQ1 model. Again, the fixed effects, random slopes and cross-level interactions were assessed and reported.

Research Question 3

The final research question asks if the U.S. census data contributes differently to the prediction models of minority student-athletes compared to white student-athletes. To evaluate this question, the final models from RQ1 were run with the addition of an interaction term between race and the significant neighborhood characteristic. The significance of the fixed effect, its random slope, and any cross-level interactions with the level 2 covariates were evaluated and reported as evidence of any predictive bias in the neighborhood characteristic variables when predicting outcomes for the two separate groups. Also, a purely exploratory model was run that included an interaction term between student-athlete race and each of the neighborhood characteristics, regardless of their significance in the final RQ1 model. Again, the fixed effects, random slopes and cross-level interactions were assessed and reported.

CHAPTER FOUR

IN WHICH THE RESULTS ARE PRESENTED

There were three primary stages to the analysis. The first was to evaluate the covariates to ensure that multicollinearity would not impact the results and to ensure that the variables intended to be used in the models were not unduly skewed. If either was an issue, data reduction and data transformations could be employed prior to analyses. Next, the relationships between the outcomes and student-level covariates were assessed to make certain that there was a relationship present before including them in the modeling. Following, multilevel modeling was used to assess the relationship between the covariates and the outcomes of interest to determine if, after controlling for individual and institutional characteristics, the characteristics of the neighborhood had a relationship with the academic outcomes. This chapter provides the results of those analyses.

Covariate Assessment and Individual Inferential Analyses

The very first step in the data analysis was to recode a few of the variables to simplify interpretation and make it a little more practical. At level one, best test was divided by 10 to reflect the SAT scoring system, and median income was recoded to the thousands so that a one unit increase in the coefficient would represent an increase of $1,000 in the neighborhood median income. At level 2, enrollment, out-of-state cost, out-of-state GIA, and total athletics expenses also were recoded to the thousands. For both the covariate assessment and the analyses examining the relationships between the individual covariates and the outcomes, IBM SPSS Statistics 24 was used.

Covariate Assessment

The first step in the analysis was to determine if a variable reduction was needed to help mitigate the risk of multicollinearity in the data. There was particular concern that the neighborhood characteristics would be so highly correlated that their individual coefficients would be less reliable. The bivariate correlations for the neighborhood characteristics ranged from a nonsignificant $r = .01$ ($p = .17$) between the proportion of females who work full-time and the median income in the neighborhood to a significant correlation $r = .95$ ($p < .01$) between the proportion of females with a college degree or greater and the proportion of males with a college degree or greater. Table 9 provides the correlation coefficients across the neighborhood covariates. Based on the high and significant correlations among the four education variables, males with less than an 8^{th} grade education, males with a bachelor or advanced degree, females with less than an 8^{th} grade education, and females with a bachelor or advanced degree, these were reduced to a single factor. Principal components analysis using a direct oblimin rotation reduced these four measures of education to one component score, named Education Attainment Factor, that ranged from -4.73 to 2.85. The correlation of this factor with the other neighborhood characteristics can be seen in the far right column of Table 9. The correlations between the factor and the two variables capturing the proportion of males and females with less than an 8^{th} grade education are strong and negative, $r = -.912$ ($p < .01$) and $r = -.899$ ($p < .01$) respectively. The correlations between the factor and the two variables capturing the proportion of males and females with a bachelor or advanced degree are strong and positive, $r = .929$ ($p < .01$) and $r =$

.921 (p < .01) respectively. A greater factor score indicates greater academic attainment for males and females.

Table 9. Pearson Correlation Coefficients among the Neighborhood Covariates

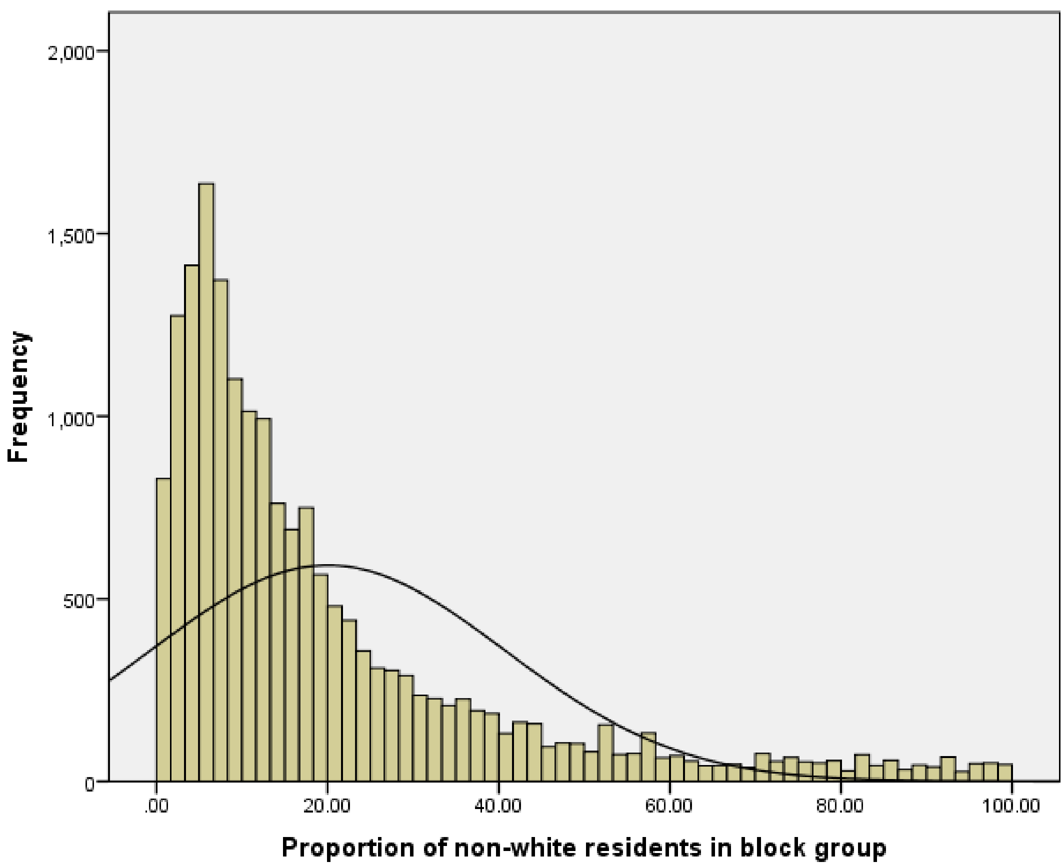
	Male Edu less than 8th grade	Male Bach or adv degree	Female Edu less than 8th grade	Female Bachor adv degree	Male Work FT	Female Work FT	Couple as head of house	Median Income	% non-white	Median years res.	Edu. Attain. Factor
Male Edu less than 8th grade	1.0	-.746*	.884*	-.713**	-.480*	-.106*	-.607*	-.642*	.470*	.099*	-.912*
Male Bachelor or advance degree		1.0	-.705*	.945*	.371*	-.028*	.503*	.769*	-.288*	-.143*	.929*
Female Edu less than 8th grade			1.0	-.708*	-.484*	-.157*	-.618*	-.644*	.480*	.080*	-.899*
Female Bachelor or advance degree				1.0	.373*	.051*	.458*	.740*	-.243*	-.119*	.921*
Male Work FT					1.0	.493*	.463*	.510*	-.304*	-.275*	.466*
Female: Work FT						1.0	-.100*	-.007	.189*	-.347*	.077*
Couple as head of house							1.0	.606*	-.753*	.067*	.596*
Median Income								1.0	-.307*	.051*	.764*
% non-white									1.0	-.118*	-.403*
Median years res.										1.0	-.121*
Edu Attain. Factor											1.0

Note: * indicates statistically significant correlation (two-tailed), p <.01

Among the institutional characteristics, the correlation between out-of-state cost and out-of-state grant in aid was r = .884 (p < .01). To correct for this, out-of-state cost was dropped from the models. Grant-in-aid can serve as both a measure of the financial aid available to the students as well as a proxy for total cost.

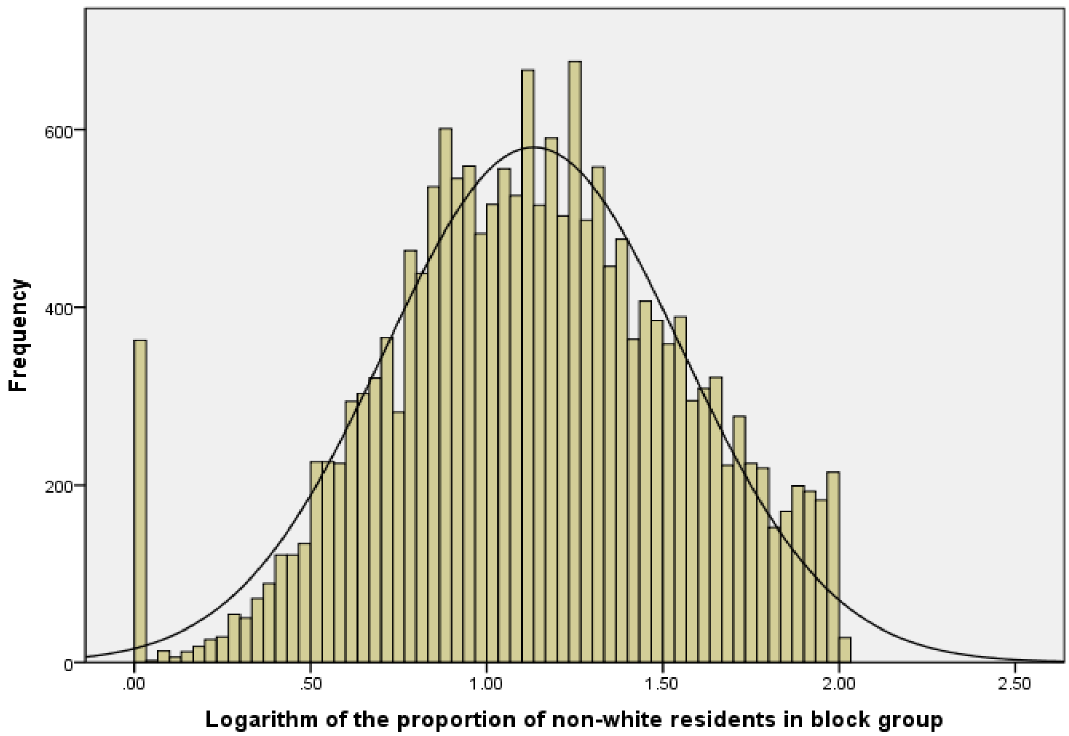
The normality of the data was assessed through histograms. Because of the large sample size used in this study, many statistical tests of normality including measures of shape (skewness and kurtosis) and a measure of the normality of the distribution (Kolmogorov-Smirnov test) are often found to be significant. The concern, however, is that this departure from normality may be so small that it would not result in biased findings using untransformed variables (Field, 2009). The histograms showed one concerning variable was skewed with a positive heavy tail: the proportion of nonwhites in the neighborhood (see Figure 1).

Figure 1. Distribution of the Proportion of Nonwhites by Census Block Group



A log transformation was used on the variable. Because some neighborhoods have a value of 0% for the proportion of non-white residents, one was added to the raw proportion before the logarithm was taken. Figure 2 shows the plot of the transformed variable.

Figure 2. Distribution of the Logarithm of the Proportion of Nonwhites by Census Block Group



Assessment of Relationships between Individual Covariates and Outcomes

Analyses, including independent sample t-tests, Pearson bivariate correlations, and chi-square analyses ensured that the intended independent variables were related to the outcomes of interest before proceeding with the multilevel modeling. Because of the large sample size, the more conservative type I error rate of .01 was used to determine statistical significance and to help mitigate the risk of making a type I error. The results of these analyses, including effect sizes, are reported by the three outcomes of interest in the study.

First-year GPA. First-year GPA was measured on a continuous scale with a sample range of 0.0 – 4.0 and a sample mean of 2.89. An independent sample t-test was conducted to compare first-year GPA by gender, race, and high-profile sport status. There were significant findings as a result of each of the t-tests. On average, female student-athletes (M = 3.07, SD = .60) reported significantly higher first-year GPAs than did males (M = 2.71, SD = .63); $t(18406) = 40.01$, $p < .01$, $d = .589$. White student-athletes (M = 3.02, SD = .60) also, on average, reported significantly higher first-year GPAs than did non-white student-athletes (M = 2.61, SD = .62); t (18415) = 43.20, $p < .01$, $d = .678$. Finally, student-athletes who participate in a sport other than men's or women's basketball, baseball or football (M = 3.01, SD = .61) reported higher first-year GPAs than did high-profile student-athletes (M = 2.67, SD = .61); $t(18415) = 36.03$, $p < .01$, $d = .557$. The effect sizes were $d = .589$, .678, and .557 respectively.

Pearson bivariate correlation was used to examine the relationship between first-year GPA and the incoming academic characteristics. The student-athletes' HSCGPA and best test have a stronger relationship with first-year GPA than does the total number of core units taken in high school; although, all are significant. The correlations between the covariates and first-year GPA are as follows: HSCPA $r = .567$ ($p < .01$); total core units $r = -.184$ ($p < .01$), and best test $r = .415$ ($p < .01$).

To assess the relationship between the neighborhood characteristics and first-year GPA, Pearson bivariate correlations again were used. Table 10 provides the results. All but one neighborhood variable, median years of residency, was significantly correlated with first-year GPA. The remaining correlations were significant but rather small ranging from a .05 to a .21.

Table 10. Pearson Correlation Coefficients: First-year GPA and Neighborhood Covariates

	Education Attainment Factor	Male Work FT	Female: Work FT	Couple as head of household	Median Income	% non-white	Median years residency
1st year GPA	.178*	.109*	-.051*	.213*	.128*	-.213*	.019
Note: * indicates statistically significant correlation, p <.01							

First-year retention. Pearson's chi-square tests were performed to evaluate the relationship between the student-athlete's demographic characteristics and first-year retention. There was a significant relationship between gender and first-year retention (χ^2 (1) = 17.64, p < .01, ϕ = -.03). Males were slightly more likely to drop out between their freshman and sophomore years than were females. The findings between race (χ^2 (1) = 2.67, p = .10) and high-profile sport status (χ^2 (1) = .14, p = .71) and first-year retention were nonsignificant.

An independent sample t-test was conducted to examine the relationship between the incoming academic characteristics, HSCGPA, total core units, and best test, and the outcome, first-year retention. Each of the tests had significant results by statistical standards. Students who were retained to their 2nd year had a higher HSCGPAs (M = 3.32, SD = .55) when compared with students-athletes who were not retained (M = 3.20, SD = .54); t(18415) = -11.69, p < .01, d = .221. Student-athletes who were retained also had, on average, greater standardized test scores using an SAT scale (M = 1066, SD = 165) than did non-retained student-athletes (M = 1048, SD = 158); t(5256) = -6.07, p < .01, d = .112. Finally, retained student-athletes (M = 17.6, SD = 2.2) had slightly fewer core units, on average, than did their non-retained peers (M = 17.8, SD = 2.1);

t(5276) = 4.30, *p* < .01, *d* = .079. Each of these had a small effect size: *d* = .221, .112, and .079 respectively.

To assess the relationship between the neighborhood characteristics and first-year retention, independent sample t-tests again were used. Table 11 provides the results. Just one neighborhood variable, median years of occupancy, was significant, but with a very small effect size.

Table 11. Independent Sample T-Test Results: First-year Retention and Neighborhood Covariates

	Mean (SD) for Retained Student-Athletes	Mean (SD) for Non-Retained Student-Athletes	t (df)	*p*	*d*
Education Attainment Factor	.01 (1.0)	-.02 (1.0)	-1.573 (18415)	.12	.030
Male Work FT	69.68 (8.8)	69.50 (9.1)	-.985 (18415)	.33	.019
Female Work FT	46.26 (8.5)	46.18 (8.4)	-.537 (18415)	.59	.010
Couple as Head of Household	80.84 (10.7)	80.65 (11.1)	-.913 (18415)	.36	.017
Median Income	57.91 (23.04)	57.26 (23.29)	-1.472 (18415)	.14	.028
Log Proportion Non-white	1.13 (.42)	1.14 (.42)	.417 (18415)	.68	.008
Years Resident	6.23 (2.92)	6.07 (2.98)	-2.846 (18415)	<.01	.054

Six-year degree attainment. Pearson's chi-square tests were performed to evaluate the relationship between the student-athlete's demographic characteristics and six-year degree attainment. There was a significant relationship between gender and graduation (χ^2 (1) = 214.05, *p* < .01, ϕ = .11). Females were more likely to have earned their baccalaureate degree within six years of enrollment than were males. Although with a very small effect size, white student-athletes were significantly more likely to have graduated within six years when compared with non-white student-athletes (χ^2 (1) = 35.53, $p < .01$, $\phi = .04$). Finally, again with a small effect size, student-athletes in a non-high-profile sport were more likely to have earned their degree within six years (χ^2 (1) = 30.67, $p < .01$, $\phi = .04$).

An independent sample t-test was conducted to examine the relationship between the incoming academic characteristics, HSCGPA, total core units, and best test, and six-year graduation. Each of the tests had significant results by statistical standards. Students who graduated within six years had a higher HSCGPAs (M = 3.38 SD = .54) when compared with students-athletes who did not graduate (M = 3.19, SD = .55); $t(18415) = -23.96$, $p < .01$, $d = .356$. Student-athletes who graduated also had, on average, greater standardized test scores using an SAT scale (M = 1076, SD = 164) than did those who did not graduate (M = 1045, SD = 161); $t(18415) = -12.88$, $p < .01$, $d = .191$. Finally, graduated student-athletes (M = 17.53, SD = 2.2) had slightly fewer core units, on average, than did their peers who did not graduate (M = 17.78, SD = 2.1); $t(17674) = 7.92$, $p < .01$, $d = .117$.

To assess the relationship between the neighborhood characteristics and six-year graduation, independent sample t-tests again were used. Table 12 provides the results. Five of the seven variables, the education attainment factor, proportion of males who work full-time, proportion of families with couple as head of household, median income, and the logarithm of the proportion of residents who are non-white, were significantly related to baccalaureate attainment within 6 years.

Table 12. Independent Sample T-Test Results: Six-year Graduation and Neighborhood Covariates

	Mean (SD) for Student-Athletes with Degree	Mean (SD) for Student-Athletes without Degree	t (df)	*p*	*d*
Education Attainment Factor	.05 (.99)	-.06 (1.0)	-7.57 (18415)	<.01	.113
Male Work FT	69.9 (8.6)	69.3 (9.1)	-4.60 (16725)	<.01	.069
Female Work FT	46.2 (8.4)	46.3 (8.6)	.817 (18415)	.41	.012
Couple as Head of Household	81.2 (10.5)	80.3 (11.2)	-5.81 (16573)	<.01	.087
Median Income	58.85 (23.26)	56.40 (22.78)	-7.15 (18415)	<.01	.106
Log Proportion Non-white	1.12 (.42)	1.15 (.43)	5.19 (17011)	<.01	.077
Years Resident	6.24 (2.96)	6.14 (2.98)	-2.21 (18415)	.03	.034

Summary of Covariate Assessment and Individual Inferential Analyses

Several actions were taken based on the assessment of the covariates and the individual inferential analyses. Based on the assessment of the individual covariates, total out-of-state cost was dropped from further analyses, and total out-of-state grant-in-aid served as a measure of the financial aid available to students as well as a proxy for total cost. The logarithm of the proportion of non-whites in the neighborhood was taken to help account for the positive skew of that variable. Also related to the neighborhood characteristics, educational attainment was transformed into a factor score that comprises the male and female proportions of attaining less than an 8th grade education and attaining a college four-year degree or greater.

Findings from the individual inferential analyses resulted in two major changes to the anticipated models. The first change was related to the neighborhood covariates. The median number of years residents live in the neighborhood was not related to either first-year GPA nor six-year degree attainment. It was statistically related to first-year retention; however, the effect size was .05. For these reasons, median tenure in the neighborhood was removed from further analyses. The second, and greatest change, was the assessment of the research questions related to first-year retention. Several covariates were unrelated to the outcome, including race, high-profile sport status, and all but one of the neighborhood characteristics. While median tenure in the neighborhood did have a statistically significant p-value, the effect size was .05. For these reasons and because the relationship between the outcome and the neighborhood characteristics was the primary interest for the current study, the analysis of RQs 1 through 3 regarding first-year retention stopped with these analyses. Chapter Five contains more discussion regarding these findings.

Multi-Level Modeling Results

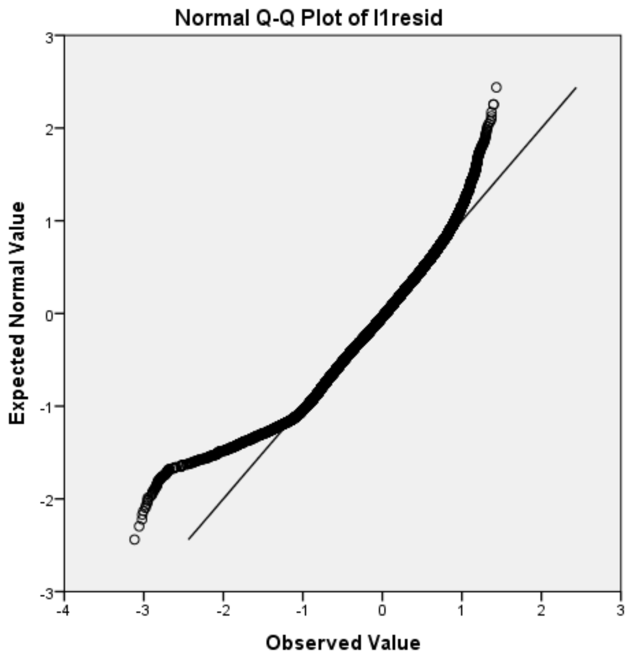
To answer the research questions put forth in this study, multilevel modeling was needed to account for the nesting of student-athletes within institutions. Hierarchical linear modeling was used to analyze the outcome, first-year GPA, while HGLM was used to analyze the relationship with the dichotomous outcome, six-year degree attainment. As was noted previously, the analysis of the research questions related to the first-year retention were confined to the preliminary analyses discussed above. The results that follow are addressed by the two remaining outcomes, first-year GPA and six-year degree attainment, and the three research questions.

Outcome: First-Year GPA

Research question 1. The first research question is: Are U.S. census block group data significantly related to student-athlete first-year college cumulative grade point average, first-year retention and eventual six-year degree attainment after controlling for student-athlete demographics and pre-college academic characteristics and college-level institutional characteristics?

Null model and hypothesis testing. The first step in conducting the multilevel linear analysis was to assess the amount of variation in student-athletes' individual first-year GPAs that can be attributed to between-group variation, and therefore, attributed to the institutions. An ICC was calculated based on the results of Model 1, an intercept-only or null model. The findings suggest that the college or university the student-athlete attended accounted for roughly 5.6% of the variation in first-year GPA. Results from Model 1 and subsequent models related to RQ1 and first-year GPA are displayed in Tables 13 and 14. Also important at this stage was an assessment of the residuals to ensure normality. Figure 3 shows a Q-Q plot of the residuals from Model 1. While there are slight tails at either end, the plot indicates that the residuals are close to normal.

Figure 3. Distribution of Residuals from Model 1with First-year GPA as Outcome.



Development of the level 1 fixed effects model. The development of the model continued with the introduction of uncentered student-athlete demographics, grand-mean centered incoming academic characteristics and grand-mean centered neighborhood characteristics in three phases. The covariates were examined to determine if any had a fixed effect on first-year GPA. The models were run using full maximum likelihood estimation, which produces deviance figures that can be compared across models to assess improvement to overall model fit. The addition of each group of variables resulted in a significant improvement to the overall model based on the likelihood ratio test and an increase in the explained variance at level 1. The addition of the pre-college academic characteristics resulted in the greatest improvement in explained variance at level 1 with an increase from Model 2 to Model 3 of 28.7%. The addition of the neighborhood characteristics improved the model slightly, χ^2 (6) = 120.82, $p < .01$, but the percent of explained variance between Model 3 and Model 4 increased by less than 1%.

It should be noted that the addition of the high school academic characteristics resulted in a substantial reduction in the variance explained at level 2. It was a concern that multicollinearity may be causing this reduction in the explained variance. To test this, several optional models were run. These included the student-athlete demographics and just HSCGPA, student-athlete demographics and best test and core units, student-athlete demographics and just best test, and gender, race, and the three high school academic characteristics. The three former models did result in some improvement to that reduction in explained variance. The additional explained variance for each was: -31.7%, -40.9%, and -44.2% respectively. The latter resulted in a similar reduction in explained variance at -65.4%. A reduction in the level 2 explained variance after the addition of level 1 covariates is possible when there are group-level differences in mean of the added level 1 covariates that are related to the outcome (Steele, 2008).

In the final fixed effects model, Model 4, the mean first-year GPA is 3.00. Practically, this means that for female, white, student-athletes, who have average incoming high school academic characteristics, and who live in an average neighborhood on the measures included in the model, their predicted first-year GPA would be a 3.00. The high-profile status of the student-athlete's sport was not significant, $\gamma_{30} = .008$, $p = .454$. The other demographic variables, male and non-white, both had a significant and negative relationship with first-year GPA. The student-athletes standardized test and their HSCGPA both positively aided in the prediction of first-year GPA, while total units had a negative relationship with first-year GPA. Finally, two of the neighborhood covariates were significant in the fixed model: the neighborhood education factor positively contributed and the proportion of females over the age of 16 who work full-time negatively contributed.

Development of the level 1 random coefficients and intercepts model. The next step in developing the level 1 model was to run a random coefficients and intercept model. Initially, a random error term was added to all covariates from Model 4. This model, however, would not converge. Ultimately, the slopes from each of the variables from Model 4 were set to vary one-by-one in a series of models labeled Models 5a through 5l. Results from these models indicate that the following two characteristics did not vary across the institutions: the percent of males over the age of 16 who work full-time u_{8j} = .000 χ^2 (326) = 375.72, p = .03 and the percent of females over the age of 16 who work full time u_{9j} = .000 χ^2 (326) = 320.69, p > .05. The remaining characteristics, including the student-athlete demographics, incoming academic characteristics, and neighborhood characteristics, had significant error terms when they were individually allowed to vary in a model.

Model 5m then included random slopes for all variables that had a significant error term in Models 5a – 5l (see Appendix A). Significant random slopes were maintained by the following covariates: student-athlete race (non-white) u_{2j} = .008 χ^2 (321) = 418.33, p < .01, high-profile sport status u_{3j} = .010 χ^2 (321) = 414.36, p < .01, and HSCGPA u_{4j} = .015 χ^2 (321) = 480.63, p < .01. Model 5n then allowed these to vary while fixing the remaining slopes. The variances of all remained significant in this model. Model 5o dropped from the model the following variables with a nonsignificant fixed effect and a nonsignificant random slope: proportion of males over the age of 16 who work full-time, the proportion with a couple head of household, the logarithm of the proportion of the neighborhood that is non-white, and the median income. Race, high-profile sport status, and HSCPA continued to vary. In Model 5o, the percent of females over the age of 16 who work full-time became nonsignificant γ_{80} = -0.000, p = .02.

The final random slopes and intercept model, Model 5p, dropped the proportion of full-time female workers. Before the addition of the level 2 covariates, the interpretation of this model is: The average first-year GPA of female, white student-athletes with average incoming high school characteristics is 3.01. Student-athletes who are male and/or non-white are predicted to have lower first-year GPAs, while HSCGPA, best standardized test and the education of the neighborhood positively contributed to a student-athlete's first-year GPA. Total core units was negatively related to first-year GPA. The reason for this will be discussed in greater detail in Chapter 5.

The random error terms of student-athlete race, high-profile sport status, and HSCGPA were significant meaning that the relationship between these variables and first-year GPA differs across schools. It should be noted that the variance of high-profile sport status was significant in spite of a nonsignificant fixed effect. The addition of these random error terms resulted in an improved model based on the likelihood ratio test, (χ^2 (19) = 231.89, p < .01).

Table 13. Level 1 Multilevel Null, Random Intercept and Random Intercept & Coefficients Models: First-year GPA as Outcome

	Model 1 Null Model	Model 2 Adding Fixed SA Demogr.	Model 3 Adding Fixed HS Academic	Model 4 Adding Fixed Neighborhood	Model 5p Random Intercept& Coefficients
FIXED EFFECTS					
γ_{00} Intercept (1st year GPA)	2.89 (.010)	3.17 (.010)	3.01 (.011)	3.00 (.011)	3.01 (.011)
γ_{10} Male		-0.265* (.011)	-0.143* (.009)	-0.139* (.009)	-0.139* (.009)
γ_{20} SA Non-white		-0.326* (.013)	-0.097* (.010)	-0.072* (.011)	-0.085* (.010)
γ_{30} High-profile sport		-0.103* (.013)	0.001 (.011)	0.008 (.009)	0.006 (.010)

γ_{40} **HSCGPA**			0.514* (.011)	0.524* (.011)	0.536* (.011)
γ_{50} **Core units**			-0.012* (.002)	-0.013* (.002)	-0.012* (.002)
γ_{60} **Best test (10s)**			0.007* (.000)	0.007* (.000)	0.007* (.000)
γ_{70} **Neighborhood Edu**				0.038* (.006)	0.042* (.004)
γ_{80} **% Male work FT**				0.002 (.063)	
γ_{90} **% Female work FT**				-0.002* (.057)	
γ_{100} **% couple head of household**				-0.000 (.066)	
γ_{110} **Log neighborhood non-white**				-0.020 (.013)	
γ_{120} **Median Income**				-0.000 (.000)	
RANDOM EFFECTS					
$\sigma^2_{r:}$ Within-school	0.3797	0.3238	0.2308	0.2291	0.2207
σ^2_{u0} Intercept	0.0226*	0.0159*	0.0263*	0.0274*	0.0279*
σ^2_{u2} SA Non-white					0.006*
σ^2_{u3} SA High-profile sport					0.008*
σ^2_{u4} HSCGPA					0.018*
σ^2_{u02} Intercept-SA Non-white cov (SE)					-0.005 (.002)
σ^2_{u02} Intercept-High-profile cov (SE)					-0.001 (.002)
σ^2_{u04} Intercept-HSCGPA cov (SE)					0.004 (.002)
GOODNESS OF FIT					
Deviance	34896.89	31912.85	25894.50	25773.68	25541.79
# of parameters	3	6	9	15	19
Chi-square results		2984.04*	6018.35*	120.82*	231.89*
Added Explained variance LV1^		14.7%	28.7%	0.7%	3.7%

Added Explained variance LV2^		29.6%	-65.4%	-4.2%	-1.8%
NOTE: * indicates statistically significant finding, p<.01; ^ indicates compared to immediately previous model; **Bold** indicates a grand-mean centered variable; SE in parentheses					

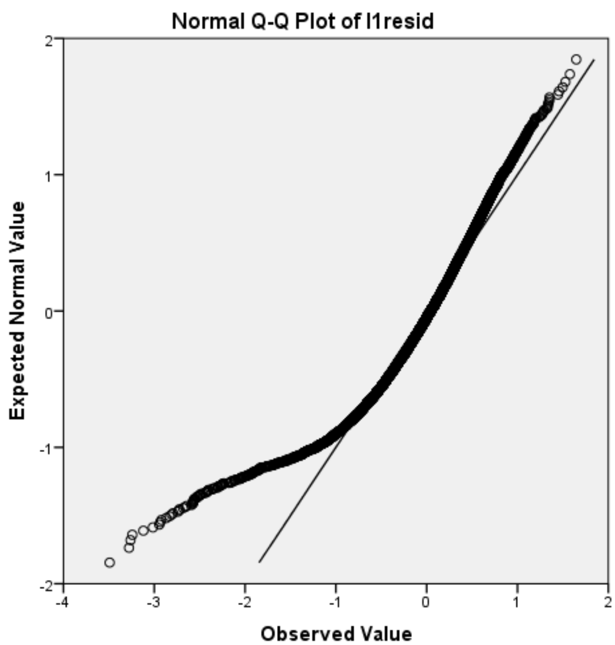
Development of the level 2 model. Although it is more common to inspect the significance of the level 2 variables on covariates with a random slope, exploratory models of level 2 variables were run for the intercept and each of the level 1 covariates, using Model 5p as the basis. Model 6a added level 2 variables to the intercept and found that the intercept was dependent upon MSI, $\gamma_{01} = -0.117$, $p < .01$ and student body graduation rate, $\gamma_{04} = -0.008$, $p < .01$. This process was then carried out for the covariates in Model 5p. High-profile sport ($\gamma_{34} = -0.002$, $p < .01$), HSCGPA ($\gamma_{44} = -0.005$, $p < .01$), and total units ($\gamma_{54} = 0.0004$, $p < .01$) had a significant cross-level interaction with student body graduation rate. The relationship between the other level 1 variables and first-year GPA were not dependent upon any of the level 2 variables. After these cross-level interactions were combined and added simultaneously to a model, Model 6i, the relationships between high-profile status and student body graduation rate ($\gamma_{31} = -.0003$, $p = .590$), and total units and student body graduation rate ($\gamma_{51} = .0002$, $p = .012$) were no longer significant. In Model 6j, the cross-level interaction between MSI and the intercept became nonsignificant, $\gamma_{01} = -0.106$, $p = .011$. Table 14 provides the results from the final HLM of first-year GPA, Model 6k.

Table 14. Multilevel Linear Model Final Results: First-year GPA as Outcome

	Model 6k – final level 2 model	Model 7 – standardized model
FIXED EFFECTS		
γ_{00} Intercept (1st year GPA)	3.02 (.009)	0.048 (.012)
γ_{01} **Student body grad rate**	-0.006* (.0006)	-0.175* (.013)
γ_{10} Male	-0.137* (.009)	-0.108* (.007)
γ_{20} SA Non-white	-0.086*(.010)	-0.063* (.007)
γ_{30} High-profile sport	0.006 (.011)	0.005 (.008)
γ_{40} **HSCGPA**	0.543* (.010)	0.472* (.009)
γ_{41} **Student body grad rate**	-0.005* (.001)	-0.078* (.009)
γ_{50} **Core units**	-0.012* (.002)	-0.041* (.006)
γ_{60} **Best test (10s)**	0.007* (.0003)	0.186* (.009)
γ_{70} **Neighborhood Edu**	0.045* (.004)	0.071* (.007)
RANDOM EFFECTS		
$\sigma^2_{r:}$ Within-school	0.2208	0.5492
σ^2_{u0} Intercept	0.0155*	0.0374*
σ^2_{u2} SA Non-white	0.0070*	0.0038*
σ^2_{u3} SA High-profile sport	0.0088*	0.0050*
σ^2_{u4} HSCGPA	0.0105*	0.0079*
σ^2_{u02} Intercept-SA Non-white cov (SE)	-0.002 (.002)	0.001 (.002)
σ^2_{u03} Intercept-SA High-profile cov (SE)	-0.002 (.002)	0.002 (.002)
σ^2_{u04} Intercept-HSCGPA cov (SE)	-0.005 (.001)	-0.008 (.001)
GOODNESS OF FIT		
Deviance	25318.07	42099.46
# of parameters	21	21
Chi-square results	223.72*	--
Added Explained variance LV1^	0.0%	--
Added Explained variance LV2^	44.4%	--
NOTE: * indicates statistically significant finding, p<.01; ^ indicates compared to immediately previous model; **Bold** indicates a grand-mean centered variable; SE in parentheses		

Model 6k could be the final model used to answer RQ1. Before settling on this model, however, it is important to reassess the normality assumption. This can be done by inspecting a plot of the residuals. Figure 4 shows those results.

Figure 4. Plot of Level 1 Residuals from Model 6k.



An analysis of the empirical Bayes residuals at level 2 for the intercept and covariates with a random error term indicate that the assumption of normality is met as well (see Appendix B).

Finally, to establish the relative strength of the neighborhood education factor when compared against the demographic and incoming academic characteristics, Model 6k was repeated using standardized coefficients (see Model 7 in Table 14).

Response to research question 1 (first-year GPA). According to the data and the multilevel results, student-athlete gender, race, HSCGPA, total core units, best standardized test, and a neighborhood education factor at the block group level help to significantly predict first-year GPA. Equations 7 and 8a – 8h show the final level 1 and level 2 models respectively. Variables that are displayed in bold were grand-mean centered.

$$FIRSTGPA_{ij} = \beta_{0j} - \beta_{1j}MALE_{ij} - \beta_{2j}\,SA\ NONWHITE_{ij} + \beta_{3j}\,HIGHPROFILE_{ij} + \beta_{4j}\,\boldsymbol{HSCGPA_{ij}} - \beta_{50}\,\boldsymbol{HS\ UNITS_{ij}} + \beta_{6j}\,\boldsymbol{BEST\ TEST_{ij}} + \beta_{70}\,\boldsymbol{NEIGHBORHOOD\ EDU_{ij}} + r_{ij} \qquad (7)$$

$$\beta_{0j} = \gamma_{00} - \gamma_{01}(\boldsymbol{SB\ GRAD\ RATE_j}) + u_0 \qquad (8a)$$

$$\beta_{1j} = \gamma_{10} \qquad (8b)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \qquad (8c)$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \qquad (8d)$$

$$\beta_{4j} = \gamma_{40} - \gamma_{41}(\boldsymbol{SB\ GRAD\ RATE}) + u_{4j} \qquad (8e)$$

$$\beta_{5j} = \gamma_{50} \qquad (8f)$$

$$\beta_{6j} = \gamma_{60} \qquad (8g)$$

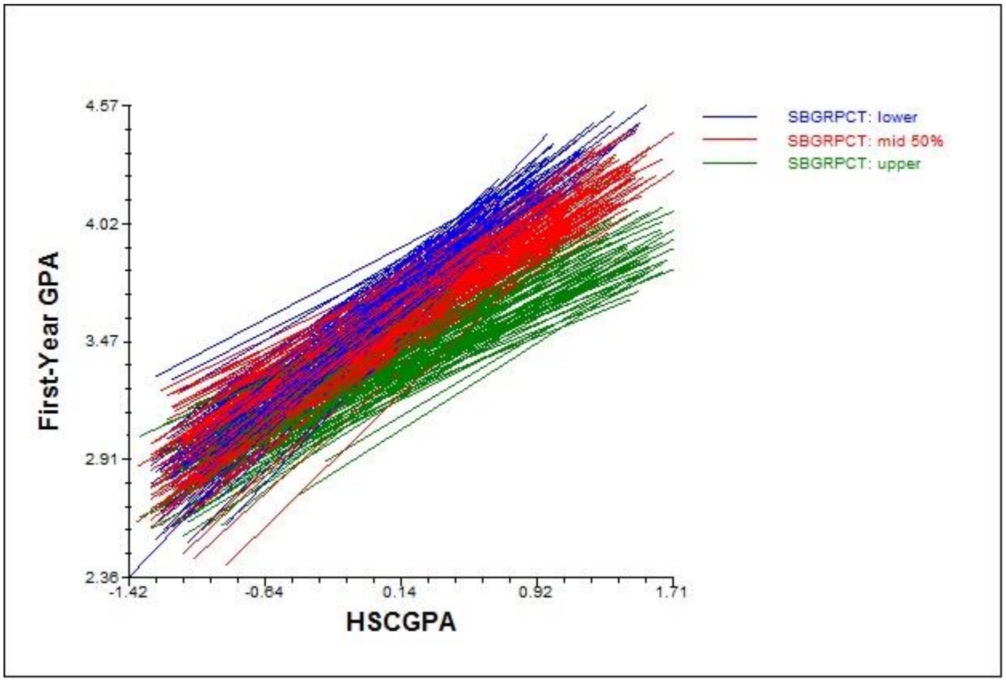
$$\beta_{7j} = \gamma_{70} \qquad (8h)$$

There was a significant variance in intercepts across institutions, var (u_{0j}) = 0.016, $\chi^2(323) = 975.11$, $p < .01$. The slopes of student-athlete race var (u_{2j}) = 0.007, $\chi^2(324) = 450.47$, $p < .01$, high-profile sport var (u_{3j}) = 0.009, $\chi^2(324) = 467.04$, $p < .01$, and HSCGPA var (u_{4j}) = 0.010, $\chi^2(324) = 536.69$, $p < .01$ also varied across institutions. The HSCGPA-intercept covariance was significant and negative suggesting that as the mean first-year GPA increased, the slope for HSCGPA flattened some.

The average first-year GPA for a female, white student-athlete with an average HSCGPA, total core units, and standardized test on the SAT scale and who lived in a census block group area with average educational attainment is 3.02. Among the demographic variables, male and non-white had a negative relationship with first-year GPA. Holding all else constant, being male is expected to result, on average, in a first-year GPA that is 0.14 lower, and holding all else constant, being a race other than white is expected to result, on average, in a first-year GPA that is 0.09 lower. All incoming high school academic characteristics significantly predicted first-year GPA. While HSCGPA and best standardized test were related positively to first-year GPA, total core units had a negative relationship with the outcome. The addition of the census block group data in Model 4 explained a little less than just 1% more of the variance in level 1. In the final model, the education level of the neighborhood was significant in predicting first-year GPA after controlling for the other variables. As the factor increased by one point, first-year GPA is expected to increase by 0.05, holding all else constant.

Model 6k shows a main effect of student body graduation rate. The negative relationship of the interaction between the intercept and student body graduation rate implies that as the graduation rate of the student body increases, the first-year GPA intercept decreases. Also, the relationship between HSCGPA and first-year GPA is at least partially dependent on the student body graduation rate. The negative coefficient of $\gamma_{41} = -0.005$ implies that as the student body graduation rate increases, the magnitude of the relationship between HSCGPA and first-year GPA lessens (see Figure 5).

Figure 5. Predicted First-year GPA from Random HSCGPA by Student Body Graduation Rate



Model 7 is a repeat of Model 6k using standardized coefficients to determine the relative strength of the fixed effects on first-year GPA. By far, HSCGPA had the strongest relationship with the outcome (γ_{40} = 0.472, $p < .01$) – nearly three times that of best standardized test (γ_{60}= 0.186, $p < .01$), which had the next strongest relationship. Following were gender (γ_{10} = -0.108, $p < .01$) and then the neighborhood educational factor (γ_{70} = 0.071, $p < .01$). Based on Model 7, the educational attainment of the neighborhood has a stronger relationship with first-year GPA than does the individual's race (γ_{30} = -0.063, $p < .01$) or the number of core courses taken in high school (γ_{40} = -0.041, $p < .01$).

Research question 2. The second research question is: Do U.S. census block group data relate to NCAA Division I student-athlete first-year college cumulative grade point average, first-year retention and eventual six-year degree attainment differently for student-athletes who participate in academically at-risk sports and their counterparts in sports not deemed academically at-risk?

Model 6k was used as the basis for the analysis of this research question. An interaction between high-profile status, which also are the sports deemed to be most academically at-risk, and the neighborhood education factor was entered into the model. Although the main effect of high-profile status is nonsignificant, the interaction term produced a statistically significant result, $\gamma_{80} = -0.030$, $p < .01$ (see Table 15). For high-profile, or academically at-risk student-athletes, the relationship between the neighborhood education factor and first-year GPA is reduced by .03 holding all else constant. Practically, this means that the relationship between first-year GPA and the level of education in the neighborhood is stronger for student-athletes who do not participate in an academically at-risk sport.

Table 15. Multilevel Linear Model with High-Profile Sport * Neighborhood Education Interaction Term: First-year GPA as Outcome

	Model 6k – final level 2 model	Model 8
FIXED EFFECTS		
γ_{00} Intercept (1st year GPA)	3.02 (.009)	3.02 (.009)
γ_{01} **Student body grad rate**	-0.006* (.0006)	-0.006* (.0004)
γ_{10} Male	-0.137* (.009)	-0.136* (.009)
γ_{20} SA Non-white	-0.086*(.010)	-0.088* (.010)
γ_{30} High-profile sport	0.006 (.011)	0.003 (.011)
γ_{40} **HSCGPA**	0.543* (.010)	0.544* (.010)
γ_{31} **Student body grad rate**	-0.005* (.001)	-0.005* (.001)

γ_{50} **Core units**	-0.012* (.002)	-0.012* (.002)
γ_{60} **Best test (10s)**	0.007* (.0003)	0.007* (.0003)
γ_{70} **Neighborhood Edu**	0.045* (.004)	0.057* (.005)
γ_{80} **High-profile sport * Neighborhood Edu**		-0.030* (.008)
RANDOM EFFECTS		
$\sigma^2_{r:}$ Within-school	0.2208	0.2206
σ^2_{u0} Intercept	0.0155*	0.0156*
σ^2_{u2} SA Non-white	0.0070*	0.0069*
σ^2_{u3} SA High-profile sport	0.0088*	0.0088*
σ^2_{u4} HSCGPA	0.0105*	0.0105*
σ^2_{u02} Intercept-SA Non-white cov (SE)	-0.002 (.002)	-0.002 (.002)
σ^2_{u03} Intercept-SA High-profile cov (SE)	-0.002 (.002)	-0.003 (.002)
σ^2_{u04} Intercept-HSCGPA cov (SE)	-0.005 (.001)	-0.005 (.002)
GOODNESS OF FIT		
Deviance	25318.07	25302.62
# of parameters	21	22
Chi-square results		15.45*
Added Explained variance LV1^		0.0%
Added Explained variance LV2^		0.0%
NOTE: * indicates statistically significant finding, p<.01; ^ indicates compared to the immediately previous model; **Bold** indicates a grand-mean centered variable; SE in parentheses		

Model 8 also was run with a random slope for the interaction term, which was nonsignificant u_{08} = .0006, $\chi^2(323) = 328.01$, p = .412. The level 2 variables were then added to the interaction covariate, and all were nonsignificant: MSI (γ_{81} = .004, p = .875), private (γ_{82} = .023, p = .275), enrollment (γ_{83} = -.001, p = .123), student body graduation (γ_{84} = -.0002, p = .589), out-of-state GIA (γ_{85} = -.002, p = .019), and total expenses (γ_{86} = .000, p = .820). Equations 9 and 10a – 10i show the final level 1 and level 2 models respectively in response to RQ2. Variables that are displayed in bold were grand-mean centered.

$$FIRSTGPA_{ij} = \beta_{0j} - \beta_{1j}MALE_{ij} - \beta_{2j}\,SA\,NONWHITE_{ij} + \beta_{3j}\,HIGHPROFILE_{ij} + \beta_{4j}\,\boldsymbol{HSCGPA_{ij}} - \beta_{50}\,\boldsymbol{HS\,UNITS_{ij}} + \beta_{6j}\,\boldsymbol{BEST\,TEST_{ij}} + \beta_{7j}\,\boldsymbol{NEIGHBORHOOD\,EDU_{ij}} - \beta_{8j}\,\boldsymbol{NEIGHBORHOOD\,EDU * HIGHPROFILE}_{ij} + r_{ij} \tag{9}$$

$$\beta_{0j} = \gamma_{00} - \gamma_{01}(\boldsymbol{SB\,GRAD\,RATE_j}) + u_{0j} \tag{10a}$$

$$\beta_{1j} = \gamma_{10} \tag{10b}$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \tag{10c}$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \tag{10d}$$

$$\beta_{4j} = \gamma_{40} - \gamma_{41}(\boldsymbol{SB\,GRAD\,RATE_j}) + u_{4j} \tag{10e}$$

$$\beta_{5j} = \gamma_{50} \tag{10f}$$

$$\beta_{6j} = \gamma_{60} \tag{10g}$$

$$\beta_{7j} = \gamma_{70} \tag{10h}$$

$$\beta_{8j} = \gamma_{80} \tag{10i}$$

For exploratory purposes, a second model was run that included an interaction term between high profile status and each of the neighborhood characteristics. Aside from the neighborhood education factor, none of the additional neighborhood variables were significant. In a second step, the slopes for each were allowed to vary, and again, none produced significant results.

Research question 3. The third and final research question is: Do U.S. census block group data relate to NCAA Division I student-athlete first-year college cumulative grade point average, first-year retention, and eventual 6-year degree attainment differently for minority student-athletes and white student-athletes?

Model 6k was used as the basis for the analysis of this research question. An interaction between student-athlete race (non-white) and the neighborhood education factor was entered into the model. Equations 11 – 12i show the model, labeled Model 9a, that was run. The interaction term was nonsignificant, γ_{80} = -0.005, p = .525. In a separate model, the slope was allowed to vary, which also was nonsignificant, u_{08} = .0018, $\chi^2(324)$ = 353.19, p = .127. Based on these results, the relationship between neighborhood education and first-year GPA does not differ by student-athlete race.

$$\begin{aligned} FIRSTGPA_{ij} = {} & \beta_{0j} - \beta_{1j} MALE_{ij} - \beta_{2j}\, SA\ NONWHITE_{ij} \\ & + \beta_{3j}\, HIGHPROFILE_{ij} + \beta_{4j}\, \boldsymbol{HSCGPA_{ij}} - \beta_{50}\, \boldsymbol{HS\ UNITS_{ij}} \\ & + \beta_{6j}\, \boldsymbol{BEST\ TEST_{ij}} + \beta_{7j}\, \boldsymbol{NEIGHBORHOOD\ EDU_{ij}} \\ & + \beta_{8j}\, \boldsymbol{NEIGHBORHOOD\ EDU * SA\ NONWHITE}_{ij} + r_{ij} \end{aligned} \tag{11}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\boldsymbol{SB\ GRAD\ RATE_j}) + u_{0j} \tag{12a}$$

$$\beta_{1j} = \gamma_{10} \tag{12b}$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \tag{12c}$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \tag{12d}$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}(\boldsymbol{SB\ GRAD\ RATE_j}) + u_{4j} \tag{12e}$$

$$\beta_{5j} = \gamma_{50} \tag{12f}$$

$$\beta_{6j} = \gamma_{60} \tag{12g}$$

$$\beta_{7j} = \gamma_{70} \tag{12h}$$

$$\beta_{8j} = \gamma_{80} \tag{12i}$$

For exploratory purposes, a second model was run that included an interaction term between student-athlete race and each of the neighborhood characteristics. None, however, were significant. In a second step, the slopes for each were allowed to vary, and again, none produced significant results.

Six-Year Degree Attainment

Six-year degree attainment was measured as a dichotomous variable – 0 for no degree and 1 for a baccalaureate degree. Because of the binary nature of the outcome, HGLM was used. The ability to compare goodness of fit across models was desired. Using Laplace estimation in HLM 7.03 provides a deviance statistic with a chi-square distribution for hypothesis testing. There is concern, however, that Laplace estimation inflates standard errors. A few trial models were run using restricted maximum likelihood prior to the model building process, and the standard errors were assessed. In all, the standard errors using Laplace were no more than a few hundredths of a point higher. The following analyses were then conducted using full maximum likelihood and Laplace estimation.

Research question 1. By way of reminder, the first research question asked whether neighborhood characteristics using U.S. census data at the block group level aid in the prediction models of six-year degree attainment after controlling for student-athlete demographics and incoming high school characteristics and college-level characteristics.

Null model and hypothesis testing. The first step in conducting the HGLM was to assess the amount of variation in student-athletes' six-year degree attainment that can be attributed to between-group variation, and therefore, attributed to the institutions. Using results from Model 10, an ICC was calculated using the simulation method (Merlo et al., 2006). The resulting ICC was 6.3%, meaning that 6.3% of the variation in six-year degree attainment is at the college-level, and HGLM is appropriate for addressing this clustering of student-athletes within institutions.

Development of the level 1 fixed effects model. The development of the model continued with the introduction of uncentered student-athlete demographics, grand-mean centered incoming academic characteristics and grand-mean centered neighborhood characteristics in three phases. The covariates were examined to determine if any had a fixed effect on six-year degree attainment. The addition of each group of variables resulted in a significant improvement to the overall model based on the likelihood ratio test. Model 13, the final fixed effects model, resulted in four significant covariates: gender (male) γ_{10} = -0.374, $p < .01$, high profile sport γ_{30} = 0.203, $p < .01$, HSCGPA γ_{40} = 0.653, $p < .01$, and total core high school units γ_{50} = -0.032, $p < .01$ (see Table 16). None of the neighborhood characteristics had a significant fixed effect on six-year degree attainment. The probability of graduation within six years for a female student-athlete who participates in a sport other than basketball and who earned an average HSCGPA and total core units is 59.3%.

Table 16. Level 1 multilevel null, random intercept and random intercept & coefficients models: Six-year Degree Attainment as Outcome

	Model 10 Null Model	Model 11 Adding Fixed SA Demogr.	Model 12 Adding Fixed HS Academic	Model 13 Adding Fixed Neighborhood	Model 14o Random Intercept & Coefficients
FIXED EFFECTS					
γ_{00} Intercept (6 yr degree)	0.237 (.032) OR: 1.268	0.479 (.036) OR: 1.614	0.343 (.037) OR: 1.409	0.337 (.038) OR:1.400	0.376 (.038) OR: 1.457
γ_{10} Male		-0.531* (.031) OR: 0.588	-0.383* (.032) OR: 0.682	-0.374* (.033) OR: 0.688	-0.400* (.042) OR: 0.671
γ_{20} SA Non-white		-0.057 (.033) OR: 0.945	0.109* (.037) OR: 1.115	0.113 (.041) OR: 1.120	

γ_{30} High-profile sport		0.133* (.036) OR: 1.143	0.196* (.038) OR: 1.217	0.203* (.039) OR: 1.225	0.243 (.032) OR: 1.275
γ_{40} **HSCGPA**			0.627* (.036) OR: 1.871	0.653* (.038) OR: 1.922	0.577 (.032) OR: 1.781
γ_{50} **Core units**			-0.031* (.008) OR: 0.969	-0.032* (.001) OR: 0.968	-0.027 (.008) OR: 0.974
γ_{60} **Best test (10s)**			-0.002 (.001) OR: 0.998	-0.004 (.001) OR: 0.996	
γ_{70} **Neighborhood Edu**				0.020 (.027) OR: 1.020	
γ_{80} **% Male work FT**				0.003 (.003) OR: 1.003	
γ_{90} **% Female work FT**				-0.002 (.002) OR: 0.998	
γ_{100} **% couple head of household**				-0.005 (.003) OR: 0.995	
γ_{110} **Log neighborhood non-white**				0.006 (.056) OR: 1.006	
γ_{120} **Median Income**				0.003 (.001) OR: 1.003	
RANDOM EFFECTS					
σ^2_{u0} Intercept	0.2216	0.2302	0.2174	0.2129	0.2485
σ^2_{u1} Male					0.1151*
σ^2_{u01} Intercept-Male cov (SE)					-.070 (.025)
GOODNESS OF FIT					
Deviance	58616.78	58368.98	57994.87	57969.82	57987.92
# of parameters	2	5	8	14	8
Chi-square results		247.80*	374.11*	25.05*	
Added Explained variance LV2^		-3.9%	5.6%	2.1%	-16.7%
NOTE: * indicates statistically significant finding, *p*<.01; ^ indicates compared to immediately previous model; **Bold** indicates a grand-mean centered variable; SE in parentheses					

Development of the random coefficients and intercepts model. The next step in developing the level 1 model was to run a random coefficients and intercept model. Initially, a random error term was added to all covariates from Model 13. This model, however, would not converge. Ultimately, the slopes from each of the variables from Model 13 were set to vary one-by-one in a series of models labeled Models 14a through 14l. Using the value of the random slope coefficient and standard error estimates to evaluate the significance of the random slopes, results from these models indicate that the following four characteristics do vary across institution when evaluated individually in models: student-athlete gender u_1 = 0.114 (SE = .029), student-athlete race u_2 = 0.106 (SE = .031), high-profile sport u_3 = 0.104 (SE = .028), and the logarithm of the proportion of the neighborhood that is non-white u_{11} = 0.072 (SE = .026).

Model 14m then included random slopes for all variables that had a significant error term in Models 14a – 14l. Laplace estimates for the standard errors could not be produced. The results, however, of the Penalized Quasi-Likelihood estimation, which has been comparable to the results of the Laplace estimation in the previous models, indicated that the random errors of student-athlete race u_2 = .083 χ^2 (324) = 368.60, p = .04, high-profile sport u_3 = .080 χ^2 (324) = 356.23, p = .11, and the logarithm of the proportion of the neighborhood that is non-white u_{11} = .077 χ^2 (324) = 348.01, p > .17 are nonsignificant. The Laplace reliability estimates for these random coefficients were 11.1, 9.7, and 2.5 respectively, which supports the assessment of nonsignificance. In Model 14n, these were removed, and the slope for male was reassessed. It continued to vary significantly across institution. Model 14o removed the covariates without a significant fixed and/or random effect, and this became the final random coefficient and intercept model (see Table 16). Core high school units and being male had negative relationships with six-year degree attainment while being a high-profile athlete and HSCGPA both increased the odds of graduating. Holding all else constant, for example, the odds of a high-profile athlete graduating within 6 years are 1.3 times that of an athlete in a non-high-profile sport.

Development of the level 2 model. Although it is more common to inspect the significance of the level 2 variables on covariates with a random slope, exploratory models of level 2 variables were run for the intercept and each of the significant level 1 covariates from Model 14o. Model 15a added level 2 variables to the intercept and found a main effect of MSI γ_{01} = -0.359, $p < .01$ and the total athletics expenses of the institution γ_{06} = 6.0E-6, $p < .01$. This process was then carried out for the covariates (see Appendix C). High-profile sport status has a significant cross-level interaction with student body graduation rate γ_{24} = 0.015, $p < .01$. The relationship between the other level 1 variables and graduation were not dependent upon any of the level 2 variables. In Model 15f, in addition to the main effects of MSI and total athletics expenses, a main effect of student body graduation rate was included in the model due to its significant cross-level interaction with high profile status. With each included simultaneously in a model, the main effect of total athletics expenses was no longer significant γ_{03} = 3.0E-6, p = .01. The main effect of student body graduation, however, was significant in this model γ_{02} = .010, $p < .01$; although, the cross-level interaction was now nonsignificant γ_{21} = .005, p = .01. Model 15g removed the main effect of total athletics expenses and the cross-level interaction between high-profile status and student body graduation rate. The results of the final HGLM of six-year degree attainment in response to RQ1, Model 15g, are provided in Table 17.

Table 17. Multilevel Logistic Model Final Results: Six-year Degree Attainment as Outcome

	Model 15g
FIXED EFFECTS	
γ_{00} Intercept (6 yr degree)	0.392 (.035) OR: 1.480
γ_{01} MSI	-0.335* (.099) OR: 0.715
γ_{02} **Student body graduation rate**	0.014* (.002) OR: 1.014
γ_{10} Male	-0.424* (.043) OR: 0.655
γ_{30} High-profile sport	0.259* (.035) OR: 1.295
γ_{40} **HSCGPA**	0.538* (.032) OR: 1.712
γ_{40} **Core units**	-0.024* (.008) OR: 0.976
RANDOM EFFECTS	
σ^2_{u0} Intercept	0.1658
σ^2_{u1} Male	0.1134*
σ^2_{u01} Intercept-Male cov (SE)	-0.0693 (.022)
GOODNESS OF FIT STATISTICS	
Deviance	57868.91
# of parameters	10
Chi-square results	119.01*
Added Explained variance LV2^	33.3%
NOTE: * indicates statistically significant finding, p<.01; ^ indicates compared to immediately previous model; **Bold** indicates a grand-mean centered variable; SE in parentheses	

Response to research question 1 (six-year degree attainment). According to the data and the multilevel results, student-athlete gender, high profile sport status, HSCGPA, and total core units help to significantly predict graduation within 6 years. Equations 13 – 15 and 16a – 16e show the final level 1 and level 2 models respectively. Variables that are displayed in bold were grand-mean centered.

$$Prob(GRADUATE_{ij} = 1|\beta_j) = \phi_{ij} \tag{13}$$

$$\log\left[\frac{\phi_{ij}}{1-\phi_{ij}}\right] = \eta_{ij} \tag{14}$$

$$\eta_{ij} = \beta_{0j} - \beta_{ij}MALE_{ij} + \beta_{2j}\,HIGHPROFILE_{ij} + \beta_{3j}\boldsymbol{HSCGPA}_{ij} - \beta_{4j}\boldsymbol{TOTUNITS}_{IJ} \tag{15}$$

$$\beta_{0j} = \gamma_{00} - \gamma_{01}(MSI_{\boldsymbol{j}}) + \gamma_{02}\boldsymbol{SB\ GRAD\ RATE_j} + u_{0j} \tag{16a}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{16b}$$

$$\beta_{2j} = \gamma_{20} \tag{16c}$$

$$\beta_{3j} = \gamma_{30} \tag{16d}$$

$$\beta_{4j} = \gamma_{40} \tag{16e}$$

There was a significant variance in intercepts across institutions, u_{0j} = 0.1658, SE = .025. The probability of graduation within six years for a female student-athlete in a sport other than basketball who had average incoming HSCGPA and core units and who attends an institution that is not designated MSI and that has an average student body graduation rate is 59.7%. The slope of student-athlete gender also varied across institutions u_{1j} = 0.1134, SE = .0286.

Among the demographic variables, males had a negative relationship with degree attainment while being in a high-profile sport had a positive relationship. Both HSCGPA and total core units were significant – HSCGPA having a positive relationship and total units a negative. Of primary interest to this study were the neighborhood characteristics, none of which were significantly related with six-year degree attainment after controlling for the other covariates.

Model 15g shows that the intercept depends on the institutional MSI designation and the grand-mean centered student body graduation rate of the institution. The negative relationship of the interaction between the intercept and MSI designation implies that student-athletes attending a minority-serving institution have a reduced probability of graduating, holding the student body graduation rate constant. The intercept also is dependent on student body graduation rate – as the graduation rate of the student body increases, so too does the probability that a student-athlete will earn a degree within six years.

Research question 2. The second research question asks whether the U.S. census block group data is related to NCAA Division I student-athlete six-year degree attainment differently for student-athletes who participate in academically at-risk sports versus their counterparts. Although none of the neighborhood characteristics were significant in Model 15g, high-profile sport was. As mentioned earlier, high-profile sport designation is equivalent to an academically at-risk designation. For exploratory purposes, Model 16a used Model 15g as a basis and added an interaction term between high-profile sport and each of the neighborhood covariates. None produced a statistically significant coefficient (see Table 18). In Model 16b, the slopes for each were allowed to vary, and again, none produced significant results (see Table 19).

Table 18. Characteristics of Interaction Terms from Model 16a

	Coefficient	Odds Ratio (CI)	Significance
High-profile * Education factor	-0.027	0.973 (0.891, 1.063)	.54
High-profile * Male work FT	0.005	1.005 (0.995, 1.015)	.35
High-profile * Female work FT	0.004	1.004 (0.995, 1.013)	.37
High-profile * Couple HOH	-0.003	0.997 (0.989, 1.005)	.50
High-profile * Logarithm of neighborhood non-white	0.143	1.153 (0.961, 1.384)	.13
High-profile * Median income	0.002	1.002 (0.998, 1.006)	.29

Table 19. Random Effects of Interaction Terms from Model 16b

	Coefficient	SE
σ^2_{u5} High-profile * Education factor	0.020	1.964
σ^2_{u6} High-profile * Male work FT	3.0E-3	.028
σ^2_{u7} High-profile * Female work FT	3.0E-3	.028
σ^2_{u8}High-profile * Couple HOH	4.0E-4	.022
σ^2_{u9} High-profile * Logarithm of neighborhood non-white	0.011	.010
σ^2_{u10} High-profile * Median Income	2.0E-3	.001

In response to RQ2, the relationships between neighborhood characteristics, defined using U.S. census block group data, and six-year degree attainment do not differ by academic risk status of the sport.

Research Question 3. The third and final research asks whether the U.S. census block group data is related to NCAA Division I student-athlete six-year degree attainment differently for white student-athletes compared with non-white student-athletes.

Model 15g again was used as the basis for the analysis of this research question. An interaction between student-athlete race (non-white) and each of the neighborhood covariates was entered into the model. None were found to be significant by statistical standards. In a second step, the slopes for each were allowed to vary, and again, none produced significant results. Results for each can be found in Tables 20 and 21 respectively.

Table 20. Characteristics of Interaction Terms from Model 17a

	Coefficient	Odds Ratio (CI)	Significance
SA non-white * Education factor	-0.055	.946 (0.871, 1.029)	.20
SA non-white * Male work FT	0.003	1.003 (0.994, 1.013)	.47
SA non-white * Female work FT	0.006	1.006 (0.993, 1.015)	.23
SA non-white * Couple HOH	-0.008	.992 (0.987, 0.998)	.01
SA non-white * Logarithm of neighborhood non-white	-0.029	0.971 (0.829, 1.139)	.72
SA non-white * Median income	0.005	1.005 (1.000, 1.010)	.05

Table 21. Random Effects of Interaction Terms from Model 17b

	Coefficient	SE
σ^2_{u5} SA non-white * Education factor	0.031	.061
σ^2_{u6} SA non-white * Male work FT	4.0E-3	.027
σ^2_{u7} SA non-white * Female work FT	2.0E-3	.002
σ^2_{u8} SA non-white * Couple HOH	2.0E-4	.014
σ^2_{u9} SA non-white * Logarithm of neighborhood non-white	0.018	.203
σ^2_{u10} SA non-white * Median Income	1.0E-3	.002

In response to RQ3, the relationships between neighborhood characteristics and six-year degree attainment do not different based on student-athlete race.

Summary

In summary, the relationship between seven neighborhood characteristics (educational attainment of the neighborhood, male and female employment, median income of the neighborhood, the proportion with a couple as head of household, the racial demographics of the neighborhood, and median tenure of the residents) and three college outcomes (first-year GPA, first-year retention, and six-year degree attainment) were tested. Simple inferential analyses indicated no meaningful relationship between the neighborhood characteristics and first-year retention. Because of that, further modeling was not done. Ultimately, only the educational attainment of the neighborhood was found to have a small, but positive and significant relationship with first-year GPA.

In response to RQ1, the educational attainment of the neighborhood is positively related to first-year GPA. Holding all else constant, a one-unit increase in the derived neighborhood education factor predicts an increase of .05 in first-year GPA. The relationship between neighborhood education and first-year GPA did not vary across institutions, nor did it have a cross-level interaction with any of the institutional characteristics. The other covariates shown to have a significant relationship with first-year GPA are male (negative), non-white (negative), HSCGPA (positive), total core units (negative), and best standardized test (positive). The intercept did vary across institutions and was dependent upon the institutional student body graduation rate. As that increased, the mean first-year GPA decreased. The relationship between race, sport, and HSCGPA also varied across institutions, and the relationship between HSCGPA and first-year GPA is dependent upon the institutional student body graduation rate. As that increases, the relationship between HSCPA and first-year GPA lessens. A standardized model shows that the relative importance of the neighborhood education factor is greater than that of student-athlete race and total core high school units, but that HSCGPA really drives the model.

Regarding six-year degree attainment, the neighborhood characteristics did not have a significant relationship – neither fixed nor random. The covariates shown to have a significant relationship with six-year degree attainment are male (negative), high-profile sport (positive), HSCGPA (positive), and total core units (negative). The intercept did vary across institutions and was dependent upon the institution's MSI status and the student body graduation rate. The likelihood of graduation is lower for student-athletes at an MSI, but as the student body graduation rate increases, the likelihood of student-athlete degree attainment also increased. The relationship between gender and degree attainment also varied across institutions.

In response to RQ2, the relationship between the neighborhood education factor and first-year GPA does differ by the academic risk status of the student-athlete's sport. For student-athletes who compete in a sport deemed at greater academic risk, the relationship between the neighborhood education factor and first-year GPA is reduced by .03 holding all else constant.

Practically, this means that the relationship between first-year GPA and the level of education in the neighborhood is stronger for student-athletes who do not participate in an academically at-risk sport. This interaction did not vary across institutions, nor was it dependent upon any of the institution-level variables. The relationships between neighborhood characteristics and six-year degree attainment do not differ by academic-risk status.

In response to RQ3, the relationships between neighborhood characteristics and first-year GPA and neighborhood characteristics and six-year degree attainment do not differ by student-athlete race.

CHAPTER FIVE

IN WHICH THE DISCUSSION OCCURS

The purpose of this study was to explore the relationship between neighborhood characteristics, using U.S. census block group data, and college academic outcomes among a representative national sample of NCAA Division I student-athletes. Secondary data and multilevel analyses accounting for the grouping of student-athletes within institutions was used. The outcomes considered in this study were student-athlete first-year GPA, first-year retention, and six-year baccalaureate degree attainment. This chapter begins with an overview of the findings. A discussion of the practical implications, the limitations of the research, and recommendations for future study follows.

Overview of Results

Overall, the results from this study provide evidence that the educational attainment of the neighborhood in which the student-athlete lived prior to enrolling in college is positively related to their first-year grades. The other neighborhood characteristics of interest, however, including adult employment, income, heads of household, racial composition of the neighborhood, and tenure of the residents were not significantly and meaningfully related to any of the three outcomes – first-year GPA, first-year retention, or six-year degree attainment. In their extensive review of the experimental, quasi-experimental and observational research that has been done concerning neighborhoods and educational outcomes, Burdick-Will and colleagues (2011) summarized that while the data indicate that neighborhoods do not *"always"* matter for children's educational outcomes, they also reject the premise that they *"never"* matter (p. 255). The findings from this study support that assertion.

The discussion of the findings is organized around two main headings. The first is a broad discussion of the limited role of the neighborhood in the outcomes of interest. This discussion includes an assesment of each neighborhood covariate included in this study. The second main heading provides a direct response to each of the RQs.

Role (and lack thereof) of Neighborhood Characteristics

Prior to addressing each of the RQs and outcomes below, it seems necessary to address separately the lack of significance found between the neighborhood characteristics and the three outcomes. While the derived educational attainment factor did have a significant, positive relationship with first-year GPA (γ_{70} = 0.045, p < .01), there was no other significant and meaningful relationship found between the neighborhood measures and the outcomes in the multilevel analyses after controlling for the student-level and institution-level characteristics. There are several potential explanations for this, including explanations of a global nature that address the data and population used in the study, as well as explanations that apply more singularly to the precise covariate-outcome relationship that was being tested. The following sections look first at the neighborhood covariates included in the study and provide a broad discussion of potential explanations for the findings presented in Chapter Four. A discussion of the results of the analyses between the neighborhood characteristics and first-year retention and the unique elements of the population used in the study follows.

Neighborhood measures included in the study. Because much of the research on neighborhood effects and education has been done at the primary and secondary level, this research largely relied upon past studies of neighborhood effects and high school academic outcomes and purported theories for relationships found between the two as the basis for an exploratory study. Those studies that do look at educational outcomes beyond high school tend to focus on college enrollment or overall number of years of education as the outcomes. Where applicable, these studies, too, helped to inform the present study.

Socio-economic status. The findings from the present study's individual inferential analyses between measures of education, employment, median income, and couple head of household and the three outcomes of first-year GPA, first-year retention, and six-year degree attainment were mixed. While all the covariates were related to first-year GPA, none were related to first-year retention, and all but the proportion of females who work full-time were related to six-year degree attainment. Those relationships that were significant by statistical standards all had small effect sizes. Within the multilevel analyses, however, the only significant relationship between a neighborhood characteristic and outcome was between the neighborhood education factor and first-year GPA.

The use of SES varies greatly within the literature, including both the measures that are used and the way in which they are operationalized – either as a composite factor or individual scores. One study found a relationship between the proportion of residents with a white-collar occupation and high school graduation, while there was no relationship between graduation and the median education of the neighborhood, median income, or percent living below the poverty line (Ensminger et al., 1996). Because of this evidence that perhaps only certain aspects of SES are predictive of educational outcomes and because this study was an exploratory study, the measures of SES were included in the model individually as opposed to a composite factor.

It is difficult to directly compare the findings from this study and those in the literature because of the different operationalizations of SES and because of the differences in the population, the models, and the outcomes studied. The present study's primary findings include a positive relationship between the general education of the neighborhood and first-year GPA and no relationship between employment, income, or head of household and first-year GPA or six-year degree attainment. Wodkte and colleagues (2011) used a composite measure of SES that included measures of poverty, unemployment, welfare, heads of household, educational attainment, and proportion in a managerial position. In their study they found that adolescents living in a disadvantaged neighborhood were significantly less likely to graduate high school compared to their peers in a more advantaged neighborhood. In addition to the differences in the measure of SES between the Wodtke et al study and the present study, their study also focused on sustained exposure to a neighborhood. The present study, on the other hand, captures snapshot information for the year prior to enrolling in college. Their educational attainment outcome, high school graduation, also is much more closely related chronologically to the period of time living in the neighborhood than is the present study.

A study by Vartanian and Gleason (1999) provides another example that highlights the difficulty in directly comparing the present study's findings to those in the literature. They studied the relationship between neighborhood SES and college graduation among a population of students who were in high school between 1968 and 1981. Socio-economic status was a composite measure of the financial health of the neighborhood, employment and head of household. They did find a significant relationship between SES and college graduation, however, only for white, affluent students. There was no relationship among the African-

American sample. While the Vartanian and Gleason study is more directly related to the present study than many others given their outcome of interest, a significant point of departure between their study and the present study is their control for familial characteristics, which led to the significant findings for a certain demographic only.

Despite limited opportunities for directly relating the results of the present study with those of past studies, past literature does help to illuminate the present study's findings. An important aspect, for example, of the present study was its interest in exploring the relationship of the main effects of the various neighborhood characteristics on the outcomes. Burdick-Will et al. (2011) uncovered in their review of the literature the importance of cross-interactions among the SES variables. They explain, for instance, the feasibility that a neighborhood with a particularly low median income may also have a relatively high proportion of families with a co-parent head of household that acts as a protectant against the negative effects (direct or indirect) of being in a low-income neighborhood. While the SES components in this study all were positive and therefore would not have a canceling out effect among them, once they all were added to a model, the average education of the neighborhood became the most relevant to the outcome of first-year GPA, and the others were not significant.

Racial composition of the neighborhood. The results of bivariate correlation and independent sample t-tests respectively in the present study provided evidence for a significant and negative relationship between the proportion of the neighborhood that is non-white and both first-year GPA and six-year degree attainment. The effect size for both was small, but particularly so for six-year degree attainment. After controlling for student-level characteristics in the multilevel analysis, however, racial composition of the neighborhood was not significantly related to either outcome. The preliminary analyses also resulted in a nonsignificant relationship between the racial composition of the neighborhood and first-year retention.

In the broad neighborhood effects literature that extends to outcomes beyond education, racial segregation was found to be a consistent characteristic of the neighborhood related to outcomes of interest that include things such as low birthweight, teenage pregnancy, and childhood delinquency (Sampson et al., 2002). Individual race also has been shown to be a significant predictor of individual academic outcomes (Casselman, 2014; U.S. Department of Commerce, 2016; U.S. Department of Education, 2016b). For these reasons, this was an included neighborhood covariate in the study. While Niewenhuis and Hooimeijer (2016) found a negative relationship between the proportion of ethnic groups and educational achievement of the neighborhood, the key difference between that study and the current one is its focus on aggregate outcomes at the neighborhood level. Studies of individual outcomes have not found a significant link between neighborhood racial composition and individual academic outcomes (Ainsworth, 2002), and this study would support those findings. Sharkey and Faber (2014) theorize that a lack of racial variation within neighborhoods and the strong correlation between racial composition and other important factors, including average income and educational attainment, helps to explain why there is not a stronger connection between the racial composition of the neighborhood and academic outcomes. The findings of Sampson et al (2002) support this theory. Their work found that one of a few consistent findings across studies of neighborhood effects is the relationship between poverty and racial segregation, particularly large concentrations of African-Americans.

***Residential tenure in the neighborhood*.** The median years of residency in the neighborhood was not significantly related to first-year GPA or six-year degree attainment in the present study's preliminary analyses. There was a significant relationship by statistical standards between residential tenure and first-year retention in an independent t-test; however, the effect size was close to zero.

There are two ways that residential tenure could be defined. The first is the length of time the individual of interest has lived in the neighborhood, and the second is the typical length of time surrounding residents have lived in the neighborhood. For both, the explanation is that the longer an individual is exposed to a phenomenon, the greater the impact it can have (Ainsworth, 2002; Sharkey & Faber, 2012). Several studies have shown a relationship between the personal length of time an individual spends in a neighborhood and academic outcomes, including high school graduation (Crowder & South, 2011; Wodtke et al., 2011) and college enrollment (Chetty at al., 2015). This study, however, operationalized residential tenure as the median number of years a typical neighborhood resident occupied a dwelling. Ainsworth (2002) used a similar definition and found no relationship between it and 10^{th} grade standardized test scores.

One potential explanation for a lack of relationship between residential tenure and the outcomes is the varying importance the students place on their neighborhoods. According to Furstenburg, Jr. and Hughes (2000), selection bias within neighborhoods applies not only to where a family chooses to live but also how they interact with their surrounding neighborhood. Families and/or children may concentrate their energies and attention to other communities outside of their neighborhood. These communities, then, may become more relevant in shaping the students' outcomes. This is particularly relevant to the population of interest in this study with sport specialization on the rise and with younger onset times (Brenner, 2016). The leagues these future NCAA student-athletes join often are year-round, involve travel, and are not linked to the student's school or neighborhood (Brenner, 2016). These leagues, then, create opportunities for the students and their families to spend more time away from the neighborhood in which they live and form connections with other communities. Another interesting potential explanation addresses the increase in social media platforms, access to social media and increased use of cell phones at younger and younger ages. These technologies are shrinking the world around us and may reduce the influence of the immediate geographical boundaries (Cook, Shagle, & Degirmencioglu, 2000).

First-year retention and an NCAA student-athlete population. The preliminary analyses assessing the individual relationships between first-year retention and the covariates indicated that there was no significant and meaningful relationship between retention and the neighborhood characteristics. When thinking about these findings, it is important to keep in mind the population on which the study is focused – NCAA Division I student-athletes – and how retention is defined. For the purposes of this study, it is defined as persisting to the second year at the original institution of enrollment *and* maintaining an athletics scholarship. It is possible that some in the sample will persist but will no longer be competing on aid. It also is possible that, with a broader college student sample that included or was restricted to non-athletes, a relationship may be found. There are several factors unique to the NCAA Division I student-athlete experience that may be preventing a relationship between neighborhood characteristics and first-year retention. Those include PTD requirements, the APR, and transfer rules.

Progress-toward-degree requirements stipulate the in-college benchmarks student-athletes must meet regarding grades, non-remedial credit hours earned, and the percent of their degree that has been completed per academic year (NCAA, nd). If followed, these requirements will set student-athletes up to graduate within no more than 5 years of their initial enrollment. While it certainly does not guarantee success, it does help to ensure student-athletes do not get unwittingly left behind. Moreover, failure to meet PTD could result in loss of the student-athlete's athletics scholarship, which in turn, could affect his or her ability to remain at the institution.

While PTD standards do not apply directly to retention persé, the association's APR standards and transfer rules are directly related to retention. The APR requires that, upon conclusion of an academic term, student-athletes are academically eligible to continue (using PTD standards) and are retained to the next term. If a student-athlete fails to meet either, points are deducted from a team composite score. If that score dips below a pre-determined threshold, team penalties, including loss of scholarships and loss of post-season competition, are implemented.

Finally, certain sports have transfer rules that require that the student-athlete take a year in residence after transferring to a new institution before they are eligible to compete. Sports that are affected by this rule include baseball, men's and women's basketball, football and men's ice hockey.[1] These student-athletes, therefore, face a direct consequence if they transfer. A consequence not faced by college students who are not NCAA athletes.

---

[1] The student-athlete may be immediately eligible to compete if s/he participate in one of the listed sports and was not recruited by the initial institution of enrollment and did not receive an athletics scholarship. Waiver opportunities to bypass this year-in-residence also are available on a case-by-case basis.

While the present study was concerned with using pre-college indicators to predict college outcomes, considering the nonsignificant findings among retention and neighborhood, it is important to consider the unique in-college circumstances of the student-athlete population that may hinder the relationship. A great deal of the research offered by Tinto (1975), Pascarella and Terenzini (1980), Astin (1993), and Kuh et al. (2008) addresses the importance of in-college activities in persistence. Social integration, in particular, has been found to be one of the most important components of retention (Kuh et al., 2007). Over three-quarters of Division I student-athletes have reported that they have sense of belonging to their campus; roughly one-half to two-thirds report frequently socializing with non-athletes, and the overwhelming majority feel supported by their faculty, student body and president or chancellor. They also feel supported by their teammates and coaches, and with over 40 hours spent with them each week, they have a significant support network as student-athletes (NCAA, 2016b).

Responses to the Research Questions

RQ1: Relationship between neighborhood characteristics and outcomes. The first RQ asks if U.S. census block group data is significantly related to NCAA Division I student-athlete first-year college cumulative grade point average, first-year retention and eventual six-year degree attainment after controlling for student-athlete demographics and pre-college academic characteristics and college-level institutional characteristics. It was hypothesized that after controlling for student-athlete demographics and pre-college characteristics that the neighborhood characteristics of SES, racial composition and residential stability would significantly contribute to first-year GPA, first-year retention and six-year graduation. It was further hypothesized that these neighborhood characteristics would remain significant after the inclusion of college-level institutional information. The discussion herein is restricted to the outcomes of first-year GPA and six-year degree attainment.

First-year GPA. The main response to RQ1 and first-year GPA is that the educational attainment of the neighborhood has a small but significant and positive relationship with student-athletes' first-year grades after controlling for individual gender, race, high-profile sport status, HSCGPA, total core HS units and best standardized test. The neighborhood variables were added to a fixed model with a random intercept that already included the student-athlete individual demographic and pre-college academic characteristics. The inclusion of the neighborhood variables in Model 4 did improve the overall model when compared with Model 3; however, the additional explained variance at level 1 was quite small – just under 1%. Practically, a strong argument could be made that the additional burden of using the U.S. census block group data does not outweigh the benefits in the predictive ability of the model. A standardized final model, Model 7, allowed for the comparison of the relative strength of the individual covariates in predicting first-year GPA. The student-athletes' HSCGPA drove the model. It was nearly three times as predictive as best standardized test, which had the second greatest standardized value. In relation to the other covariates, neighborhood education had a slightly stronger relationship with first-year GPA than did individual race or the total core units earned in high school. The findings of the positive relationship between educational attainment in the neighborhood and first-year GPA support the findings of Brooks-Gunn and colleagues (1993) that it is the presence of affluence within a neighborhood, as opposed to the deprivation, that is most relevant for academic outcomes.

Past models have shown the importance of both high school academic preparation and individual demographics in college academic performance. The results of this study confirm those general findings, and they confirm the findings regarding the relative importance of past academic performance in relation to individual demographics (Adelman, 2006; Geiser & Santelices, 2007). While the negative relationship between the total core units a student-athlete takes in high school and first-year GPA contradicts the work of Pike and Saupe (2002), it very likely is due to unique traits of that covariate and the difference in the measurement of core units used in the present study and those used in the literature. Pike and Saupe used a dichotomous variable that captured whether the student had completed the high school's core curriculum. In the present study, the total number of core units on a continuous scale was used. Two unique aspects of this variable are worth noting. First, there is minimal variation in the covariate. At the time these student-athletes were admitted, to be academically eligible to participate in NCAA Division I sports, a student-athlete must have earned 16 core courses by the time of their high school graduation (Petr & Paskus, 2009). The range of core units in this dataset was 6.5 – 30; however, 85% have a value of 14 – 20. Of perhaps greater note, however, is the way in which the HSCGPA is calculated and its relationship with total core units. The calculation of HSCGPA typically is restricted to the 16 core units needed for eligibility and incorporates the 17th unit and beyond only if helps the student-athlete's HSCGPA. The negative correlation between HSCGPA and total core units (r = -.203, p < .01) provides evidence that those student-athletes who have greater core units on their record may take these courses because they are using these extra units to increase their HSCGPA. A negative relationship, therefore, between total core units and first-year GPA could be the product of student-athletes who were less academically prepared and/or historically struggle with their GPAs.

While there is a great deal of literature on the prediction of student-athlete academic performance, very few have considered the multilevel nature of the data and used multilevel analyses to account for the nesting of student-athletes within institutions. This study explored the need for a multilevel analysis by running an intercept-only or null model and found that 5.6% of the variation in the outcome of first-year GPA could be attributed to the clustering of student-athletes within institution. As part of this multilevel analysis, it was found that the intercept, student-athlete race, high-profile sport status, and HSCGPA all varied across institutions. At level 2, the student body graduation had a significant main effect and had a significant cross-level interaction with HSCGPA. As the student body graduation rate increased, the mean first-year GPA decreased. This is consistent with the findings of McArdle et al. (2013), which is the one study that most closely resembles the present study. They theorized that perhaps institutions with greater student body graduation rates are more selective with stricter grading practices; therefore, as the graduation rate increases, the intercept of the first-year grades would decrease. Although, not found in the McArdle et al study, in the present study, the student body graduation rate also had a negative relationship with the slope of HSCGPA. As the graduation rate increased, the relationship between HSCGPA and first-year GPA lessened. A potential explanation for this could be that once the student successfully enrolls in an institution, their first-year grades become more a reflection of the rigor of the institution in which they enroll (defined by the student body graduation rate) and less of a reflection of their academic performance in high school. Like McArdle et al (2013), no other level 2 variable had a significant cross-level interaction once the student body graduation rate was included. This could be due to the relationship between the graduation rate and other institutional characteristics. There was a strong relationship between the graduation rate and MSI (M = 34.8, SD = 11.5) vs others (M = 65.2, SD = 17.0), $t(39) = 12.74$, $p < .01$, d = 1.83, and between graduation rate and public (M = 56.9, SD = 17.7) vs private (M = 74.5, SD = 14.6), $t(325) = -8.91$, $p < .01$, d = 1.05. There also were significant bivariate correlations, which can be seen in Table 19.

Table 22. Pearson correlation coefficients: Level 2 institutional covariates

	Enrollment	Out of State GIA	Total Athletics Expenses
Student body graduation rate	.157*	.665*	.463*
Note: * indicates statistically significant correlation (two-tailed), p <.01			

Six-year degree attainment. Findings from the present study indicate that after controlling for student-athlete gender, high-profile sport status, HSCGPA, and total core units, neighborhood characteristics are not predictive of six-year degree attainment among the student-athlete sample used in the study. The significant findings do support prior work that found a negative relationship with being male (U.S. Department of Education, 2016b) and a positive relationship with HSCGPA (Geiser & Santelices, 2007). Like the first-year GPA model, total core units had a negative relationship with six-year degree attainment; although the effect size is quite small OR = 0.976. Student-athletes who participate in a high-profile sport have an increased likelihood of graduation, all else constant. The work of McArdle and Hamagami (1994) support this finding. In their multilevel study of student-athlete degree attainment, student-athletes in the sports of men’s basketball and football were more likely to graduate after controlling for HSCGPA, best test, race, gender, and whether they were on the travel squad, which was a proxy for athletics ability. The reason for this could be due to the relative support offered to high-profile student-athletes when compared with other student-athletes. The sports of men's and women's basketball and football, along with a few others, are full scholarship sports, meaning that student-athletes receive the full tuition and room and board for their athletics participation. Data also shows that tutorial support is provided and required more frequently for student-athletes in high-profile compared with the others. Finally, race and best test both were nonsignificant. McArdle and Hamagami (1994) also found race to be nonsignificant in their multilevel model of student-athlete degree attainment. Best test, however, was significant in their final model after controlling for gender, sport, HSCGPA, and travel team status. One key difference between the McArdle and Hamagami study and the present study that likely accounts for the difference in best test significance is the presence of initial eligibility criteria. McArdle and Hamagami's study was done on a sample of student-athletes who were admitted prior to the introduction of initial eligibility. The guidelines required at the time the present sample was admitted required a 16-core course minimum and best test-HSCGPA combined minimum. These student-athletes also were held to PTD standards and the APR criteria. Taken together, these likely diminished the effect of best test on six-year degree attainment.

As part of the multilevel analysis, it was found that the intercept and gender varied across institutions. At level 2, MSI status and student body graduation rate had a significant relationship with the intercept. Likelihood of graduation is reduced for student-athletes attending an MSI, which is counter to what has been reported in the literature. In their review of the literature, for example, Kuh at al (2007) report a positive relationship between Historically Black College and University (HBCU) status and degree attainment. These findings, however, are most often restricted to an African-American sample and are compared to what the outcomes are for African-Americans at Predominantly White Institutions. Within the NCAA, student-athletes at HBCUs have historically underperformed compared to their peers at non-HBCU institutions using the APR as a common metric (Paskus, 2012). Resources often are highlighted as a likely function of this underperformance (Paskus, 2012). Similar to the model of first-year GPA, the intercept also had a significant relationship with student body graduation rate. Unlike the first-year GPA model, however, the relationship between the intercept and student body graduation rate was positive. As posited by McArdle et al (2013), the negative relationship between first-year GPA and graduation rate may be due to stricter grading practices at more selective schools. A positive relationship, however, would be expected between the graduation rates of the student body and student-athletes. Data for the 2007 – 2010 entering classes show that four-year average graduation rates for the student body and student-athletes are highly correlated, $r = .786$, $p < .01$, perhaps indicating an institutional culture or expectation around degree attainment.

RQ2: The role of sport. The second RQ asks if the relationship between the U.S. census block group data and the outcomes are different for student-athletes who participate in a sport deemed at high academic risk compared with their peers in other sports. For the purposes of this study, the high-risk sports were baseball, men's and women's basketball, and football. It was hypothesized that the neighborhood characteristics of SES, racial composition and residential stability will be related to student-athletes' first-year GPA, first-year retention, and six-year graduation comparably for student-athletes participating in academically at-risk sports and their counterparts in sports not deemed academically at-risk. The discussion herein is restricted to the outcomes of first-year GPA and six-year degree attainment.

First-year GPA. In response to RQ2, the relationship between the neighborhood education factor and first-year GPA does differ by the academic risk status of the student-athlete's sport. For student-athletes who compete in a sport deemed at greater academic risk, the relationship between the neighborhood education factor and first-year GPA is reduced by .03 holding all else constant. Practically, this means that the relationship between first-year GPA and the level of education in the neighborhood is stronger for student-athletes who do not participate in an academically at-risk sport. One potential explanation for this differential relationship could be due to differences in academic support provided to these high-risk sports compared to the others. A 2008 study of Division I academic support services, for example, found that tutorial services are required more often for student-athletes deemed academically at-risk, and regardless of risk, are required more often for men's and women's basketball players (T. Petr, personal communication, February 20, 2018). This same study also found that over one-quarter of tutorial support budgets are directed to the support of men's football programs, and one-half of the support given to women's teams goes to women's basketball with the remaining one-half spread among the other 20 teams. With the comparative amount of support given to the high-risk sports, the role of the neighborhood in predicting first-year GPA is diminished. This interaction did not vary across institutions, nor was it dependent upon any of the institution-level variables. The inclusion of the interaction did not reduce the amount of explained variance at level 1, but it did improve the overall model when compared with the final model in response to RQ1, Model 6k.

Six-year degree attainment. Unlike first-year GPA, none of the neighborhood characteristics had a significant main effect on six-year degree attainment, nor was there an interaction effect between any of the neighborhood characteristics and academic-risk status of the sport when modeling six-year degree attainment. As previously posited, a lack of a main effect between neighborhood and six-year degree attainment potentially can be explained by the time lapse between the measure of the neighborhood and the measure of the outcome. Once enrollment and the requisite GPA to continue athletics eligibility were achieved, the differential role of the neighborhood in predicting at-risk student-athlete success was no longer relevant.

RQ3: The role of race/ethnicity. The third RQ asks if the relationship between the U.S. census block group data and the outcomes are different for white student-athletes compared with non-white student-athletes. It was hypothesized that the neighborhood characteristics of SES, racial composition and residential stability will be related to student-athletes' first-year GPA, first-year retention, and six-year graduation comparably for white and non-white student-athletes. The discussion herein is restricted to the outcomes of first-year GPA and six-year degree attainment.

As hypothesized, there was no differential relationship between the neighborhood characteristics and student-athlete race for either first-year GPA or six-year degree attainment. As a main effect, student-athlete race had a comparatively small relationship with first-year GPA (see Model 7) and had a nonsignificant relationship with six-year degree attainment. While studies with a general student population have shown race to be a significant contributor to degree attainment (Casselman, 2014; U.S. Department of Education, 2014), NCAA research has shown that the effect of race is generally accounted for by the pre-college academic characteristics (Petr & McArdle, 2012). This study's findings regarding the importance of individual student-athlete race and academic outcomes supports the general findings of past NCAA research.

Practical Implications and Contributions

When considering the practical implications of and contributions from this research, it is worthwhile to revisit the intended purpose of the study. This was exploratory in nature, aimed at assessing whether the addition of neighborhood characteristics using U.S. census block group data improved the predictive validity of college academic outcome models for an NCAA Division I student-athlete sample. The motivation behind the research questions was to determine if there was a way to improve identification of academically at-risk student-athletes with the hopes that once at-risk students are identified that then interventions can be delivered to help them succeed. The intent was not to better understand why the neighborhood characteristics may be predictive of academic outcomes or to offer policy recommendations at the neighborhood level. Both are beyond the scope of this study but are addressed in future research recommendations. There are two primary practical contributions from this study: a contribution to the NCAA and a contribution to the literature on neighborhood effects.

Contribution to NCAA

The NCAA has been making data-driven decisions regarding student-athlete academic initial and continuing eligibility in an increasing manner since the 1980s. In many ways, these initial academic eligibility decisions are akin to an institutional admission decision. A recent change to the NCAA initial eligibility rules allows for student-athletes who are deemed to be at an elevated academic risk but still show potential for academic success to be granted partial eligibility. In athletics-lingo, these student-athletes are granted an academic redshirt year in which they are awarded an athletics scholarship and are permitted to practice with the team but cannot travel or compete (NCAA, 2016d). Early identification of those who are academically at-risk and labeling them as such allows for an opportunity to address potential deficits in the student-athletes' academic preparation. Properly and thoroughly identifying risk is important not only from an individual student-athlete standpoint but also from an institutional standpoint. According to Paskus (2012), managing the collective risk being assumed by the institution will help to ensure that the institution is adequately prepared to address the risk. One suggestion offered by Paskus is that institutions cap the number of high-risk student-athletes they admit. Another is to enhance the academic support that is available to student-athletes. Many institutions try to balance building an athletically competitive program with admitting student-athletes who have the potential to succeed academically. This potential often is supported by extensive institutional academic support programs, aspects of which frequently are offered only to their student-athlete population (Rubin & Moses, 2016).

The findings from this study will be shared with the NCAA research staff, and from there, potentially with the policy bodies that decide upon the academic bylaws of the Association. While the academic world celebrates significant and meaningful findings, in this case, it is the lack of a significant and/or meaningful finding that will be the headline. There was no meaningful relationship between any of the neighborhood characteristics and first-year retention and no significant relationship between any of the neighborhood characteristics and six-year degree attainment. There was a small and positive relationship between the educational attainment of the neighborhood and first-year GPA, and that relationship was stronger for student-athletes not in an academically high-risk sport. That relationship, however, was nominal when compared with the precollege academic characteristics and added little, overall, to the explained variance of the model. When creating the initial eligibility standards, the NCAA staff and members sought to balance efficacy with simplicity (Petr & McArdle, 2012). It is important that these rules are clear to the prospective student-athletes and their families, to the high schools who help to guide them, and to the public who holds the NCAA accountable for their decisions. A cost-benefit analysis would need to evaluate the added predictive validity to the prediction model of first-year GPA against the added burden of including U.S. census block group data to the models and their transparency.

A secondary practical implication for NCAA research is the use of multilevel modeling. Although, this is not the first time multilevel modeling has been used in predictive validity studies using NCAA academic data, it still contributes to the body of knowledge in the area. The objective of using this method is to provide less biased results by considering the dependence among student-athletes who are clustered within institutions. Results from the intercept-only models in the present study provided evidence that there was an effect of the clustering of student-athletes within institutions, but that it was relatively small – 5.6% for first-year GPA and 6.3% for six-year degree attainment. McArdle et al (2013) found similar results when using pre-college academic characteristics to predict first-year GPA. In their study, the use of multilevel modeling "led to only minor alterations in the traditional regression estimates of fixed effects" (p. 89). Currently, NCAA research uses single-level multiple linear regression and generalized linear models in their predictive validity studies. The results of the current study, which support the findings of McArdle and colleagues, provides evidence that the cost-benefit analysis of multilevel modeling in establishing initial eligibility criteria may prove to be too costly.

Contribution to the Literature

The use of U.S. census data in neighborhood effects research largely has been restricted to outcomes up to and including high school graduation (Harding, 2003; Leventhal & Brooks-Gunn, 2000). The few studies that have examined college outcomes, have used college enrollment or academic attainment as the outcomes (Brattbak, 2014; Chetty & Hendren, 2015; Harris et al., 2010; Vartanian & Gleason, 1999). No known studies have used in-college outcomes, which this study addressed with the use of first-year GPA and first-year retention. Generally, the findings from the present study support some of what has been done in the literature; although, as mentioned previously, comparisons are challenging due to the differences in measures and in the population of interest. This study, for example, does support the findings of Brattbak (2014), who found, using a sample from Norway, that the average education of the neighborhood had a relationship with educational attainment after controlling for individual and familial characteristics. And, similar to Ainsworth (2002), the present study did not find a significant relationship between racial composition or residential tenure and the academic outcomes. The primary contribution of this research to the literature, however, is not in supporting evidence but in a making a unique contribution to what has been done prior by using a large and national sample that accounts for individuals’ academic backgrounds.

Limitations of the Present Study and Directions for Future Research

The findings of the present study are subject to several limitations. Perhaps the greatest limitation inherent in all non-experimental neighborhood effects research is selection bias of families into neighborhoods. According to Duncan and Raudenbush (1999), while the impact of selection bias is likely, its specific impact is uncertain. The reason for this largely goes back to the earlier discussion of interaction effects, particularly within measures of SES. Some families may be more equipped to withstand the effects of a more disadvantaged neighborhood or less equipped to take advantage of a resource-rich neighborhood (Brook-Gunn et al., 1993). This has been referred to as getting a different "dose" of the neighborhood (Harding, Gennetian, Winship, Sanbonmatsu, & Kling, 2010). Essentially, depending on personal and familial circumstances, the same neighborhood can have a very different effect on individuals.

It is difficult to discuss the effects of selection bias without also discussing familial characteristics. A second limitation of the present study is the lack of control for family characteristics, most notably measures of SES. One measure purported by Brooks-Gunn and colleagues (1993) was an ability to capture parental resilience. Likely, this could be quantified through things such as parental education, income, and employment. As they attest, without it, the "estimated effects of bad neighborhoods…will be smaller than they would have been if parents had been randomly allocated across neighborhoods," (p. 358) which speaks to both selection bias and a need to account for familial characteristics. Others have stipulated that not fully controlling for SES will inflate the role of the neighborhood (Jencks & Mayer, 1990).

While the tenure of the residents was not significant in the models, a third limitation was the inability to control for how long the student-athlete lived in the neighborhood and to account for the characteristics of other neighborhoods in which s/he may have resided. Chetty et al. (2015) and Crowder and South (2011) found a significant relationship between personal tenure in a neighborhood and educational attainment, high school graduation and college enrollment. Furstenbrug, Jr. and Hughes (2000) explain that the "impact of a particular community on a child will likely depend on the child's duration of exposure to the characteristics of that community, the ages at which it occurs, and, perhaps, the types of neighborhoods that precede and follow it" (p. 28).

Finally, there were two noteworthy methodological limitations to this study. The first is this study used outcomes that are predicated on one another. To be retained on an athletics scholarship, a certain GPA must be achieved, and to graduate, you must be retained. The dependent nature of the outcomes was not accounted for in the present study and is a recommendation for future research. Another limitation and opportunity for future research is accounting for the nesting of student-athletes within census block group area. In the current study, this variable was treated as an individual, level 1 covariate. It is possible; however, that not only did multiple student-athletes reside in the same block group, violating the independence of error assumption, but also even more likely is that student-athletes resided in very like-block group areas that may have had a similar impact on outcomes and again violated the assumption of independence of error.

The other limitations presented here also offer opportunities for future research. While accounting for selection bias of families into neighborhoods is best done experimentally (see the MTO study), and thus likely too great a burden or cost for most researchers, feasible opportunities for future research could include a more thorough accounting of the residential history and a better ability to control for familial characteristics, including parental education, income, and employment.

There were several plausible explanations for a lack of significance between the neighborhood characteristics and the outcomes of interest that were related to the population used for this study. Expanding these models to a non-athlete population would remove some of the potential confounds discussed within this chapter, most notably the comparative wealth of academic, financial and social resources afforded to student-athletes.

As stated earlier, this research is not concerned with the process – the *why* the neighborhood influences one way or another; however, this is an important next step in the research connecting neighborhood characteristics to individual academic outcomes. The present study did find a significant and positive relationship between neighborhood educational attainment and first-year GPA. What are the mechanisms that facilitate that relationship? Cook et al (2000) point out that most neighborhood research that uses neighborhood demographics assumes a mediational relationship. The theories discussed in Chapter Two all necessitate a mediational relationship. Collective socialization theory, for example, focuses on the role of neighborhood adults in creating norms and examples the neighborhood children follow. According to the theory, these adults can indirectly affect the children's outcomes by affecting their motivations and priority-setting. These mediational processes, however, are rarely tested on a grand scale due to practical limitations with data collection. Those who are imbedded in the neighborhood effects field continue to tout the importance of the structural components of the neighborhood having both direct and indirect effects (Morenoff et al., 2001) and in their usefulness when devising policy implications (Furstenberg Jr. & Hughes, 2000; Sampson et al., 2002). Harding and colleagues (2010) present a detailed examination of ways in which these processes can be captured.

SUMMARY

There are over 170,000 student-athletes who compete annually for Division I institutions. In many ways, they mirror the general university student population. In other ways, however, they are a truly unique subpopulation with distinctive features. One of the most popularly contested topics in American higher education today is the academic well-being of the student-athletes. Early identification of at-risk student-athletes can help to, upon matriculation, mitigate their risk factors by directing relevant academic support services, and in some cases, considering the appropriateness of competition in their first year.

The objective of the current study was to explore if and how neighborhood characteristics improve the predictive validity of college academic outcome models using first-year GPA, first-year retention and six-year degree attainment as the outcomes of interest. The findings indicate that while consideration of the educational attainment of the neighborhood adds to the predictive ability of first-year GPA, the meaningful impact is quite small. Cost-benefit analyses may reveal that the added burden of data collection and reduction in transparency is not worth the minimal addition of explained variance in the outcome, particularly in light of the lack of a significant relationship with the other outcomes.

APPENDIX A

DEVELOPMENT OF LEVEL 1 MODELS WITH FIRST-YEAR GPA AS OUTCOME

Models 5a – 5l were run allowing a single covariate to vary across institutions. Below are the variance components, chi-square values, and significance.

Model	Covariate	Variance Component	Chi-square	*p*-value
Model 5a	Gender (Male)	0.008	479.91	<.01
Model 5b	SA Non-white	0.014	575.70	<.01
Model 5c	High-profile sport	0.014	563.59	<.01
Model 5d	HSCGPA	0.022	786.12	<.01
Model 5e	Core units	2.0E-3	397.03	<.01
Model 5f	Best test	2.0E-4	589.21	<.01
Model 5g	Education factor	0.002	463.68	<.01
Model 5h	Male work FT	1.0E-4	375.72	.03
Model 5i	Female work FT	0.000	320.69	>.05
Model 5j	Couple HOH	0.191	506.23	<.01
Model 5k	Logarithm Neighborhood Non-white	0.009	451.79	<.01
Model 5l	Median Income	0.000	425.95	<.01

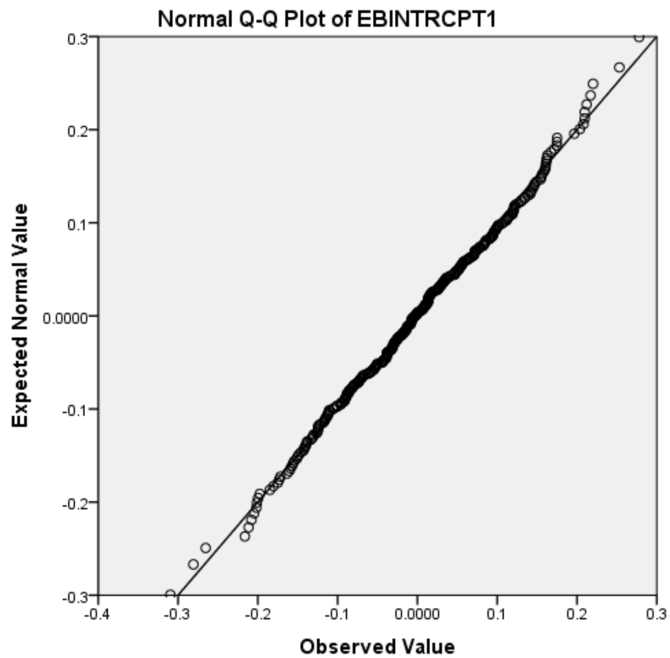
Model 5m included each of the significant variance components above into a single model.

Covariate	Variance Component	Chi-square	*p*-value
Gender (Male)	0.003	363.25	.04
SA Non-white	0.008	418.24	<.01
High-profile sport	0.010	413.98	<.01
HSCGPA	0.015	480.09	<.01
Core units	1.0E-3	348.53	.12
Best test	0.000	353.07	.09
Education factor	0.002	359.71	.06
Couple HOH	.236	376.60	.01
Logarithm Neighborhood Non-white	0.005	354.37	.08
Median Income	0.000	386.96	.01

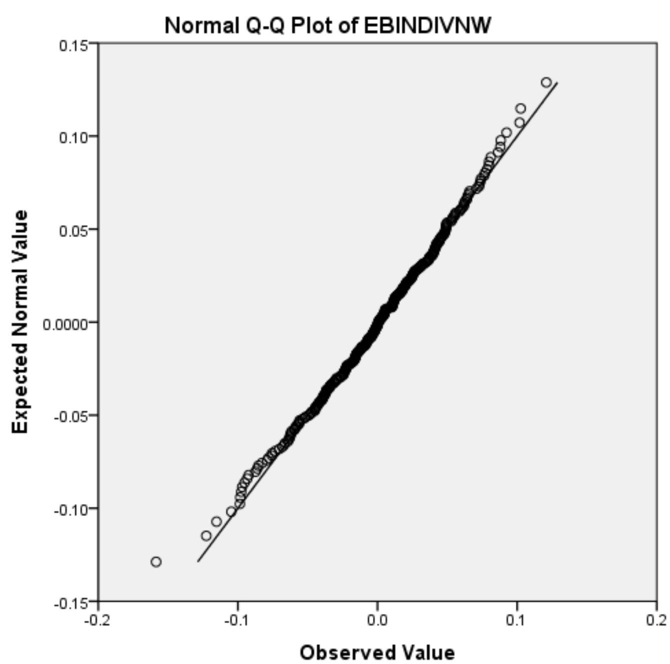
APPENDIX B

Q-Q PLOTS OF EMPIRICAL BAYES RESIDUALS FROM MODEL 6K

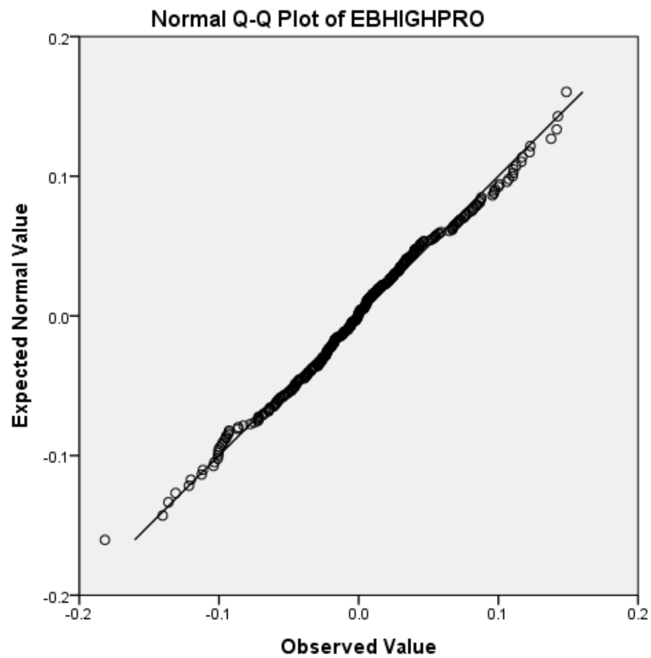
Q-Q plot of empirical Bayes residuals for the intercept



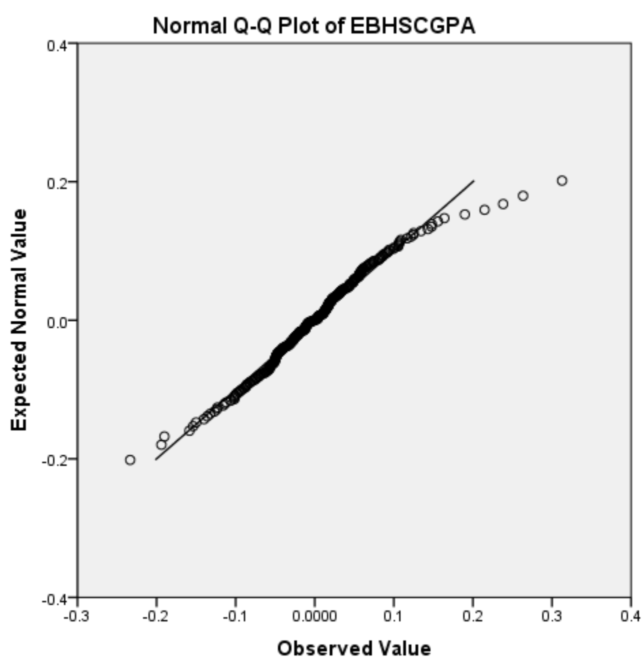
Q-Q plot of empirical Bayes residuals for student-athlete non-white



Q-Q plot of empirical Bayes residuals for high-profile sport



Q-Q plot of empirical Bayes residuals for HSCGPA



APPENDIX C

DEVELOPMENT OF LEVEL 2 MODELS WITH SIX-YEAR DEGREE ATTAINMENT AS OUTCOME

Models 15a – 15e were run adding level 2 covariates to the intercept and level 1 covariates separately. Below are the coefficients, standard errors, and significance.

Model	Level 2 Covariate	Coefficient	Standard Error	*p*-value
Model 15a: Intercept	MSI	-.359	.092	<.01
	Private	.190	.086	.03
	Enrollment	-.002	.004	.64
	Student Body Graduation Rate	.006	.002	.01
	GIA	.005	.003	.15
	Total Athletics Expenses	6.0E-6	1.0E-6	<.01
Model 15b: Male	MSI	-.267	.121	.03
	Private	.167	.093	.07
	Enrollment	-.004	.004	.31
	Student Body Graduation Rate	.007	.003	.01
	GIA	.001	.003	.81
	Total Athletics Expenses	2.0E-6	2.0E-6	.13
Model 15c: High-profile sport	MSI	-.168	.151	.27
	Private	-.245	.110	.03
	Enrollment	.004	.006	.56
	Student Body Graduation Rate	.015	.003	<.01
	GIA	.003	.006	.60
	Total Athletics Expenses	-4.0E-6	2.0E-6	.01
Model 15d: HSCGPA	MSI	.034	.174	.85
	Private	.183	.111	.01
	Enrollment	2.0E-4	.007	1.0
	Student Body Graduation Rate	-.002	.003	.43
	GIA	-.010	.005	.04
	Total Athletics Expenses	1.0E-6	2.0E-6	.56
Model 15e: Total units	MSI	-.045	.040	.27
	Private	.022	.029	.45
	Enrollment	.002	.001	.16
	Student Body Graduation Rate	-2.0E-3	.001	.68
	GIA	-2.0E-4	.001	.99
	Total Athletics Expenses	-1.0E-6	.000	.21

REFERENCE LIST

Aaronson, D. (1998). Using sibling data to estimate the impact of neighborhood on children's educational outcomes. *The Journal of Human Resources, 33*(4), 915-946. doi: 10.2307/146403

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, D. C.: U.S. Department of Education.

Ainsworth, J. W. (2002). Why does it take a village? The mediation of neighborhood effects on educational achievement. *Social Forces*, *81*(1), 117-152.

Astin, A. W. (1993). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education.* Phoenix, AZ: The Oryx Press.

Astin, A. W., Tsui, L., & Avalos, J. (1996). *Degree attainment rates at American colleges and universities: Effects of race, gender, and institutional type.* Los Angeles: Higher Education Research Institute, UCLA.

Austin, A. (2013, 22 July). African Americans are still concentrated in neighborhoods with high poverty and still lack full access to decent housing. *Economic Policy Institute*. Retrieved from http://www.epi.org/publication/african-americans-concentrated-neighborhoods/

Bauer-Wolf, J. (2017, October 16). NCAA: No academic violations at UNC. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2017/10/16/breaking-ncaa-finds-no-academic-fraud-unc

Bettinger, E. P., Evans, B. J., & Pope, D. G. (2013). Improving college performance and retention the easy way: Unpacking the ACT exam. *American Economic Journal: Economic Policy, 5*(2), 26-52. doi: 10.1257/pol.5.2.26

Berger, J. B., & Milem, J. F. (2000). Organizational behavior in higher education and student outcomes. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (268-338). New York, NY: Agathon Press.

Brattbakk, I. (2014). Block, neighbourhood or district? The importance of geographical scale for area effects on educational attainment. *Geografiska Annaler, Series B Human Geography, 96*(2), 109-125.

Braxton, J. M. (2006, June). *Faculty professional choices in teaching that foster student success*. Retrieved from https://nces.ed.gov/npec/pdf/Braxton_Report.pdf

Brenner, J. (2016). Sports specialization and intensive training in young athletes. *Pediatrics, 138*(3), e1-e8. doi: 10.1542/peds.2016-2148

Bridgeman, B., Pollack, J., & Burton, N. (2004). Understanding what SAT reasoning test scores add to high school grades: A straightforward approach. College Board Research Report 2004-4. New York: College Entrance Examination Board. Retrieved from http://files.eric.ed.gov/fulltext/ED563090.pdf

Brooks-Gunn, J., Duncan, G. T., Klebanov, P. K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology*, *99*(2), 353-395.

Brown, G. (2015, October 28). Why the GSR is a better methodology. *NCAA*. Retrieved from http://www.ncaa.org/about/resources/research/why-gsr-better-methodology

Brutlag Hosick, M. (2016, April 16). DI council adopts academic integrity proposal. *NCAA News*. Retrieved from http://www.ncaa.org/about/resources/media-center/news/di-council-adopts-academic-integrity-proposal

Burdick-Will, J., Ludwig, J., Raudenbush, S. W., Sampson, R. J., Sanbonmatsu, L., & Sharkey, P. (2011). Converging evidence for neighborhood effects on children's test scores: An experimental, quasi-experimental, and observational comparison. In G. J. Duncan & R. J. Murnane (Eds). *Whither Opportunity? Rising inequality, schools, and children's life chances* (255-276). New York: Russell Sage Foundation.

Burton, N. W., & Ramist, L. (2001). Predicting success in college: SAT studies of classes graduating since 1980. Retrieved from https://files.eric.ed.gov/fulltext/ED562836.pdf

Camara, W. (2003, March). College persistence, graduation, and remediation. *The College Board: Research Summary*. Retrieved from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchinreview-2003-9-college-persistence-graduation-remediation.pdf

Camara, W., & Echternacht, G. (2000). *The SAT[R] I and high school grades: Utility in predicting success in college*. Retrieved from http://files.eric.ed.gov/fulltext/ED446592.pdf

Casselman, B. (2014, April 30). Race gap narrows in college enrollment, but not in graduation. *FiveThirtyEight.* Retrieved from https://fivethirtyeight.com/features/race-gap-narrows-in-college-enrollment-but-not-in-graduation/

Chetty, R., Hendren, N., & Katz, L. F. (2015). *The effect of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment*. Retrieved from http://www.nber.org/papers/w21156

Cho, D. (2007). The role of high school performance in explaining women's rising college enrollment. *Economics of Education Review, 26*, 450-462. doi: 10.1016/j.econedurev.2006.03.001

Cierniak, K., Stewart, M., & Ruddy, A-M. (2015, March). *Mapping the growth of statewide voucher programs in the United States*. Retrieved from http://ceep.indiana.edu/pdf/Statewide_Vouchers_CEEP_EPB.pdf

Comeaux, E., & Harrison, K. C. (2011). A conceptual model of academic success for student-athletes. *Educational Researcher, 40*(5), 235-245. doi: 10.3102/0013189X11415260

Cook, T. D., Shagle, S. C., & Degirmencioglu, S. M. (2000). Capturing social process for testing mediational models of neighborhood effects. In J. Brooks-Gunn, G. J. Duncan & J. L. Aber (Eds.), *Neighborhood poverty: Policy implications in studying neighborhoods, Vol. II* (pp. 94-119). New York: Russell Sage Foundation.

Crowder, K., & South, S. J. (2011). Spatial and temporal dimensions of neighborhood effects on high school graduation. *Social Science Research, 40*, 87-106. doi: 10.1016/j.ssresearch.2010.04.013

Cunningham, G. B. (2012). Diversity issues in academic reform. *Journal of Intercollegiate Sport, 5*, 54-59.

Dornbusch, S. M., Ritter, P. L., & Steinberg, L. (1991). Community influences on the relation of family statuses to adolescent school performance: Differences between African Americans and Non-Hispanic Whites. *American Journal of Education, 99*(4), 543-567.

Duncan, G. J. (1994). Families and neighbors as sources of disadvantage in the schooling decisions of white and black adolescents. *American Journal of Education, 103*(1), 20-53.

Duncan, G. J., & Raudenbush, S. W. (1999). Assessing the effects of context in studies of child and youth development. *Educational Psychologist, 34*(1), 29-41.

Edelman, M. (2015, February 28). Don't feed (or shelter) the athletes: The absurdity of NCAA amateurism in 2015. *Forbes*. Retrieved from https://www.forbes.com/sites/marcedelman/2015/02/28/dont-feed-or-shelter-the-athletes-the-absurdity-of-ncaa-amateurism-in-2015/#1c7527602529

Ensminger, M. E., Lamkin, R. P., & Jacobson, N. (1996). School leaving: A longitudinal perspective including neighborhood effects. *Child Development, 67*(5), 2400-2416.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications, Ltd.

Fields, S. K. (2012). Are we asking the right questions?: A response to the academic reforms research the NCAA. *Journal of Intercollegiate Sport, 5,* 60-64.

Firebaugh, G., & Acciai, F. (2016). For black in America, the gap in neighborhood poverty has declined faster than segregation. *Proceedings of the National Academy of Sciences, 113*(47), 13372-13377. doi: 10.1073/pnas.167220113

Flowerdew, R., & Goldstein, W. (1989). Geodemographics in practice: Developments in North America. *Environment and Planning A*, *21*(5), 605-616.

Furstenberg, Jr., F. F., & Hughes, M. E. (2000). The influence of neighborhoods on children's development: A theoretical perspective and a research agenda. In J. Brooks-Gunn, G. J. Duncan & J. L. Aber (Eds.), *Neighborhood poverty: Policy implications in studying neighborhoods, Vol. II* (pp. 23-47). New York: Russell Sage Foundation.

Gansemer-Topf, A. M., & Schuh, J. H. (2006). Institutional selectivity and institutional expenditures: Examining organizational factors that contribute to retention and graduation. *Research in Higher Education, 47*(6), 613-642. doi: 10.1007/s111-62-006-9009-4

Geiser, S., & Santelices, M. V. (2007, June 13). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes*. Retrieved from: http://www.cshe.berkeley.edu/publications/validity-high-school-grades-predicting-student-success-beyond-freshman-yearhigh-school

Geiser, S., & Studley, R. (2003). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment, 8*(1). 1-26.

Go, A. (2008, December 30). Athletes show huge gaps in SAT scores. *U.S. News & World Report*. Retrieved from https://www.usnews.com/education/blogs/paper-trail/2008/12/30/athletes-show-huge-gaps-in-sat-scores

Harding, D. J. (2003). Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology, 109*(3), 676-719.

Harding, D. J., Gennetian, L., Winship, C., Sanbonmatsu, L., & Kling, J. R. (2011). Unpacking neighborhood influences on educations outcomes: Setting the stage for future research. In G. J. Duncan & R. J. Murnane (Eds). *Whither Opportunity? Rising inequality, schools, and children's life chances* (277-298). New York: Russell Sage Foundation.

Harris, R., Grose, D., Longley, P., Singleton, A., & Brunsdon, C. (2010). Grid-enabling geographically weighted regression: A case study of participation in higher education in England. *Transactions in GIS, 14*(1), 43-61. doi: 10.1111/j.1467-9671.2009.01181.x

Harrison, W. (2012). NCAA Academic Performance Program (APP): Future directions. *Journal of Intercollegiate Sport, 5*, 65-82.

Hiltzik, M. (2015, March 20). Big-time college sports should split from their universities. *Los Angeles Times*. Retrieved from http://www.latimes.com/business/hiltzik/la-fi-hiltzik-20150320-column.html

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Huta, V. (2014). When to use hierarchical linear modeling. *The Quantitative Methods for Psychology*, *10*(1), 13-28. doi: 10.20982/tqmp.10.1.p013

Iceland, J., & Steinmetz, E. (2003). *The effects of using census block groups instead of census tracts when examining residential housing patterns*. Retrieved from https://www.census.gov/hhes/www/housing/resseg/pdf/unit_of_analysis.pdf

Jargowsky, P. (2016, October 14). Are minority neighborhoods a disaster? *The Century Foundation*. Retrieved from https://tcf.org/content/commentary/minority-neighborhoods-disaster/

Jencks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In L. E. Lynn & M. G. H. McGeary (Eds.), *Inner-city poverty in the United States* (111-186). Washington, D.C.: National Academy Press.

Johnson, G. (2014, August 22). Grant program aids limited-resource schools. *Champion Magazine*. Retrieved from http://www.ncaa.org/champion/grant-program-aids-limited-resource-schools

Kamens, D. H. (1971). The college "charter" and college size: Effects on occupational choice and college attrition. *Sociology of Education, 44*(3), 270-296.

Keilman, J., & Hopkins, J. S. (2015, March 26). College athletes routinely sign away rights to be paid for names, images. *Chicago Tribune*. Retrieved from http://www.chicagotribune.com/sports/college/ct-ncaa-waivers-met-20150326-story.html

Krogstad, J. M., & Fry, R. (2014, April 24). More Hispanics, blacks enrolling in college, but lag in bachelor's degrees. Retrieved from http://www.pewresearch.org/fact-tank/2014/04/24/more-hispanics-blacks-enrolling-in-college-but-lag-in-bachelors-degrees/

Kuh, G. D. (2009). What Student Affairs Professionals Need to Know About Student Engagement. *Journal of College Student Development* 50(6), 683-706.

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

Kuh, G. D., Kinzie, J., Buckley, J., Bridges, B., & Hayek, J. C. (2007). Piecing together the student success puzzle: Research, propositions, and recommendations. ASHE Higher Education Report, 32(5). San Francisco: Jossey-Bass.

LaForge, L., & Hodge, J. (2011). NCAA academic performance metrics: Implications for institutional policy and practice. *The Journal of Higher Education, 82*(2), 217-235.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126*(2), 309-337. doi: 10.1037//0033-2909.126-2-309

Levine, J., Etchison, S., & Oppenheimer, D. M. (2014). Pluralistic ignorance among student-athlete populations: A factor in academic underperformance. *Higher Education, 68*, 525-540. doi: 10.1007/s10734-014-9726-0

Long, M. C., Conger, D., & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal, 49*(2), 285-322. doi: 10.3102/0002831211431952

Lopez, M. H., & Gonzalez-Barrera, A. (2014, March 6). Women's college enrollment gains leave men behind. Retrieved from http://www.pewresearch.org/fact-tank/2014/03/06/womens-college-enrollment-gains-leave-men-behind/

Lynn, L. E., Jr., & McGeary, M. G. H. (Eds.). (1990). *Inner-city poverty in the United States*. Washington, D.C.: National Academy Press.

McArdle, J. J. (2008). Studies of impacts of minimum academic standards (Prop 48) on the academic achievements of college student-athletes. In P. C. Kyllonen, R. D. Roberts & L. Stankov (Eds.), *Extending intelligence: Enhancement and new constructs* (115-134). New York, NY: Taylor & Francis Group, LLC.

McArdle, J. J. & Hamagami, F. (1994). Logit and multilevel logit modeling of college graduation for 1984-1985 freshman student-athletes. *Journal of the American Statistical Association, 89*(427), 1107-1123.

McArdle, J. J., Paskus, T. S., & Boker, S. M. (2013). A multilevel multivariate analysis of academic performances in college based on NCAA student-athletes. *Multivariate Behavioral Research, 48*, 57-95. doi: 10.1080/00273171.2012.715836

Mayer, S. E., & Jencks, C. (1989). Growing up in poor neighborhoods. How much does it matter? *Science, 243*(4897), 1441-1445.

Merlo, J., et al. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health, 60*, 290-297. doi: 10.1136/jech.2004.029454

National Collegiate Athletic Association (2009, February). *Facilitating learning and achieving graduation: A report prepared for the NCAA Division I Academic Membership.* Retrieved from https://web1.ncaa.org/GRO/pages/pdf/report_facilitate.pdf

National Collegiate Athletic Association (2011, November). *Division I results from the NCAA GOALS study on the student-athlete experience* [PowerPoint slides]. Retrieved from http://www.ncaa.org/sites/default/files/DI_GOALS_FARA_final_0.pdf

National Collegiate Athletic Association (2016a). *NCAA sports sponsorship and participation rates report*. Retrieved from http://www.ncaapublications.com/productdownloads/PR1516.pdf

National Collegiate Athletic Association (2016b). *Results from the 2015 GOALS study of the student-athlete experience* [PowerPoint slides]. Retrieved from http://www.ncaa.org/sites/default/files/GOALS_convention_slidebank_jan2016_public.pdf

National Collegiate Athletic Association (2016c). *Trends in graduation success rates and federal graduation rates at NCAA Division I institutions* [PowerPoint slides]. Retrieved from http://www.ncaa.org/sites/default/files/2016RES_GSRandFedTrends-Final_sc_20161114.pdf

National Collegiate Athletic Association (2016d). *2016-17 Guide for the college-bound student-athlete*. Retrieved from http://www.ncaapublications.com/productdownloads/CBSA17.pdf

National Collegiate Athletic Association (2016e). *NCAA Division I Graduation Rates Report*. Retrieved from http://www.ncaa.org/sites/default/files/2016RES_GSR_report_20161114.pdf

National Collegiate Athletic Association. (2017a). NCAA graduation rates institution report. Retrieved from https://web3.ncaa.org/aprsearch/public_reports/instAggr2017/1_0.pdf

National Collegiate Athletic Association. (2017b). [Graph illustration the Average APR by Sport: Four-Year APR Plot October 5, 2017]. *National and Sport-Group APR Averages and Trends*. Retrieved from http://www.ncaa.org/about/resources/research/apr-averages-and-trends

National Collegiate Athletic Association. (2017c). NCAA Division I Committee on Academics operating manual. Retrieved from http://www.ncaa.org/sites/default/files/2017-18DICOA_PoliciesandProceduresManual_20171018.pdf

National Collegiate Athletic Association. (nd). *Division I Progress-Toward-Degree Requirements*. Retrieved from http://www.ncaa.org/about/division-i-progress-toward-degree-requirements

Nieuwenhuis, J., & Hooimeijer, P. (2016). The association between neighbourhoods and educational achievement, a systematic review and meta-analysis. *Journal of Housing and the Build Environment, 31*, 321-347. doi: 10.1007/s10901-015-9460-7

Nguyen, M. (2012, February). *Degreeless in debt: What happens to borrowers who drop out*. Retrieved from https://eric.ed.gov/?id=ED529758&gt

Oriard, M. (2012). NCAA academic reform: History, context and challenges. *Journal of Intercollegiate Sport, 5*, 4-18.

Osborne, J. W. (2000). Advantages of hierarchical linear modeling. *Practical Assessment, Research, and Evaluation, 7*(1).

Osterman, K. F. (2000). Students' need for belonging in the school community. *Review of Educational Research, 70*(3), 323-367.

Pascarella, E. T., & Terenzini, P. T. (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model. *The Journal of Higher Education, 51*(1), 60-75.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. San Francisco, CA: Jossey-Bass, Inc.

Paskus, T. S. (2012). A summary and commentary on the quantitative results of current NCAA academic reforms. *Journal of Intercollegiate Sport, 5*, 41-53.

Petr, T. A., & McArdle, J. J. (2012). Academic research and reform: A history of the empirical basis for NCAA academic policy. *Journal of Intercollegiate Sport, 5*, 27-40.

Petr, T. A., & Paskus, T. S. (2009). The collection and use of academic outcomes data by the NCAA. *New Directions for Institutional Research, 144*, 77-92. doi: 10.1002/ir.315

Pike, G. R., & Saupe, J. L. (2002). Does high school matter? An analysis of three methods of predicting first-year grades. *Research in Higher Education, 43*(2), 187-207.

Quillian, L. (2003). How long are exposures to poor neighborhoods? The long-term dynamics of entry and exit from poor neighborhoods. *Population Research and Policy Review, 22*, 221-249.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, second edition*. Newbury Park, CA: Sage.

Rubin, L. M., & Moses, R. A. (2017). Athletic subculture within student-athlete academic centers. *Sociology of Sport, 34*(4), 317-328. https://doi.org/10.1123/ssj.2016-0138

Ryan, J. F. (2004). The relationship between institutional expenditures and degree attainment at baccalaureate colleges. *Research in Higher Education, 45*(2), 97-114.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing “neighborhood effects”: Social processes and new directions in research. *Annual Review of Sociology, 28*, 443-78. doi: 10.1146/annurev.soc.28.110601.141114

Sharkey, P., & Faber, J. W. (2014). Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects. *Annual Review of Sociology, 40*, 559-579. doi: 10.1146/annurev-soc-071913-043350

Shulman, J. L., & Bowen, W. G. (2000). *The game of life: College sports and educational values*. Princeton, NJ: Princeton University Press.

Singleton, A. D., & Spielman, S. E. (2014). The past, present, and future of geodemographic research in the United States and United Kingdom. *The Professional Geographer, 66*(4), 558-567. doi: 10.1080/00330124.2013.8487642345

Singleton, A.D., Wilson, A.G., & O’Brien, O. (2012). Geodemographics and spatial interaction: An integrated model for higher education. *Journal of Geographical Systems, 14*, 223-241. doi: 10.1007/s10109—010-0141-5

Steele, F. (2008). Module 5: Introduction to multilevel modeling. *Learning Environment for Multilevel Methodology and Applications*. Retrieved from https://www.cmm.bris.ac.uk/lemma/course/view.php?id=13#_ga=2.237109952.1549267532.1522361971-1819344592.1520892156

Stinebrickner, T., & Stinebrickner, R. (2013). *Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model*. Retrieved from http://www.nber.org/papers/w18945

Taylor, J. (2012). Comparing the acceptance rates of athletes to the general public. *The Husky Haul*. Retrieved from http://thehuskyhaul.com/2012/05/16/comparing-the-acceptance-rates-of-athletes-to-the-general-public/

Tinto, V. (1975) Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89-125.

Titus, M. A. (2004). An examination of the influence of institutional context on student persistence at 4-year college and universities: A multilevel approach. *Research in Higher Education, 45*(7), 673-699.

Tracy, M. (2017, September 26). North Carolina looking to slip through holes in NCAA rules. *The New York Times.* Retrieved from https://www.nytimes.com/2017/09/26/sports/north-carolina-ncaa-academic-scandal.html

Umbach, P. D., Palmer, M. M., Kuh, G. D., & Hannah, S. J. (2006). Intercollegiate athletes and effective educational practices: Winning combination or losing effort?. *Research in Higher Education*, *47*(6), 709-733.

U.S. Department of Commerce. (1994). *Geographic areas reference manual*. Prepared by the U.S. Census Bureau. Retrieved from https://www.census.gov/geo/reference/garm.html

U.S. Department of Commerce. (2001a). *Summary file 1. 2000 census of population and housing: Technical documentation*. Prepared by the U.S. Census Bureau. Retrieved from https://www.census.gov/prod/cen2000/doc/sf1.pdf

U. S. Department of Commerce. (2001b). *Summary file 3. 2000 census of population and housing: Technical documentation*. Prepared by the U.S. Census Bureau. Retrieved from https://www.census.gov/prod/cen2000/doc/sf3.pdf

U.S. Department of Commerce. (2016). *Percentage of recent high school completers enrolled in in 2- and 4-year colleges, by race/ethnicity: 1960 through 2015* [Data file]. Retrieved from https://nces.ed.gov/programs/digest/d16/tables/dt16_302.20.asp

U.S. Department of Education. (2007). *Characteristics of minority-serving institutions and minority undergraduates enrolled in these institutions: Postsecondary education descriptive analysis report*. Retrieved from https://nces.ed.gov/pubs2008/2008156.pdf

U.S. Department of Education, National Center for Educations Statistics (2012). *Profile of 2007-08 first-time bachelor degree recipients in 2009*. Retrieved from https://nces.ed.gov/pubs2013/2013150.pdf

U.S. Department of Education, National Center for Education Statistics. (2014). [Table illustration Integrated Postsecondary Education Data System (IPEDS), Fall 2004, Fall 2013, and Fall 2014, Completions component]. *Digest of Education Statistics.* Retrieved from https://nces.ed.gov/programs/raceindicators/indicator_ree.asp

U.S. Department of Education, National Center for Education Statistics. (2016a). [Graph illustration Integrated Postsecondary Education Data System (IPEDS), Spring 2016, Fall Enrollment component; and Fall 2014, Institutional Characteristics component]. *Digest of Education Statistics.* Retrieved from https://nces.ed.gov/programs/coe/indicator_ctr.asp

U.S. Department of Education, National Center for Education Statistics. (2016b). [Graph illustration Integrated Postsecondary Education Data System (IPEDS), Spring 2001 through Spring 2016, Fall Enrollment component; and Enrollment in Degree-Granting Institutions Projection Model, 1980 through 2026]. *Digest of Education Statistics*. Retrieved from https://nces.ed.gov/programs/coe/indicator_cha.asp

U.S. Department of Education, National Center for Education Statistics. (2017). *The Condition of Education 2016* (NCES 2017-144). Retrieved from https://nces.ed.gov/pubs2017/2017144.pdf

U.S. Department of Housing and Urban Development. (2010). *HUD-USPS ZIP crosswalk files* [Data file]. Retrieved from https://www.huduser.gov/portal/datasets/usps_crosswalk.html

Vartanian, T. P. & Gleason, P. M. (1999). Do neighborhood conditions affect high school dropout and college graduation rates? *Journal of Socio-Economics, 28*(1), 21-41.

Venezia, A., Callan, P. M., Finney, J. E., Kirst, M. W., & Usdan, M. D. (2005, September). *The governance divide: A report on a four-state study on improving college readiness and success.* Retrieved from http://files.eric.ed.gov/fulltext/ED508097.pdf

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago: The University of Chicago Press.

Wilson, M., & Daly, M. (1997). Life expectancy, economic inequality, homicide, and reproductive timing in Chicago neighborhoods. *British Medical Journal, 314*(7089), 2171-1274.

Wodtke, G. T., Harding, D. J., & Elwert, F. (2011). Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. *American Sociological Review, 76*(5), 713-736. doi: 10.1177/0003122411420816

Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology, 8*(1), 52-69. doi: 10.20982/tqmp.08.1.p052

VITA

Ann Kearns Davoren earned her baccalaureate degree from Saint Mary's College, Notre Dame, IN in 1998 and a Master of Public Administration with concentrations in policy analysis and comparative international affairs from Indiana University, Bloomington in 2001. She was admitted into the doctoral program in research methodology at Loyola University, Chicago for the Fall 2011 term. Dr. Davoren has been employed with the research department of the National Collegiate Athletic Association since October of 2004. In this capacity, her work has focused on student-athlete survey research and academic, governance, and health and safety research across all three membership divisions. She also has been an active independent research and statistical consultant with organizations such as the National Catholic Educational Association, the Andrew M. Greeley Center for Catholic Education, Pass the Torch for Women, and several collegiate athletic conferences and associations. Her academic interests are in research design, quasi-experimental design, survey research, higher education research, student outcomes, multilevel modeling, and structural equation modeling.