GEOGRAPHY AND COLLEGE CHOICE: FOUNDATIONAL ANALYSES TOWARDS A GEOSPATIAL FRAMEWORK

Nestor Alexis Ramirez

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Education (Policy, Leadership, and School Improvement).

Chapel Hill
2020

Approved by:

Eric A. Houck

Thurston Domina

Cynthia P. Demetriou

Paul D. Umbach

Robert Martinez

ABSTRACT

Nestor Alexis Ramirez: Geography and college choice: Foundational analysis towards a geospatial framework
(Under the direction of Eric Houck)

College access and choice are topics that have been studied extensively for several decades; scholars have employed multiple theoretical and conceptual frameworks to inform their research on access and choice, with the traditional models framing college choice as a multistage decision-making process. Geography, and the importance of geographic characteristics in the formation of college choice sets as well as the enrollment decision itself, has been overlooked in much of the literature. Heeding the call to incorporate geospatial considerations in college choice research (Hillman, 2016; Turley, 2009), this three-article dissertation develops an empirical foundation through which a new conceptual framework can be created—a framework that places the onus of college decision-making on geographic proximity and a willingness to be mobile to attend college. Through a combination of quantitative methods and secondary data analysis, this dissertation builds upon extant literature on geography and college choice and examines the role that geography and place through multiple angles, including student preferences to remain close to home, exploration of student mobility patterns, and geographic characteristics of college choice sets. Analyses from the first study found that having stronger preferences to remain close to home for college was not associated with college enrollment but was associated with how far from home they ultimately traveled to attend college. The second study found that traditional and nontraditional first-time beginning students, defined by age when starting college, recency of high school graduation, and institutional control of college attended, differed significantly from one another in the distance between home and college. The final study explored college choice sets, and identified significant variables that predicted the distance between students and the colleges they chose to include in their choice sets.

To my beautiful wife Jamie, who pushes me every day to be as good as I can be, and my family, who stressed education above all else.

ACKNOWLEDGEMENTS

Despite having written over 180 pages for this dissertation (as I am writing this at the end of the project), I never knew how difficult acknowledgements would be for me to write. There were many times I thought I'd give up, or not have the wherewithal to see such a daunting task through. I truly believe that the only reason I managed to drag myself across the finish line was thanks to my amazing support network. My family—including my wonderful wife, whom I met while in graduate school, and her incredible family that has welcomed me with open arms—has seen me through the highs and the lows, and were always willing to excuse a lack of attention or time spent with them. Same goes for my friends, who I've had to neglect during this entire endeavor, and are likely anticipating how I'm going to pay them back for all of this lost time (and probably still don't get why I put myself through this all). Lastly, I'd give a special shoutout to all of the professors, mentors, administrators, and colleagues who offered advice throughout these last six years, support when I wavered, or challenging my thinking. I can only hope that with this dissertation, I can begin a career of impactful and meaningful work in education that shows all of my family, friends, professors, and mentors that their investment of time and energy in me was worth it in the end. Thank you for everything, and I love you with all my heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BPS	Beginning Postsecondary Students Longitudinal Study
BPS:12/17	2012/17 Beginning Postsecondary Students Longitudinal Study
BY	Base year
CCD	Common Core of Data
Coeff.	Coefficient
CRS	Coordinate Reference System
CZ	Commuting zone
F1	First follow-up
HSLS	High School Longitudinal Study
HSLS:09	High School Longitudinal Study of 2009
HSLS:09/12	High School Longitudinal Study of 2009/12
IPEDS	Integrated Postsecondary Education Data System
M	Mean
MLM	Multilevel modeling
NCES	National Center for Education Statistics
OLR	Ordered logistic regression
OLS	Ordinary least squares regression
OR	Odds ratio
PPO	Partial proportional odds modeling
PSS	Private School Survey
Ref.	Reference group
SD	Standard deviation
SES	Socioeconomic status
U13	2013 update

LIST OF SYMBOLS

χ	Chi-statistic
γ	Multilevel model coefficient
†	Not applicable
ε	Observation error
β	Regression coefficient
#	Rounds to zero

CHAPTER 1: INTRODUCTION

Dissertation Overview

Deciding where to attend college—and whether to attend at all—involves complex negotiations for students and their families. In order to better examine these negotiations, the literature on college access and choice has traditionally conceptualized the college choice process using frameworks and models, with the most common being a multistage process comprised of three steps: predispositions, search, and choice (Cabrera & La Nasa, 2000; Hossler & Gallagher, 1987; Nora, 2004). These models are incorporated into research to identify the factors involved during the choice process as well as their strength in influencing the college decisions that students and their families make. College choice frameworks are particularly useful for studying the connections between a predictor and a college access or choice decision that can be affected by policy. This body of research include policy-relevant associations, including the influence of economic capital and financial aid on access (Bailey & Dynarski, 2011; Dynarski & Scott-Clayton, 2013; Paulsen & St. John, 2002), the impact of high school context on access and increasing access through college-going cultures (Engberg & Wolniak, 2010; Roderick, Coca, & Nagaoka, 2011; Sherwin, 2012), using behavioral nudging and simplifying complex logistical work to ease the transition process (Bettinger, Long, Oreopoulos, & Sanbonmatsu, 2012; Castleman & Page, 2015), and the role of social capital as information sources (Ceja, 2006; Perez & McDonough, 2008; Sherwin, 2012). Many of these models acknowledge students and families contend with geographic characteristics as well, such as the distance to college options and rurality of both the home community and where the colleges are located; however, geographic characteristics have not been the focus of much of the college choice literature despite several scholars who have argued that they are extremely important—if not among the most important—factors in constraining or encouraging postsecondary access (Hillman, 2016; Hillman & Weichman, 2016; Turley, 2009).

There are several avenues available to scholars to study geographic influencers on college choice. Some scholars approach geography and access at the student level; students' proximity to postsecondary options, for instance, has been found to correlate with a likelihood of attending a postsecondary institution (Frenette, 2004; Turley, 2009). As the distance between the student's home and an institution increases, the student is more likely to attend closer options when controlling for family income, though low-income students are more likely to be affected by proximity (Frenette, 2004). Cultural characteristics can influence how important the role of geography plays for students' college choices. Students and their families hold weaker or stronger preferences for living close to home during college, which is constructed based on a variety of socioeconomic and academic characteristics (Ovink & Kalogrides, 2015). The preference to remain close to home is associated with postsecondary enrollment, where stronger preferences decrease the likelihood of college enrollment; the evidence suggests this association exists both from the students' preferences to stay home (Desmond & Turley, 2009) and the parents' preferences for their children to stay home (Ovink & Kalogrides, 2015; Turley, 2006). Patterns of student mobility, and the factors associated with mobility have also received attention by researchers (Alm & Winters, 2009; Mixon, 1992; Mixon & Hsing, 1994; Niu, 2015)

Beyond student-level characteristics, scholars have examined geographic factors at the institution- and state-level. González Canché (2017, 2018) has published extensively on macro-level topics such as the variations in institutional price setting explained by the proximity between institutions, student migration patterns across states along with gravity models to explain the relative "push" and "pull" states and institutions have on interstate migrants, and on methodological advances in geospatial research in higher education. Hillman (2016, 2017) has published on the availability of college options for different communities, finding that low income, rural, Hispanic-majority and Black-majority communities often have fewer available postsecondary options than other communities. Institutional amenities and institutional quality and their interactions with distance have also been examined in relation to their impact on college choices (Sá, Florax, & Rietveld, 2004).

Despite the growing body of geospatial and geographic research in higher education, researchers have yet to fully merge the extant literature on college access and choice with the important and relevant research on the role of geography in higher education. In traditional models of college choice, geography and mobility have been treated as footnotes within larger frameworks involving human, social, cultural, and economic capital—usually by including a measure of rurality or regionality, if at all. However, like many of the scholars cited above, I argue that the role of geography and student mobility is much more impactful to college decision-making than previously considered.

In this three-paper dissertation, I add to the existing body of research by laying the foundation for a geospatial conceptual framework of college choice, wherein the decision to attend college is framed as one shaped by geographic context. This framework will approach college decision-making as a process heavily constrained or expanded by a willingness to relocate or travel farther distances for college. Moreover, policy contexts and institutional characteristics will make certain colleges more or less appealing for students. I will begin by describing the multidisciplinary conceptual frameworks that are most common in the college choice literature. The overview will inform the impetus of the three articles by outlining where the conceptual frameworks work well and how geography remains underexplored when applying them in research. Then, I will conclude with a description of the three dissertation studies and how they contribute to the establishment of the geospatial conceptual framework.

Multidisciplinary Conceptual Frameworks in College Choice Research

Instead of applying individual theories, it is common for higher education scholars to utilize conceptual frameworks and models that incorporate multiple theories or theoretical concepts discussed here in studies of college choice. These conceptual frameworks borrow from sociology, economics, psychology, and other fields. The three-stage model of college choice in particular (Cabrera & La Nasa, 2000; Hossler, Braxton, & Coopersmith, 1989; Hossler & Gallagher, 1987; Nora, 2004) remains salient in the literature today, providing a foundation through which to understand the components that influence access to higher education. Using these conceptual frameworks allow scholars to focus on a holistic set of factors that affect access, choice, and match where theoretical frameworks may only focus on a subset.

Multistage Models of College Choice

Among the most commonly used conceptual frameworks in college choice research involve the three-stage or multi-stage models. The three-stage model of college choice was first developed by Hossler and Gallagher (1987) and expanded in Hossler, Braxton, and Coopersmith (1989). Cabrera and La Nasa (2000) offer a great review of the three-stage model with insights from the literature about each component of the model. The three stages of the college choice process are *predispositions*, *search*, and *choice*. In the predispositions stage, children develop educational aspirations and expectations and the academic outcomes necessary for college enrollment, driven by factors like parental encouragement, academic resources, demonstrated academic ability, SES, and others (Cabrera & La Nasa, 2000). Once students approach the end of high school, they begin the search phase, wherein they begin collecting and narrowing a tentative list of prospective colleges. The search phase largely relies on social and cultural capital—need access to information about individual colleges, understanding of the financial aid and college application process, and informed preferences among institutions and potential choices of majors all inform the list of prospective colleges (Cabrera & La Nasa, 2000). Lastly, students enter the choice phase, where they ultimately decide to enroll or not enroll in a college.

Litten (1982) posited a similar multistage process to describe college choice, electing to separate the process into five stages. According to Litten, the process begins with college aspirations—aspiring to attend college is affected by the student's background (demographics, socioeconomic status, etc.), environmental factors, academic performance and school characteristics, and psychological/noncognitive factors. Once the aspirations are established, eligibility of aid and policy context influences whether or not the college choice process begins in earnest. Should the student proceed with the college choice process, they engage in information gathering. During this third step, the student receives information from colleges and information channels to craft a list of institutions to which they wish to apply. Step four involves applying to those colleges based on institutional characteristics that are preferable or not preferable—then, after the colleges provide an admissions decision to the student as well as a financial aid package, the student decides where or if to enroll.

The primary benefit that these multistage frameworks is the explicit delineation of the decision-making process. By describing the choice process temporally—that is, as a linear process over time—these frameworks can serve as an asset for advocates, institutions, and policymakers looking to increase college enrollments efficiently. For example, an institution seeking to draw more students to their campus would likely not do so during the predispositions/aspirations stage, as it may not be cost effective to reach these students at this point in the process given the numerous other factors that influence them at this point; further, at least where Litten (1982) is concerned, the student has yet to make the decision to even begin the process of college choice. Instead, an institution would target students who are actively gathering information—the search stage (Hossler & Gallagher, 1987) or information gathering stage (Litten, 1982)—to influence their application behaviors.

Perna's Multilayered Choice Framework

The multilayered choice framework—also known as the integrated model of college choice (Salisbury, Umbach, Paulsen, & Pascarella, 2009)—attempts to take a multidisciplinary approach to the study of college choice. Citing Hossler, Braxton, and Coopersmith (1989) and Paulsen (1990), Perna argued that college choice research relies on both economic and sociological theories, and thus created a conceptual framework that merges the two into a cohesive decision-making process centered on human capital investment (Perna, 2006). In this framework, the college choice is conceptualized as one in which individuals compare the perceived costs and benefits, as described by human capital theory. However, the framework extends the human capital theory by situating this calculus within four contextual layers—habitus, school and community context, higher education context, and social, economic, and policy context (Perna, 2006). These layers influence the internal calculus by changing how each individual perceives their expected costs and expected benefits of higher education, and in turn influencing their decision to attend college.

The first layer, habitus, reflects the immediate, individual context of the prospective student (Perna, 2006). This includes demographic considerations like gender and race, social capital, dispositions, beliefs, attitudes, and other cultural capital. As mentioned previously, there is an extensive literature on the importance of habitus and individual characteristics on college choice (e.g. Byun et al., 2012; Paulsen & St.

John, 2002; Perna & Titus, 2005). This innermost layer of the framework is nested within layer two, school and community context, which refers to what Perna cites as "organizational habitus (Perna, 2006)." The quantity and quality of school and community resources as well as the availability of support structures are situated within this layer and can either help or harm the college choice process (Perna, 2006).

The third layer consists of the higher education market context—institutional characteristics like distance and program offerings, competition between institutions, admissions criteria, and other college-level factors fall within this layer (Perna, 2006). This context is important to the decision-making process in several different ways. For example, lower geographic distances from home to a college (which will be covered later) will often make students consider that college more strongly when making choice decisions. The final layer is social, economic, and policy context (Perna, 2006). All of the prior layers are affected by the condition of policy when the students engage in the college decision-making process. National level policy like the maximum amount of Pell Grant aid may affect whether some colleges are more or less accessible for low-income students, for example. Furthermore, state and local level college promise programs that leverage different funding sources to provide college tuition-free to students may encourage access for many students that would otherwise not consider higher education (Pluhta & Penny, 2013).

College Match

The literature on academic undermatch has taken a different approach on college choice by focusing on the compatibility between an individual's academic credentials and the selectivity of institutions they choose to attend. Academic undermatch frameworks operate under the assumption—supported by research—that students who enroll in more selective institutions are more likely to attain a postsecondary credential, and thus students should enroll in an institution that is as selective as their academic credential allows (Smith et al., 2013). Students who do not make that optimal choice, whether because of a lack of information, aspirations, or finances, are considered "undermatched."

Low-income students undermatch more often than middle- or high-income students (Roderick et al., 2011; Smith et al., 2013). Roderick et al. (2011) argue that high schools shape match and undermatch through college-going cultures and by providing sources of college knowledge—both conceptualizations of social

capital—that are either sufficient or insufficient for different students. Per their review of the literature, underserved students rely more on "traditional feeder patterns" (p. 181) instead of considering other colleges that may better meet their needs and boost their college outcomes. However, there has been pushback to this perspective stemming from skepticism of how selectivity is defined by researchers, and assumptions that matching based on academic measures will reduce inequality in access for low-income students, despite the body of work (some of which has been cited herein) examining the role of academic measures like standardized test scores in driving stratification and marginalization for higher education institutions (Bastedo & Flaster, 2014).

Other Frameworks

Of course, there are many other conceptual frameworks available to researchers that I have not discussed in this literature review. Some scholars develop conceptual frameworks tailored for studying nontraditional college students. For example, Lansing (2017) developed a conceptual framework for studying college choice specific to distance learners—this framework truncates the three-step choice model and focuses on the search and choice steps. Lansing's study wherein they develop the distance learner framework indicates that distance learners care about similar choice factors like availability of program and quality of education, while other components like location of the college have less of an impact on the choices made. The financial nexus model ties together college choice and college persistence and examines how early decisions related to college choice can affect persistence rates later, and how financial perceptions and experiences are tied to both (Paulsen & St. John, 2002). The researchers behind the financial nexus model suggest through applications of the model that financial perceptions and financial resources affect persistence differently across social class, where poorer students experience higher likelihoods of having financial perceptions serving as a barrier for access, and fewer financial resources hindering persistence (Paulsen & St. John, 2002; St. John, 2001; St. John et al., 2001). Frameworks that focus on college outcomes other than enrollment choice can also be adapted to college choice studies; for example, validation theory—which stipulates that validating experiences are particularly important for nontraditional students to learn and thrive in college settings that can often invalidate their cultural diversity—can be adapted from a persistence and

attainment model of student development to one of college access by shifting the context from validating college experiences to validating K-12 ones (Rendon, 1994).

Commonalities across Access and Choice Frameworks

This review of theory and conceptual frameworks is far from exhaustive; however, the literature presented can help us understand some commonalities across lenses and examine where gaps in traditional thinking may lie. There are three commonalities across many of these access and choice frameworks that I wish to highlight: establishing college enrollment as a decision-making process with multiple discernable stages; the importance of multiple contexts on the decision-making process; and the integration of multiple theories and disciplines to explain decision-making.

As described here, most of the more popular theoretical frameworks describe college access and choice as one that extends beyond the decision to enroll in a college or not. Prior to the enrollment decision, there are other stages where students and families develop a knowledge base about postsecondary education, predispositions for where and whether to attend college, leveraging multiple sources of capital to determine whether college is affordable, and a selection of a choice set. Both the three-stage model of college choice and the five-stage model span as far back as when preferences and predispositions first start developing—while Cabrera and La Nasa (2000) suggest that predispositions and early college planning begin in eighth grade, the timing likely varies across different subgroups including race (Litten, 1982). Thus, these frameworks incorporate pre-college characteristics into the decision-making process. Of course, the role of child development in postsecondary outcomes is not a scandalous claim by any means, but if a researcher took a purely theoretical human capital theory approach to college choice, it is more likely early pre-college characteristics will be neglected over the costs and benefits components of the decision-making process. In other words, the conceptual frameworks extend the process beyond the final choice decision to incorporate years of child development and educational development in order to examine how these pre-college forces—academic, demographic, financial, cultural, etc.—play a role in that ultimate decision and, more specifically, how these forces are more influential for some than others.

While some frameworks—specifically the multilayered choice framework (Perna, 2006)—place a greater emphasis on context than others, they all incorporate familial, communal, school, and policy contexts in some way. Litten's (1982) college choice model incorporates context via school characteristics, family characteristics, economic conditions, and cultural conditions affecting college aspirations. Hossler and Gallagher's original three-stage choice framework places little attention to context explicitly but incorporates individual and familial contexts often throughout the framework (Hossler & Gallagher, 1987). Social class, economic capital (in the form of available financial resources), cultural capital (in the form of perceptions of affordability and educational values), and financial aid policy (in the form of amount of aid students qualify for) all play a role in the framework's predispositions, search, and choice stages (Hossler & Gallagher, 1987). Later work on Hossler and Gallagher's three-stage model highlights rurality and urbanicity—defined as family residence location—as an important contextual factor (Hossler et al., 1989), as well as the dispersion of college information within familial and school contexts (Cabrera & La Nasa, 2000). College match frameworks place extra emphasis on school and community context in particular through their work on developing college-going cultures and how they affect academic undermatch (Roderick et al., 2011; Smith et al., 2013). And, of course, Perna's (2006) model places the onus of the decision-making process on individual, school, community, and policy contexts to describe how these factors affect the perceptions of benefits and costs in the final choice.

Ultimately, what the frameworks suggest is that there are several contexts that affect individual decision-making related to college choice that makes the final choice incredibly nuanced and variant across individuals. However, it also avoids suggesting that context is the *only* part that matters; in this way, frameworks do not approach college access, choice, and match in the same way that SRT might. Context matters, but only as far as it affects the individual's formulation of which college to consider more than others—thus, the frameworks incorporate context without removing all of the individual agency in the decision-making process.

Lastly, as mentioned in previous sections, college access and choice frameworks integrate multiple disciplines within their formulation of the decision-making process. They are best suited for scholars who

wish to examine access and choice through multiple concurrent lenses. One of the best examples of this is in the way Perna justifies the utility of a framework that melds sociology and economics:

> When considered separately, neither rational human capital investment models nor sociological approaches are sufficient for understanding differences across groups in student college choice… By reflecting differences in expectations, preferences, tastes, and certainty about higher education investment decisions, measures of social and cultural capital appear to be particularly important for understanding differences across groups in college enrollment decisions that are not explained by human capital investment models. (Perna, 2006, pp. 115–116)

In other words, sociological concepts like cultural and social capital, alongside human capital and rational decision-making ideas from economics, are vital for studying diverse students and student subgroups. Thus, the frameworks that combine these factors are better suited for explaining choice decisions and the decision-making process for our current population of college-bound students than any one theory alone.

Geography and its Reduced Role in College Choice Frameworks

Now, we turn to the role of geography in both current frameworks and college choice as a whole. The attention paid to geographic considerations and its perceived importance can best be summarized by Litten, who said "for the most part, geographic considerations are of little consequence (Litten, 1982, p. 395)." The theories and models discussed herein are not adequately formulated to explore critical issues related to geography for underserved student populations. While some geographic considerations are taken into account in the literature through these approaches, such as rurality (Byun et al., 2012) and urban-suburban differences (Roderick et al., 2011), the theories and frameworks are less able to explain how geographic opportunity structures shape college access, transition, and match. These theories operate under the assumption that students are highly mobile: for example, proponents of the mismatch hypothesis and human capital theory might suggest that, should students have perfect access to information about the benefits of attending elite universities and labor market outcomes, they would be willing to attend the most selective or most economically beneficial institution they are able to get accepted to, regardless of where it is located. This assumption has little basis in reality.

For years, scholars have been calling for an increased focus on student mobility and geographic proximity to college options (Hillman, 2016; Paulsen & St. John, 2002; Turley, 2009). It can be argued that

the traditional models and theories do not adequately focus on the "geography of opportunity" (Hillman, 2016, p. 990). Without considerable focus on student mobility and the geography of opportunity through traditional theories and frameworks, the studies that utilize them will not adequately describe college choice patterns and inequities in access, choice, transition, and match. As such, there is a considerable need for incorporating geography in a more salient way than researchers have until now; instead of looking at geography merely as urbanicity or rural-urban differences, researchers should integrate geography in more creative and relevant ways. For example, the impact that geographic location has on developing college choice sets may differ across students based on their preference towards whether they are willing to travel farther from home for college. And while some scholars have studied how the number of and distance from postsecondary institutions can affect college enrollment behaviors, the literature has not gone as far as establishing whether proximity to college options play a primary role in developing college choice sets across student groups.

Contents and Structure of the Dissertation

The three papers work together to build evidence for the framework by approaching the college choice decision through three research angles: a study of the preference to stay home after high school, descriptions of the individual and contextual characteristics that are associated with mobility for students attending college for the first time, and an exercise in predictive modeling of student mobility operationalized via the college choice set for each student. These three research agendas complement each other by explicitly searching for associations and meanings discussed in prior literature to understand how a willingness or ability to leave the home community—or an unwillingness or inability to do so—impacts which schools prospective students consider in their college decisions. Furthermore, the studies allow for insights into characteristics beyond the student-level factors which may play a part in how mobile or nonmobile students are willing to be—financial aid availability, proximity of college options, postsecondary institutional characteristics—and how these characteristics may help us understand whether a willingness to leave the home for college influences the choice sets prospective students consider.

The following three chapters are independent articles meant to be self-contained and of publishable quality. Each chapter, therefore, has its own introduction, literature review, methods, data, and analysis. Though independent, the studies are meant to inform the development of a geospatial conceptual framework that is driven both by these findings and the extant literature. The framework will be presented at the conclusion of this dissertation as a component of the synthesis of all three studies.

The first paper examines the connection between a stated desire to remain home for college, college access related characteristics, enrollment in college and the distance of the enrolled college. The paper is driven by three guiding questions: 1) what factors and characteristics are associated with a preference to stay near home for college? 2) is the preference to remain close to home associated with college enrollment net of other characteristics? 3) And if so, is it also associated with the distance of the enrolled college?

The connection between the preference to remain home, as stated previously, has been explored in the literature before; this paper adds to our understanding of its connection on college access in several ways. First, it uses a newer data set, the High School Longitudinal Study of 2009 and its follow-ups (HSLS:09; HSLS:09/12, U13). HSLS:09 is a nationally representative study of 9th graders that follows them through high school and beyond, with particular focus on their pathways from high school to postsecondary education (Ingels et al., 2015). One of the benefits of this data set, apart from giving a nationally representative sample of high schoolers, is that it asks students in the first follow-up whether staying close to home is important to them when considering their college options. This item is well suited for exploratory analysis to pinpoint the student characteristics associated with this preference.

Second, this study pushes our knowledge forward by using complex sampling analysis and more accurate definitions of distance to see whether the preference to stay home is associated both with enrollment more generally, and with the distance of the chosen colleges. HSLS:09/12 and U13 provide sampling weights and replicate weights that allow for accurate analysis of the population of interest without requiring exceedingly large sample sizes. Studies that incorporate complex sampling analysis often neglect to account for components of sampling design that can compromise the generalizability and validity of their findings (West, Sakshaug, & Aurelien, 2016). Furthermore, this study (and the others in this dissertation) uses

geographic information systems (GIS) software and open-source maps to generate distances between home and college. By beginning with an examination of student preferences for mobility, the dissertation establishes a foundation to determine how important mobility is when deciding what college to attend. It is hypothesized that students who express that staying close to home when choosing a college will choose—or consider—closer colleges. Those who do not think it is important to stay home, or express an interest in being farther from home, are hypothesized to consider farther colleges on average.

The second paper will continue to expand the framework foundation by using a dataset of beginning college students and finding student-level, institutional-level, and geographic characteristics associated with students along the mobility spectrum. Where the first paper is concerned with the desire and preference to remain home for college (prior to attending and deciding on a college), the second paper will use a nationally representative sample of postsecondary education participants who have chosen colleges both near and far from where they originate. This study is also focused on two areas of inquiry: 1) what individual, institutional, and contextual factors are associated with the distance between a student's enrolled college and their home community, and 2) are there "meaningful" categories of mobility researchers can derive from the data?

Paper two contributes to the literature first by setting the scope of college choice beyond the traditional college-aged student. The data set used for study two, the Beginning Postsecondary Students (BPS:12/17) study, is a representative sample of all beginning college students—this includes adults and older students beginning college for the first time. Nontraditional students remain understudied and overlooked in traditional models of college choice (Cabrera & La Nasa, 2000; Schuetze & Slowey, 2002). By using a representative data set in this analysis, the examination of student mobility will reflect the larger population of college students rather than primarily traditional-aged students. Further studies can be derived from this larger, more representative framework to explore subpopulations—including traditional students, but also others—while establishing a cohesive and accurate understanding of the decision-making process as it relates to mobility.

Based on prior research and literature, the second paper extends our current knowledge by identifying predictors that can be tested in a model of college choice and student mobility. This paper will

seek to further understand the spectrum of student mobility as it relates to beginning college students nationwide: the distance they travel from home to school and the various factors and characteristics associated with that distance. The goal of research question two is to codify and describe the mobility spectrum based on these associations; once established, these predictors will be used to form a testable, empirical model of student mobility that will predict the range of distance students are willing to consider when engaging in the college choice process. The model will be tested and explored further in paper three.

Paper three completes the groundwork for the geospatial framework and fully formulates the proposed model for examining student mobility. Drawing from the findings of papers one and two, the final component of the dissertation develops and tests a predictive model of student mobility by drawing from identified predictors collected through literature reviews and prior results and examining their associations with the distances of schools in the choice set. This process relies on self-reported "choice" schools by students in HSLS:09/12 and U13, which are used to calculate an average distance of institutions the students considered for college. The average distance of self-reported college options, then, serves as a proxy for the distance the students were willing to travel for college. Using ordinary least squares regression (OLS) and multilevel modeling (MLM), the proposed predictors are analyzed alongside the average distance of college options, thereby providing evidence for the viability of the conceptual framework as they relate to student mobility and college choice. In short, paper three asks two questions: 1) is the distance of colleges in students' choice sets predicted by background, academic, familial, and other characteristics? and 2) what institutional factors can predict the distance between students and the colleges in their choice set? The variables utilized in this study are to be included in the prospective conceptual framework.

Like paper one, paper three utilizes HSLS:09/12 and U13 to test the predictive model. This data set is useful for this set of analyses due to the inclusion of up to three postsecondary institutions for respondents to identify as options they considered for college irrespective of cost. These institutions are then matched with institutional data from the Integrated Postsecondary Education Data System (IPEDS), which allows for calculating distances from the home community to the schools in each choice set. However, HSLS suffers in several respects: 1) the sample is traditional-aged students only, as the base year consists of ninth graders in

2009; 2) not all of the variables from BPS in paper two may exist in HSLS, as they are two different survey instruments. Despite these limitations, these analyses can provide more utility in the study of college choice sets and predicted student mobility.

In all of these papers, distance is covered and discussed extensively. In every paper, when discussing the distance between two points (for example, high school to college or permanent household to college), the text is referring to the geodesic distance between those two points. Geodesic distance refers to the shortest line between two points "as the crow flies." While it is possible to calculate distance using methods more accurate to road travel—for example, via road networks—these studies instead calculate a straight line using ellipsoids in QGIS and determine the distance in miles of that line. This form of distance was chosen because of the computer resources necessary to run calculations via road maps or other types of distance beyond shortest line (for example, time to travel between two points). While losing analytic utility by having a distance measure that is less realistic to the lived experiences of students, the sheer amount of computer resources needed to run distance calculations on a national scale was simply unrealistic given the time constraints of this dissertation.

REFERENCES

Alm, J., & Winters, J. V. (2009). Distance and intrastate college student migration. *Economics of Education Review*, *28*(6), 728–738. https://doi.org/10.1016/j.econedurev.2009.06.008

Bailey, M., & Dynarski, S. (2011). *Gains and gaps: Changing inequality in U.S. college entry and completion* (No. w17633). National Bureau of Economic Research. https://doi.org/10.3386/w17633

Bastedo, M. N., & Flaster, A. (2014). Conceptual and methodological problems in research on college undermatch. *Educational Researcher*, *43*(2), 93–99. https://doi.org/10.3102/0013189X14523039

Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *Quarterly Journal of Economics*, *127*(3), 1205–1242. https://doi.org/10.1093/qje/qjs017

Byun, S., Meece, J. L., Irvin, M. J., & Hutchins, B. C. (2012). The role of social capital in educational aspirations of rural youth. *Rural Sociology*, *77*(3), 355–379. https://doi.org/10.1111/j.1549-0831.2012.00086.x

Cabrera, A. F., & La Nasa, S. M. (2000). Understanding the college-choice process. *New Directions for Institutional Research*, *2000*(107), 5–22. https://doi.org/10.1002/ir.10701

Castleman, B. L., & Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, *115*, 144–160. https://doi.org/10.1016/j.jebo.2014.12.008

Ceja, M. (2006). Understanding the role of parents and siblings as information sources in the college choice process of Chicana students. *Journal of College Student Development*, *47*(1), 87–104. https://doi.org/10.1353/csd.2006.0003

Desmond, M., & Turley, R. N. L. (2009). The role of familism in explaining the Hispanic-White college application gap. *Social Problems*, *56*(2), 311–334. https://doi.org/10.1525/sp.2009.56.2.311

Dynarski, S., & Scott-Clayton, J. (2013). *Financial aid policy: Lessons from research* (No. w18710). National Bureau of Economic Research. https://doi.org/10.3386/w18710

Engberg, M. E., & Wolniak, G. C. (2010). Examining the effects of high school contexts on postsecondary enrollment. *Research in Higher Education*, *51*(2), 132–153. https://doi.org/10.1007/s11162-009-9150-y

González Canché, M. S. (2017). The Heterogeneous non-resident student body: Measuring the effect of out-of-state students' home-state wealth on tuition and fee price variations. *Research in Higher Education*, *58*(2), 141–183. https://doi.org/10.1007/s11162-016-9422-2

González Canché, M. S. (2018). Geographical network analysis and spatial econometrics as tools to enhance our understanding of student migration patterns and benefits in the U.S. higher education network. *The Review of Higher Education*, *41*(2), 169–216. https://doi.org/10.1353/rhe.2018.0001

Hillman, N. W. (2016). Geography of college opportunity: The case of education deserts. *American Educational Research Journal*, *53*(4), 987–1021. https://doi.org/10.3102/0002831216653204

Hillman, N., & Weichman, T. (2016). Education deserts: The continued significance of "place" in the twenty-first century. *Viewpoints: Voices from the Field.*

Hossler, D., Braxton, J., & Coopersmith, G. (1989). Understanding student choice. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 5, pp. 231–281). Agathion Press.

Hossler, D., & Gallagher, K. S. (1987). Studying student college choice: A three-phase model and the implications for policymakers. *College and University*, *62*(3), 207–221.

Ingels, S. J., Pratt, D. J., Herget, D. R., Bryan, M., Fritch, L. B., Ottem, R., Rogers, J. E., & Wilson, D. (2015). *High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript Data File Documentation* (NCES 2015-036). National Center for Education Statistics.

Lansing, J. (2017). A new model of college choice for distance learners. *Journal of Educational Technology Systems*, *45*(3), 365–389. https://doi.org/10.1177/0047239516673183

Litten, L. H. (1982). Different strokes in the applicant pool: Some refinements in a model of student college choice. *The Journal of Higher Education*, *53*(4), 383–402. https://doi.org/10.1080/00221546.1982.11780470

Mixon, F. G. (1992). Factors affecting college student migration across states. *International Journal of Manpower*, *13*(1), 25–32. https://doi.org/10.1108/EUM0000000000900

Mixon, F. G., & Hsing, Y. (1994). College student migration and human capital theory: A research note. *Education Economics*, *2*(1), 65–73. https://doi.org/10.1080/09645299400000005

Niu, S. X. (2015). Leaving home state for college: Differences by race/ethnicity and parental education. *Research in Higher Education*, *56*(4), 325–359. https://doi.org/10.1007/s11162-014-9350-y

Nora, A. (2004). The role of habitus and cultural capital in choosing a college, transitioning from high school to higher education, and persisting in college among minority and nonminority Students. *Journal of Hispanic Higher Education*, *3*(2), 180–208. https://doi.org/10.1177/1538192704263189

Ovink, S. M., & Kalogrides, D. (2015). No place like home? Familism and Latino/a–white differences in college pathways. *Social Science Research*, *52*, 219–235. https://doi.org/10.1016/j.ssresearch.2014.12.018

Paulsen, M. B. (2001). The economics of human capital and investment in higher education. In M. B. Paulsen & J. C. Smart (Eds.), *The Finance of Higher Education: Theory, Research, Policy, and Practice*. Algora.

Paulsen, M. B., & St. John, E. P. (2002). Social class and college costs: Examining the financial nexus between college choice and persistence. *Journal of Higher Education*, *73*(2), 189–236. https://doi.org/10.1080/00221546.2002.11777141

Perez, P. A., & McDonough, P. M. (2008). Understanding Latina and Latino college choice: A social capital and chain migration analysis. *Journal of Hispanic Higher Education*, *7*(3), 249–265. https://doi.org/10.1177/1538192708317620

Perna, L. W. (2006). Studying college access and choice: A proposed conceptual model. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 21, pp. 99–157). Springer Netherlands. https://doi.org/10.1007/1-4020-4512-3_3

Perna, L. W., & Titus, M. A. (2005). The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences. *Journal of Higher Education*, *76*(5), 485–518. https://doi.org/10.1080/00221546.2005.11772296

Pluhta, E. A., & Penny, G. R. (2013). The effect of a community college promise scholarship on access and success. *Community College Journal of Research and Practice*, *37*(10), 723–734. https://doi.org/10.1080/10668926.2011.592412

Rendon, L. I. (1994). Validating culturally diverse students: Toward a new model of learning and student development. *Innovative Higher Education*, *19*(1), 33–51. https://doi.org/10.1007/BF01191156

Roderick, M., Coca, V., & Nagaoka, J. (2011). Potholes on the road to college: High school effects in shaping urban students' participation in college application, four-year college enrollment, and college match. *Sociology of Education*, *84*(3), 178–211. https://doi.org/10.1177/0038040711411280

Sá, C., Florax, R. J., & Rietveld, P. (2004). Determinants of the regional demand for higher education in the Netherlands: A gravity model approach. *Regional Studies*, *38*(4), 375–392. https://doi.org/10.1080/03434002000213905

Salisbury, M. H., Umbach, P. D., Paulsen, M. B., & Pascarella, E. T. (2009). Going global: Understanding the choice process of the intent to study abroad. *Research in higher education*, *50*(2), 119-143.

Schuetze, H. G., & Slowey, M. (2002). Participation and exclusion: A comparative analysis of non-traditional students and lifelong learners in higher education. *Higher Education*, *44*(3), 309–327. https://doi.org/10.1023/A:1019898114335

Sherwin, J. (2012). *Make me a match: Helping low-income and first-generation students make good college choices*. MDRC.

Smith, J., Pender, M., & Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review*, *32*, 247–261. https://doi.org/10.1016/j.econedurev.2012.11.001

St. John, E. P. (2001). The impact of aid packages on educational choices: High tuition-high loan and educational opportunity. *Journal of Student Financial Aid*, *31*(2), 35–54

St. John, E. P., Asker, E. H., & Hu, S. (2001). The role of finances in student choice: A review of theory and research. In M. B. Paulsen & J. C. Smart (Eds.), *The Finance of Higher Education: Theory, Research, Policy, and Practice*. Algora.

Turley, R. N. L. (2006). When parents want children to stay home for college. *Research in Higher Education*, *47*(7), 823–846. https://doi.org/10.1007/s11162-006-9017-4

Turley, R. N. L. (2009). College proximity: Mapping access to opportunity. *Sociology of Education*, *82*(2), 126–146. https://doi.org/10.1177/003804070908200202

West, B. T., Sakshaug, J. W., & Aurelien, G. A. S. (2016). How big of a problem is analytic error in secondary analyses of survey data? *PLOS ONE*, *11*(6), 1–29. https://doi.org/10.1371/journal.pone.0158120

CHAPTER 2: THE RELATIONSHIP BETWEEN WANTING TO STAY HOME AND COLLEGE ENROLLMENT PATTERNS

Introduction

One of the most important components of the college choice decision-making process is *habitus*, or the predispositions towards college-going that a student holds (Hossler et al., 1989; Nora, 2004; Perna, 2006). It is obvious to understand why this would be the case—the preferences we hold influence what choices seem more or less attractive to us when we are forced to make decisions and enrollment in postsecondary education is no different. The literature on habitus and its connection to college access and attainment is well-explored (e.g. Cerna, Pérez, & Sáenz, 2009; Hamrick & Stage, 2004; Massé, Perez, & Posselt, 2010; Nora, 2004). Thanks to decades of research, we take it for granted that predispositions are important in studying college access. However, scholars have not fully explored a crucial precollege preference that likely impacts the availability of college options and enrollment decisions: the preference to stay close to home while attending college.

Literature Review

Factors that Affect College Choice

A large contingency of the college access and choice literature explores how various factors—both at the individual level and beyond—affect college-going patterns college choice decision-making. Socioeconomic status, racial and ethnic identification, parental education, and geographic origin are all associated with differential outcomes in college access, aspirations, persistence, and educational attainment (Bailey & Dynarski, 2011; Byun, Meece, Irvin, & Hutchins, 2012; Cabrera & La Nasa, 2000; Castleman & Page, 2014; Cho, Hudley, Lee, Barry, & Kelly, 2008; Desmond & Turley, 2009; Goldrick-Rab & Pfeffer, 2009; Hillman, 2016; Mayhew et al., 2016; Nora, 2004; Perna, 2000; Smith, Pender, & Howell, 2013). Studies on

college choice have sought to answer questions about what factors contribute to these differences—in particular, college choice examines the characteristics that influence students when they consider higher education as an option, where they wish to attend, and whether they ultimately enroll in a college (Cabrera & La Nasa, 2000). These factors exist within and across several contexts and affect the decision-making process (Perna, 2006). When students make the mental decision to attend college in general or specific institutions, these contexts incorporate themselves in the choice process in a multifaceted way—students draw from their individual characteristics, family characteristics, school-level characteristics and educational context, and geographic context to ultimately decide whether postsecondary education is a right fit for them, and if so, what institutions are reasonable options for them.

There are many well-explored student and family characteristics that are associated with college choice. First, as mentioned previously, there are many demographic and socioeconomic factors associated with college choice. Beyond these background characteristics, *college readiness* and academic achievement have implications for college choice. College readiness refers to the preparation possessed by students that will allow them to enroll and persist in higher education without the need for remediation (Conley, 2008; Tierney & Sablan, 2014). Often, college readiness is assessed through measurable, quantifiable academic traits from students' educational experiences (Tierney & Sablan, 2014). For example, taking math courses beyond Algebra 1 (Horn, Kojaku, & Carroll, 2001) or AP courses (Attewell & Domina, 2008) are often considered signals of curriculum intensity that can better propel students through higher education. Higher levels of academic achievement—often measured as high school GPA, higher SAT and ACT scores, and college-level course-taking such as AP courses—are associated with higher rates of college enrollment, and enrollment in four-year colleges specifically. College readiness extends beyond markers of academic achievement, however. Other, less measurable traits such as self-efficacy can have implications for college readiness (Conley, 2008). In a similar vein, lower student expectations can make it more difficult for students to consider themselves prepared for college and less willing to engage in postsecondary education.

At the family level, parental involvement and expectations (Cabrera & La Nasa, 2000; Perna & Titus, 2005), parental education (Nora, 2004), and family income (Bailey & Dynarski, 2011; Goldrick-Rab & Pfeffer,

2009; Paulsen & St. John, 2002; Perna, 2000) are all associated with students' college enrollment rates as well. In general, when parents are more involved in their children's education, have higher household incomes, and are more educated themselves, their children will have a higher probability of choosing to attend postsecondary institutions. School context matters in the college choice process as well. Research conducted through the college match perspective focus attention on the development of *college-going cultures* within high schools, finding that schools that cultivate strong cultures of college-going decrease the levels of academic mismatch and increase levels of educational aspirations among matriculating students (Roderick et al., 2011; Smith et al., 2013). High schools serve as a source of resources for students on the path towards postsecondary education—schools with higher levels of parental involvement, higher average SES of the student population, and college-relevant peer and parental social capital networks are better situated to send students to college (Engberg & Wolniak, 2010; Roderick et al., 2011).

Habitus and Predispositions Towards College-Going

Habitus, a sociological concept posited by Pierre Bourdieu, is defined as "an enduring, internal system of values, attitudes, beliefs, and actions, which is derived from the student's immediate family, community, and school environments and is common to members of one's social class" (Paulsen & St. John, 2002, p. 196). An individual's habitus drives the subconscious determination of what college pathways are possible and which are unreasonable (Perna & Titus, 2005). Habitus can facilitate college access through personal educational aspirations and perceptions of the capacity to afford college (Cabrera & La Nasa, 2000; Perna, 2006). These dispositions do not materialize on their own; as mentioned in the definition, the student's context and surroundings play a substantial role in developing habitus. Byun and colleagues (2012) found that school and family social capital were related to educational aspirations. Similarly, Roderick and colleagues (2011) noted that stronger "college-going-climates" (p. 179) in school were associated both with higher aspirations and higher likelihoods of college match or overmatch. On the other hand, habitus can create barriers to access—low-income students and first-generation college students differ in their perceptions of college cost compared to their peers, believing college to be unaffordable or holding the expectation that they will not have the means to pay for college (Bailey & Dynarski, 2011; Paulsen & St. John, 2002).

Geospatial Considerations

More recently, scholars have explored how geography and place have an impact on college enrollment. Current research on the role of geography and space in college-going patterns suggests that distance to college options is an important component of the college decision-making process (Alm & Winters, 2009; Desmond & Turley, 2009; Frenette, 2004; Klasik et al., 2018; Niu, 2015; Perez & McDonough, 2008; Sá et al., 2004; Turley, 2006). The availability of postsecondary options within proximity to students—or a lack thereof—can serve as a structural factor for college choice that may help or harm their likelihood of attending college. The *geography of opportunity* (Hillman, 2016; Hillman & Weichman, 2016) would suggest that choosing whether to attend college is one situated within the geographic context of local, accessible college options. A dearth of college options is a form of geographic inequality that is associated with racial, ethnic, and socioeconomic characteristics (Hillman, 2016). For example, communities with high Latinx populations have fewer four-year, and selective institutions nearby compared to Black and White communities, while larger Asian and high-income areas tend to have more two- and four-year options in proximity (Hillman, 2016). Regions also differ from each other in terms of college options, where students in the Eastern region of the United States have more options than students in the West, Midwest, and South (Turley, 2009).

The extant literature provides evidence that some students are more sensitive to geographic factors when making college choices. The sensitivity of enrollment and choice patterns relative to the distance between a student and their postsecondary options is known as *distance elasticity* (Hillman 2016; 2017). Low-income students, Black and Latinx students, and those with familial obligations are more distance elastic, while higher-income students, White students, and Asian students tend to be more distance inelastic (Alm & Winters, 2009; Mixon, 1992; Niu, 2015; Tuckman, 1970; Turley, 2006, 2009). Within group comparisons also provide evidence for the importance of distance when choosing to attend college. For example, there is also evidence that distance has an interactive effect with income; students from lower income families who do not have college options nearby are less likely to attend a four-year college than low-income students who do have college options nearby (Frenette, 2004).

What is less understood is how, or if, the student's preference for remaining local after high school is associated with college-going behaviors. It is a reasonable to assume that students who want to live close to home for college will seek out opportunities nearby—however, there is very little research that exists examining whether there is a real association between this preference and with college attendance. Moreover, should an association exist between the preference to remain close to home and college enrollment, there remains to be seen whether the association differs based on the availability of local college options. Lastly, there is no research that explores whether students who enrolled in college and wished to live close to home ultimately enrolled in colleges nearby.

The Nexus Between Habitus and Geography

Aside from educational aspirations and perceptions of affordability, another component of habitus that may be influential in the college-choice process is a desire to live close to home while pursuing higher education. There are several reasons why students may prefer to stay home while attending college. The economic burden of a college education can be reduced substantially for students who remain with family if their family is able to adequately provide food, shelter, and other necessities for the student. For example, there is evidence that one of the drivers of millennial interdependence and reliance on family housing prior to and after college stems from student loan debt (Bleemer, Brown, Lee, & van der Klaauw, 2014). There is also a cultural argument to be made about staying home: it is a common cultural value for familial desires or needs can often supersede the desires or needs of the student (also known as *familism*). Though this trait is often associated with Latinx students in the literature, it can be applied to more populations (Desmond & Turley, 2009; Keefe, 1984; Schwartz, 2007). Unfortunately, the familial desire for students to remain close to home are negatively associated with college-going (Desmond & Turley, 2009; Turley, 2006).

What is less understood is how, or if, the student's preference for remaining local after high school is associated with college-going behaviors. It is a reasonable to assume that students who want to live close to home for college will seek out opportunities nearby—however, there is very little research that exists examining whether there is a real association between this preference and with college attendance. Moreover, should an association exist between the preference to remain close to home and college enrollment, there

remains to be seen whether the association differs based on the availability of local college options. Lastly, there is no research that explores whether students who enrolled in college and wished to live close to home ultimately enrolled in colleges nearby.

Research Questions

In order to explore how preferences related to location correlate with college choices, this study will posit three questions to guide the research. First, what factors and characteristics are associated with a preference to stay near home for college? This research question serves as an exploration into the levers that might drive or are connected to the predispositions of college-aspiring students. Second, is the preference to remain close to home associated with college enrollment net of other characteristics? While controlling for other factors associated with college enrollment, the study will address whether the preference is tied to enrollment behaviors. Lastly, is it also associated with the distance of the first enrolled college for those students who chose to attend college? There is evidence that, under certain circumstances such as a lack of two-year college options nearby, students who do not live near colleges will seek opportunities outside of their communities (Frenette, 2004). Therefore, it is possible that students who have a desire who remain close to home for college will nevertheless vary in the distances of their enrolled colleges.

Through a combination of descriptive statistics and multilevel modeling, the questions will be explored in the context of traditional-aged, college-aspiring high school students. Both the age and aspirational qualifiers to the analytic population are a result of the data utilized for this study—more information on the data and the sample are below. I hypothesize that the preference to remain home will be associated with college going behaviors—both attending college, and the distance of college attended. Distance, for the purposes of this study, will be defined as the geodesic distance in miles between the high school attended in 2011 (the closest approximation to the students' homes available in the data) to the college attended in the fall of 2013.

Data

Data Source and Description

The proposed study examines the preference to remain at home and its association with college application behaviors. In order to provide the broadest generalizability, the study utilizes nationally representative data found in the High School Longitudinal Study of 2009 (HSLS:09). HSLS:09 is a federally mandated survey developed by the National Center for Education Statistics (NCES) that seeks to answer questions related to the transition between secondary and postsecondary education (Ingels et al., 2011; Ramirez, Lacy, Duprey, & Jones, 2019). The longitudinal study focuses primarily on educational decision-making—for example, science, math, engineering and technology (STEM) coursetaking and the decision to enter STEM fields later in life—and how the decision-making process is affected by a variety of sources including social context and student background (Ingels et al., 2011).

As a longitudinal study, HSLS combines multiple years of data across several data collection times into a single analytic panel. The proposed study utilizes data from the base-year (BY) data collection conducted in 2009, first follow-up (F1) conducted in 2012, and the brief data updated conducted in 2013 (U13). As of this proposal, further follow-ups have been conducted, including a second follow-up in 2016 (Ingels et al., 2015); however, for the purposes of this particular study, data up to U13 are sufficient for examining the research questions proposed.

HSLS:09 and its follow-ups contain a variety of data. First and foremost are student questionnaire data collected via computer assisted self-administered interviews (CASI) during school hours, or computer assisted telephone interviews (CATI) if the respondent was unable to complete the former (Ingels et al., 2011). This student questionnaire included questions regarding demographic information, school experiences, course-taking, self-efficacy and attitudes, and postsecondary college and career plans. Further questionnaires were conducted—an interview with the parent/guardian "most familiar with the [student's] school situation" (p. 15), an interview with all teachers who had the student in one of their math or science courses, an interview with a school administrator, and with a school counselor (Ingels et al., 2011). These data were all linked to the individual student records. Similarly, the F1 data collection consists of student, parent, school administrator, and counselor questionnaires with adjustments to the questionnaires to account for dropouts and early graduates (Ingels et al., 2013). The U13 follow-up asked students about high school completions,

employment, postsecondary education goals and influencers, financial aid, and other behaviors—the questionnaire was then combined with high school transcript data through a separate data collection effort (Ingels et al., 2015).

These data were combined with other publicly available sources of data. The Common Core of Data (CCD) and the Private School Survey (PSS), two public data sets also provided by NCES, were utilized to merge longitude and latitude coordinates of the last high school attended by the first follow-up. These coordinates were later used to derive distance measures and other geographic variables. The Integrated Postsecondary Education Data System (IPEDS) was also utilized in this manner, allowing longitude and latitude of the postsecondary institutions attended by students in the sample to be merged onto their records. Not all records in HSLS were able to be merged in with CCD, PSS, and IPEDS, which reduced the analytic sample.

Sample

The analytic sample was developed by identifying students within HSLS that met certain criteria based on the nature of the data needed for the models that will be discussed in the following section. First, students in the sample needed to have been respondents for the BY, F1 and U13 survey data collections, as well as high school transcript respondents. More than 15,000 students met these criteria. Next, students needed to have identified in their F1 survey responses that they aspired to at least some postsecondary education after high school. This requirement was necessary due to the survey logic utilized for the item that measured the preference to remain close to home—respondents only saw this item if they indicated they aspired to postsecondary education. Only 3% of the overall F1 respondent group indicated they did not aspire to postsecondary education, and thus did not receive the survey item for the preference. About 14,700 students met these criteria and had nonmissing values for the close to home preference item. Of the 14,700 students at this stage, approximately 10,540 had complete data for all analytic variables included in the final model. The last requirement for inclusion in the analytic sample was an adequate merge onto CCD/PSS and IPEDS for geographic location data. This was needed in order to properly address research question 3—whether the preference to remain close to home impacted the distance of the first postsecondary institution

attended. Approximately 7,860 students remained in the analytic sample after adding the final requirement, due primarily to the nature of the research question requiring college attendance to be included in the analytic sample. Given the large drop of students in the analytic sample after applying this restriction, research question 1 and 2 were examined using the roughly 10,700 student sample, while research question 3 was examined with the more restrictive sample of 7,860 students. Based on these sample definitions, the population of interest for research questions 1 and 2 consisted 9th grade students in 2009 who aspired to postsecondary education, and the population of interest for research question 3 consisted 9th grade students in 2009 who aspired to postsecondary education and ultimately enrolled in college in the a Title IV-eligible postsecondary institution in the United States.

Data Processing

Data processing procedures were conducted using a combination of R, Stata, QGIS and Python. The HSLS:09/12 and U13 data were merged with CCD/PSS and IPEDS using the appropriate IDs for each student's F1 high school and first postsecondary college attended. These merges attached longitude and latitudes for the high school and college, respectively. Using the GEOTOOLS package in Stata, the longitude and latitude of each were used to create a shapefile containing lines between both schools. This shapefile was then ported into QGIS, where a Python script was run to calculate the length of each line in miles using the WGS 84/Pseudo-Mercator Coordinate Reference System (EPSG:3857). This CRS was chosen over the more common WGS 84 Coordinate Reference System (EPSG:4326) to better estimate the geodesic distance—that is, the straight line distance or "as the crow flies"—in miles rather than degrees. This form of distance was chosen over network distance to reduce computational burden. The calculated distances were then attached to the student record corresponding to the paired schools. During this stage of data processing, the Python script also utilized shapefiles containing 2010 commuting zone delineations—delineations of local economies and communities—alongside IPEDS to calculate the number of active postsecondary institutions within each commuting zone in 2013. Like the distance measure, these counts were attached to the HSLS student records based on the commuting zone their high school was located within. Finally, these data were imported to Stata 15 for final data cleaning, variable derivations, and statistical analyses.

Methods

Analysis Plan

I rely on ordered logistic regression (OLR) modeling, logistic regression and OLS in order to answer this study's research questions. OLR is a form of regression which is used to model categorical outcomes in which the levels of the outcome are ordered in a meaningful way (Fullerton, 2009; Williams, 2016). The primary assumption made for OLR involves the assumption that the difference between each pair of outcome categories are not equivalent—this assumption is called the proportional odds assumption or the parallel odds assumption (Fullerton, 2009; Williams, 2016). When this assumption is not fully met, partial proportional odds (PPO) modeling is an alternative model that allows the assumption to be relaxed for a subset of independent variables. During model specification, two independent variables (one variable and one category of another independent variable) did not meet the assumption—however, while testing a PPO version of the OLR model, the results were roughly equivalent when allowing the assumption to relax. Therefore, to simplify the interpretation of the results, OLR was used for the final analysis of research question 1. Logistic regression and OLS regression are more simple methods than OLR and utilized to model dichotomous and continuous outcomes, respectively. These models were deemed sufficient to address the outcomes of interest for research questions 2 and 3.

Several statistical models were tested for each research question.

$$CLOSEHOME_i = \beta_0 + \beta_1 D_i + \beta_2 A_i + \beta_3 C_i + \beta_4 F_i + \beta_6 S_i + \varepsilon_i \quad (1)$$

$$ENROLL_i = \beta_0 + \beta_1 CLOSEHOME_i + \beta_2 D_i + \beta_3 A_i + \beta_4 C_i + \beta_5 F_i + \beta_6 S_i + \varepsilon_i \quad (2)$$

$$Log(DISTANCE_i) = \beta_0 + \beta_1 CLOSEHOME_i + \beta_2 D_i + \beta_3 A_i + \beta_4 C_i + \beta_5 F_i + \beta_6 S_i + \varepsilon_i \quad (3)$$

Equations 1, 2, and 3, are used for research questions 1, 2, and 3, respectively. Each model has a different dependent variable: $CLOSEHOME_i$ refers to the stated importance of remaining close to home for college for student *i*; $ENROLL_i$ refers to enrollment in college for student *i*; and $DISTANCE_i$ refers to the log distance between home and the enrolled college for student *i*. In each model, β_0 is the constant; D is the vector of demographic and background variables; A is a vector of academic variables; C is a vector of social

and cultural capital variables; F is a vector of family variables; and S is a vector of school-level variables. ε_i refers to observation error. In research questions 2 and 3, $CLOSEHOME_i$ becomes the predictor of interest.

Variables

The variables used for this study included derived variables delivered in HSLS:09/12 and the U13 follow-up, as well as additional variables derived using these data and secondary sources. Variables delivered as part of HSLS:09/12 and U13 will include their variable name as found in the restricted-use file (RUF). There are three dependent variables corresponding to each of the research questions.

Dependent Variables

The first is a measure of a preference for being close to home for college (S2CLOSEHOME). This measure comes from an F1 survey item that asks the survey respondent: "How important to you [will/would] each of the following characteristics be when choosing a school or college to attend after high school? (Ingels et al., 2013, p. A-41)" Several characteristics are presented to the respondent, of which one of them is "close to home." The three response options are "very important," "somewhat important," and "not at all important. (Ingels et al., 2013, p. A-42)" The second dependent variable is enrollment in postsecondary classes (X3CLASSES). The variable is derived from a U13 survey instrument that asks "Which of the following activities [will/were/was] [you/your teenager] [be] doing on or around November 1st? (Ingels et al., 2015, p. B-12)" The first activity presented is "Taking classes from a college, university, community college, trade school, or other occupational school (such as a cosmetology school or a school of culinary arts) (Ingels et al., 2015, p. B-12)" and the respondent could choose "yes," "no," or "don't know." The X3CLASSES variable includes both respondent data as well as imputed data for item nonrespondents. The final dependent variable is a measure of distance to the postsecondary institution enrolled. This variable was derived using the longitude and latitude coordinates of both the last high school attended at the time of the first follow-up and the first postsecondary institution attended. Shapefiles containing lines between the high school and postsecondary institution were created using the GEOTOOLS package in Stata, then the geodesic distance in miles was calculated by measuring the line produced in Stata using QGIS, an open source GIS software application.

Predictor Variables

The predictor variables used for the analyses will correspond to multicontextual factors and characteristics identified by scholars in the college access and choice literature as important considerations in the college decision-making process. The first group of variables are student-level *background and demographic characteristics*. These include race and ethnicity (X2RACE), sex (X2SEX), socioeconomic status, or SES (X2SES), first-generation college student status (derived from X2PAREDU) and number of high schools attended (X2NUMHS). The last variable is not a factor that has been identified in the literature previously; however, it may be associated with a student's preference for being close to home—specifically, students who have previously been mobile between schools may be more inclined to be mobile for college, as well

There are several student-level *academic* variables as well: transcript GPA (X3TGPA11TH), an indicator for whether the student took any AP courses (S2ANYAP), and several items related to college prep or admission behaviors—taking a course to prepare for a college admissions exam (S2CLGEXAMPREP), number of times the student took the PSAT or PLAN (S2PSATNUM), number of times they took the SAT or ACT (S2SATNUM), and the number of times they took any AP test (S2APEXAMNUM). Lastly, two measures of academic self-efficacy for math (X2MTHEFF) and science (X2SCIEFF) were included. These measures were added given their relevance to college readiness (Conley, 2008).

The next group of variables are related to the student's *social and cultural capital*. This group of variables is a combination of expectations, perceptions, and use of social capital related to college. The social capital indicators capture whether the student talked to several individuals about college in 9th grade: their parents (an indicator combining S1MOMTALKCLG and S1DADTALKCLG), friends (S1FRNDTLKCLG), teachers (S1TCHTALKCLG), and school counselor (S1CNSLTLKCLG). Cultural capital variables include student's education expectations (S2EDUEXP) and a school motivation scale (X2BEHAVEIN). The school motivation scale was included as part of HSLS, and measures whether the student engaged in certain behaviors, including school absenteeism, tardiness, and attending class without completing homework, that could be argued to indicate lower school motivation. In other words, lower scores on this scale suggests the student may be less motivated in school. Despite the problematic nature of the scale—tying behaviors that

may be out of the control of high school students to motivation—the measure was included under the assumption that the scale was tested as part of HSLS and validated as an accurate measure of school motivation. The last group of variables are related to the student's *family* and correspond with background characteristics as well as cultural and economic capital resources. These variables include the number of household members (X2HHNUMBER), total family income (X2FAMINCOME), and parental expectations for the student to attend college in the fall of 2013 (derived from S2MOSTIMP2013).

The second level of variables are related to the school the student attended in F1. These items include school control (X2CONTROL), the rural-urban locale code (X2LOCALE), the percent of 11th graders with free lunch (X2FREELUNCH), and the state where the school is located (X2STATE). Several items related to college going culture and college resources were also included, which pulled from reports made by the school counselor. These items included the availability or participation in college-going resources like college fairs (C2CLGFAIR), college information sessions (C2INFOSESSN), assistance with college applications (C2CLGAPPS), access to college information (C2CLGINFO), and assistance with selecting colleges to apply to (C2CLGSELECT). It is worth noting that for all of the students in both analytic samples attended the same high school in the base year as in F1; as such, while the variable labels differentiate between schools attended in the base year and in F1, they refer to the same institution for all students examined. Lastly, the model for research question 3 contains a derived variable identifying the number of postsecondary institutions in the same commuting zone—a delineation of local economies and communities—as the high school attended. This variable was derived using publicly available shapefiles and the longitude and latitude coordinates of high school and postsecondary institution in QGIS.

Weighting and Subpopulation Considerations

In order to account for the complex sampling design inherent in HSLS, analyses were conducted using the svyset command in Stata 15. This command allows Stata to calculate means and estimate variances while incorporating sampling weights and multi-stage sampling designs. HSLS:09 provides a multitude of sampling weights, sampling design variables, and balanced repeated replicate (BRR) weights for analysts to use to account for nonresponse and to provide more precise estimates that are generalizable to the student

target population. Similarly, they provide variables to identify the stratum (STRAT_ID) and the primary sampling unit (PSU). These analyses utilized W3W1W2STUTR as the primary sampling weight; this longitudinal weight covers school nonresponse, student questionnaire nonresponse, and high school transcript nonresponse for the BY, F1, and U13 data collections (Duprey et al., 2018). Additionally, the sampling design variables—the primary sampling unit identifier and strata identifier variables listed above—were incorporated. Taylor series linearization was used for variance estimation in accordance with these choices. Lastly, it must be noted that subpopulations in complex survey data should be treated differently than a traditional subset in a statistical program. Ideally, when accounting for complex sampling designs, you would treat members that are outside of the subpopulation as if they had a weight of zero and keeping them as part of the overall sample (Lumley, 2004). The svyset command in Stata provides an option to do this by using the subpop option, which was included in the analyses described herein.

Results

Descriptive Statistics

Details for both analytic samples—the sample for research questions 1 and 2, and the sample for research question 3—are presented in tables 2.1 and 2.2. Table 2.1 shows the weighted and unweighted percentages for each value of every categorical variable analyzed. The statistics described in this section will refer to the weighted percentages, means, and standard deviations unless otherwise specified.

Table 2.1: Frequency Statistics for Categorical Variables, by Analytic Sample

	Analytic Sample					
	Research question 1 and 2			Research question 3		
Variable	**N**	**Unweighted percent**	**Weighted percent**	**N**	**Unweighted percent**	**Weighted percent**
Dependent Variables						
Preference to remain close to home for college						
Very important	2,370	29.17	23.69	1,610	20.48	20.88
Somewhat important	5,090	48.33	48.62	3,860	49.07	49.45
Not at all important	3,080	29.17	27.69	2,390	30.45	29.66
Taking postsecondary classes in fall 2013						
Yes	8,510	80.77	76.98	7,860	100	100
No	2,030	19.23	23.02	†	†	†
Independent Variables						
Race						
White	7,250	68.82	68.49	5,410	68.76	69.39
Black/African American	950	8.98	12.07	680	8.59	10.97
Asian American	930	8.78	4.11	790	10.02	4.9
Native Hawaiian/Pacific Islander	110	1.08	1.46	80	1	1.43
American Indian/Alaska Native	230	2.17	3.14	150	1.93	2.73
Multiracial	1,070	10.16	10.73	760	9.69	10.58
Hispanic or Latinx						
Yes	1,420	13.44	18.72	980	12.4	17.75
No	9,120	86.56	81.28	6,890	87.6	82.25
First-generation college student						
FGCS	3,660	34.72	42.9	2,270	71.1	35.85
Non-FGCS	6,880	65.28	57.1	5,590	28.9	64.15

See notes at end of table

Table 2.1: Frequency Statistics for Categorical Variables, by Analytic Sample—Continued

Variable	Analytic Sample: Research question 1 and 2 N	Research question 1 and 2 Unweighted percent	Research question 1 and 2 Weighted percent	Research question 3 N	Research question 3 Unweighted percent	Research question 3 Weighted percent
Number of high schools attended						
One	10,300	97.7	97.4	7,740	98.51	98.44
Two	230	2.14	2.42	110	1.4	1.38
Three	20	0.15	0.19	10	0.09	0.18
Any AP courses taken						
Yes	4,340	41.13	38.5	3,710	47.25	45.93
No or not sure	6,210	58.87	61.5	4,150	52.75	54.07
Took a course to prepare for college admissions exam						
Yes	4,980	47.2	41.41	4,020	51.18	45.8
No	5,570	52.8	58.59	3,840	48.82	54.2
Number of times PSAT or PLAN taken						
Never or don't know	2,830	26.8	32.9	1,690	21.52	26.75
Once	4,250	40.28	38.88	3,260	41.52	41.01
Twice	2,510	23.76	21.22	2,110	26.87	24.46
3 or more times	970	9.15	7	790	10.09	7.79
Number of times SAT or ACT taken						
Never or don't know	5,670	53.79	61.43	3,890	49.45	57.45
Once	3,410	32.38	27.35	2,730	34.78	30.03
Twice	970	9.19	7.45	830	10.52	8.4
3 or more times	490	4.64	3.76	410	5.25	4.12
Talked to parents about going to college: 9th grade						
Yes	9,140	86.7	85.32	7,000	89.03	88.09
No	1,400	13.3	14.68	860	10.97	11.91

See notes at end of table.

Table 2.1: Frequency Statistics for Categorical Variables, by Analytic Sample—Continued

Variable	Analytic Sample: Research question 1 and 2 N	Research question 1 and 2 Unweighted Percent	Research question 1 and 2 Weighted Percent	Research question 3 N	Research question 3 Unweighted Percent	Research question 3 Weighted Percent
Talked to friends about going to college: 9th grade						
Yes	5,990	56.81	55.32	4,660	59.24	58.28
No	4,550	43.19	44.68	3,200	40.76	41.72
Talked to teacher about going to college: 9th grade						
Yes	2,180	20.7	21.47	1,680	21.33	21.76
No	8,360	79.3	78.53	6,180	78.67	78.24
Talked to counselor about going to college: 9th grade						
Yes	1,860	17.65	17.65	1,430	18.2	18.39
No	8,680	82.35	82.35	6,430	81.8	81.61
Postsecondary education expectations						
Less than certificate or don't know	1,490	14.1	16.05	800	10.1	11.18
Complete a certificate	480	4.58	5.22	200	2.57	2.92
Some 2-year college or 2-year degree completion	930	8.83	10.05	550	7.02	8.06
Some 4-year college or 4-year degree completion	3,120	29.56	29.71	2,430	30.95	31.56
Some Master's education or Master's completion	2,690	25.5	24.1	2,290	29.11	28.6
Some doctoral/terminal education or doctoral/terminal degree completion	1,840	17.43	14.88	1,600	20.25	17.67
Family income (categorical)						
Less than or equal to $15,000	810	7.68	9.9	470	6.03	7.44
$15,001 - $35,000	1,580	14.94	17.95	980	12.44	15.1
$35,001 - $55,000	1,690	16.03	17.47	1,170	14.82	16.6
$55,001 - $75,000	1,520	14.38	14.57	1,150	14.65	14.86
$75,001 - $95,000	1,290	12.19	11.46	990	12.62	12.39
$95,001 - $115,000	1,040	9.85	9.1	840	10.71	10.27

See notes at end of table.

Table 2.1: Frequency Statistics for Categorical Variables, by Analytic Sample—Continued

Variable	Analytic sample: Research questions 1 and 2 N	Research questions 1 and 2 Unweighted percent	Research questions 1 and 2 Weighted percent	Research question 3 N	Research question 3 Unweighted percent	Research question 3 Weighted percent
Family income (categorical)						
$115,001 - $135,000	730	6.93	5.73	610	7.8	6.38
Greater than $135,000	1,900	17.99	13.81	1,650	20.93	16.96
Parents think is most important for student to do in fall 2013						
Continuing education after high school	9,530	90.4	88.66	7,460	94.87	93.95
Something other than continuing education	1,010	9.6	11.34	400	5.13	6.05
First follow-up school: control						
Public	8,480	80.49	92.47	6,180	78.64	91.51
Private	2,060	19.51	7.53	1,680	21.36	8.49
First follow-up school: locale (urbanicity)						
City	2,880	27.35	27.82	2,270	28.83	28.15
Suburb	3,120	29.62	28.79	2,410	30.61	30.13
Town	1,370	13.02	12.63	970	12.29	12.05
Rural	3,160	30.02	30.77	2,220	28.28	29.69
First follow-up school: percent free lunch (categorical)						
Zero	1,390	13.18	5.63	1,160	14.77	6.37
0 - 9%	930	8.82	6.88	830	10.51	8.58
10 - 19%	1,190	11.32	10.03	1000	12.54	12.03
20 - 29%	1,340	12.71	14.06	1,040	13.23	15.14
30 - 39%	1,300	12.29	12.69	930	11.88	12.64
40 - 49%	1,290	12.25	15.03	890	11.36	14.54
50 - 59%	1,100	10.39	11.62	720	9.2	10.06
60 - 69%	870	8.25	8.9	580	7.42	7.97
Greater than or equal to 70%	1,140	10.79	15.15	720	9.1	12.68

See notes at end of table.

Table 2.1: Frequency Statistics for Categorical Variables, by Analytic Sample—Continued

Variable	Analytic sample: Research questions 1 and 2 N	Unweighted percent	Weighted percent	Research question 3 N	Unweighted percent	Weighted percent
Base year school: school holds or participates in college fairs						
Yes	9,680	91.83	92.18	7,220	91.83	92.32
No	860	8.17	7.82	640	8.17	7.68
Base year school: school holds college information sessions						
Yes	9,900	93.94	93.31	7,390	93.96	93.17
No	640	6.06	6.69	480	6.04	6.83
Base year school: School helps with completing college applications						
Yes	10,250	97.19	96.26	7,620	96.97	95.71
No	300	2.81	3.74	240	3.03	4.29
Base year school: School provides access to information on college						
Yes	10,500	99.37	99.45	7,810	99.36	99.48
No	70	0.63	0.55	50	0.64	0.52
Base year school: School helps with selecting colleges to apply to						
Yes	10,380	98.47	98.16	7,770	98.82	98.45
No	160	1.53	1.84	90	1.18	1.55

† Not applicable.
NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

The distribution for the preference to remain close to home for college, the first dependent variable of the study, was relatively even across all three categories. Approximately 24 percent of the first analytic sample indicated that living close to home was "very important" for considering college, while approximately 21 percent of the second analytic sample indicated the same. About 49 percent for both samples indicated it was somewhat important, while about 28 and 30 percent of both sample 1 and sample 2, respectively, indicated that it was not at all important. The second dependent variable, enrollment in postsecondary classes by the fall of 2013, was less even across samples. In sample 1, about 77 percent were enrolled in postsecondary classes by the fall of 2013, while 23 percent were not. Therefore, three quarters of the sample pursued higher education on a traditional timetable—enrolling in classes after completing high school. Given that research question 2 examines students who chose to enroll in college by the fall of 2013, all students in sample 2 were enrolled in the fall of 2013. The final dependent variable was the distance in miles from the high school to the enrolled college in the fall of 2013. For analytic sample 2, the minimum distance between high school and college was .108 miles, while the maximum was 4975.19 miles. The mean distance was 138.01 miles and the standard deviation was 8.28. However, the unweighted standard deviation was 341.43, suggesting that the data were heavily skewed.

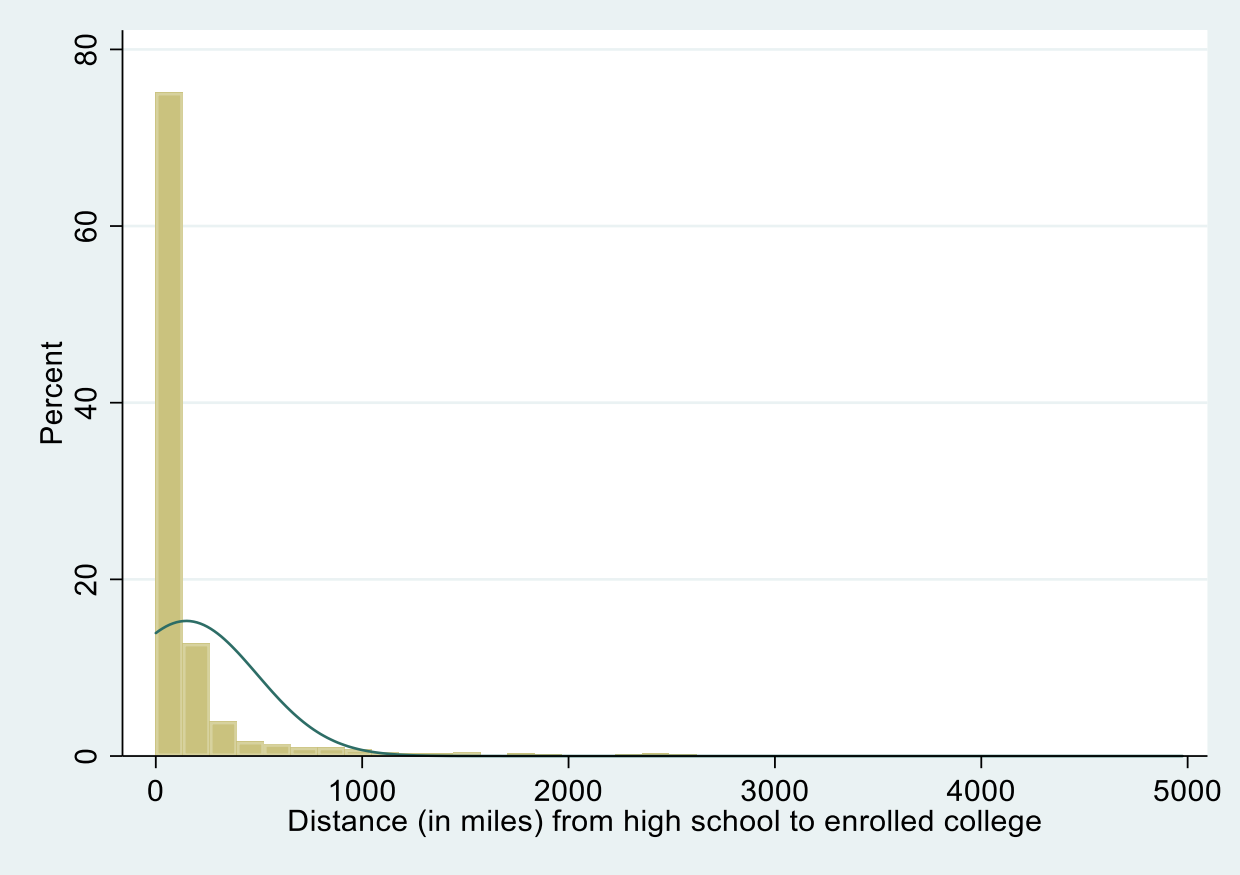
Table 2.2: Summary Statistics for Continuous Variables, by Analytic Sample

	Analytic sample											
	Research questions 1 and 2						Research question 3					
			Unweighted		Weighted				Unweighted		Weighted	
Variable	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**
Distance from high school to college attended in 2013	†	†	†	†	†	†	0.108	4975.186	148.322	341.427	138.01	8.279
Log-transformed distance	†	†	†	†	†	†	-2.226	8.512	3.604	1.757	3.442	0.062
Number of high schools attended	1	3	1.024	0.164	1.028	0.003	1	3	1.016	0.132	1.017	0.003
Socioeconomic status	-1.75	2.282	0.201	0.746	0.026	0.018	-1.715	2.282	0.311	0.731	0.151	0.019
Age	16	23	18.336	0.569	18.349	0.011	16	23	18.283	0.516	18.283	0.01
11th grade GPA reported on transcript	0	4	2.829	0.951	2.688	0.233	0	4	3.02	0.836	2.924	0.02
Math self-efficacy	-2.5	1.73	0.076	0.992	0.032	0.016	-2.5	1.73	0.134	0.977	0.094	0.018
Science self-efficacy	-2.47	1.64	0.075	0.99	0.04	0.018	-2.47	1.64	0.116	0.977	0.084	0.021
School Motivation Scale	-5.64	1.21	0.124	0.875	0.064	0.014	-5.64	1.21	0.214	0.796	0.173	0.014
Number of household members	1	13	4.227	1.396	4.236	0.024	1	13	4.213	1.361	4.234	0.026

† Not applicable.
NOTE: SD = Standard Deviation
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

Figure 2.1 shows the unweighted distribution of the distance from high school to enrolled college for the second analytic sample along with the predicted normal distribution of distance given the unweighted mean and standard deviation. The unweighted distribution of this variable was positively skewed to an extreme degree (M = 148.32, SD = 341.43); about 75 percent of the sample had less than 150 miles of distance between their high school and enrolled college and the maximum value for any student was about 4975 miles. Given the highly skewed data, the models for research question 3 were examined using the log-transformed value of distance to normalize the data.

Figure 2.1: Distribution of Distance from High School to Enrolled College for Analytic Sample 2, in Miles

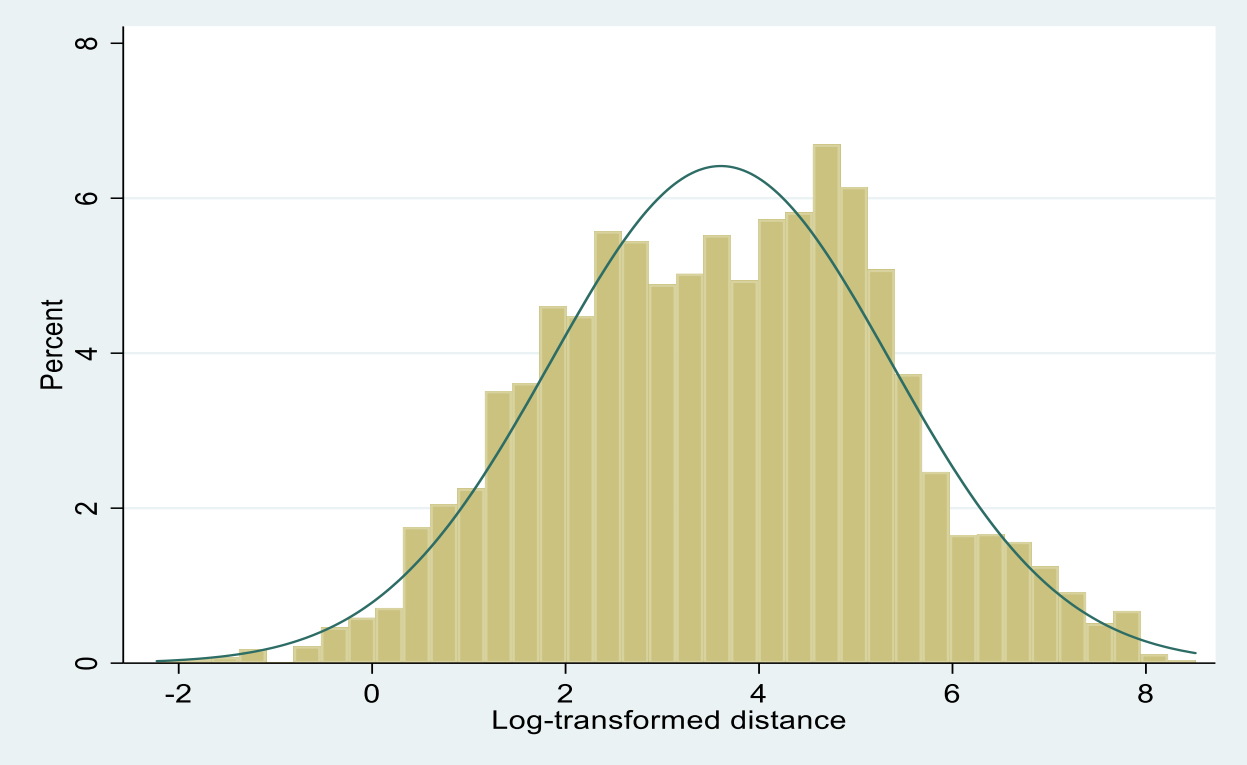


SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

The distribution of the distance outcome after conducting a log transformation was less skewed than the original values. The minimum and maximum values for the log of distance were -2.226 and 8.512, and the

weighted mean was 3.442 with a standard deviation of .062. As shown in figure 2.2, the distribution was normalized, allowing for more precise analysis.

Figure 2.2. Distribution of Log-Transformed Distance from High School to Enrolled College for Analytic Sample 2



SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

Regression Results

Models were created in a stepwise fashion, with progressively more variables added in subsequent models. The final model included background characteristics, academic characteristics, social and cultural capital, family characteristics, and school characteristics, as well as a state fixed effect to control for the state that the student resided in. Interactions were examined for several variables within the models, with special attention paid between the interaction between race and ethnicity—in this case, an indicator of *Hispanidad* as no other ethnicities were measured in HSLS. This interaction was tested as the literature has pointed to traits among Hispanic and Latinx students (specifically, familism) as impacting postsecondary enrollment behaviors

and habitus (Desmond & Turley, 2009; Keefe, 1984; Schwartz, 2007). For ease of interpretation of the results, reported odds from the OLR models will describe comparisons of the difference between the "very important" preference category and the combined "somewhat important" and "not at all important" categories, and the difference between the combined "very important" and "somewhat important" categories versus the "not at all important" category simply in terms of higher or lower preferences. OLR models presents a single set of results despite describing multiple relationships; the models reinterpret the outcome of interest as a dichotomous variable, wherein one value is a single category of the outcome, and the other is a combined set of the remaining categories (Williams, 2016). In this case, as the preference to remain close to home is an outcome with three categories, the results describe both derivations of the outcome noted previously. Instead of introducing additional confusion in the interpretation of the results as the proportional odds assumption was assumed to be met, the results will simply be described as increasing or decreasing the odds of a stronger preference.

Modeling the Preference to Live Close to Home

Research question 1 explored the factors that were associated with the preference to live close to home while attending college. The results of the ordered logistic regression models for research question 1 are listed in table 2.3, with the outcomes of each predictor presented as odds ratios. The full OLR model, weighted and accounting for survey design, explained a significant portion of the variance in the preference to remain close to home for college, $F(40, 452) = 10.00$, $p < .001$.

Table 2.3: Ordered Logistic Regression Results for Variables Associated with the Preference to Remain Close to Home, by Model

	Ordered logistic regression model									
	Model 1 (Background only)		Model 2		Model 3		Model 4		Model 5 (Full model)	
Covariates	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**
Background Characteristics										
Race (ref. = White)										
Black or African American	0.951	(0.096)	0.92	(0.097)	0.939	(0.096)	0.923	(0.093)	0.89	(0.093)
Asian	0.882	(0.095)	0.991	(0.112)	0.973	(0.114)	0.958	(0.11)	1.001	(0.123)
Native Hawaiian/Pacific Islander	2.517	(1.338)	2.584	(1.365)	2.845	(1.531)	2.812	(1.524)	2.755*	(1.401)
American Indian/Alaska Native	0.839	(0.224)	0.793	(0.215)	0.882	(0.256)	0.861	(0.249)	0.794	(0.228)
Multiracial	0.916	(0.095)	0.906	(0.091)	0.909	(0.096)	0.898	(0.095)	0.887	(0.092)
Hispanic or Latinx status	0.832	(0.079)	0.83	(0.080)	0.84	(0.081)	0.834	(0.081)	0.828*	(0.079)
Race and Hispanic Interaction Effects (ref. = non-Hispanic)										
Black or African American, Hispanic	2.232*	(0.911)	2.082	(0.88)	1.965	(0.731)	1.963	(0.714)	1.982*	(0.652)
Asian, Hispanic	0.371	(0.311)	0.341	(0.273)	0.312	(0.247)	0.318	(0.252)	0.293	(0.241)
Native Hawaiian/Pacific Islander, Hispanic	0.221*	(0.132)	0.213*	(0.129)	0.181**	(0.112)	0.183**	(0.114)	0.179**	(0.107)
American Indian/Alaska Native, Hispanic	2.041*	(0.696)	2.1*	(0.727)	1.842	(0.658)	1.898	(0.677)	2.031*	(0.726)
Multiracial, Hispanic	0.838	(0.238)	0.871	(0.262)	0.908	(0.286)	0.921	(0.295)	0.963	(0.307)
First-generation college student	1.003	(0.075)	1.011	(0.078)	1.003	(0.076)	1.029	(0.08)	1.018	(0.081)
Number of high schools attended	0.876	(0.133)	0.771	(0.117)	0.788	(0.113)	0.784	(0.122)	0.749	(0.118)
Socioeconomic Status	0.655***	(0.037)	0.731***	(0.043)	0.745***	(0.045)	0.797**	(0.062)	0.829*	(0.065)
Age	1.232***	(0.063)	1.151**	(0.055)	1.132*	(0.056)	1.13*	(0.056)	1.116*	(0.055)
Academic Characteristics										
11th grade GPA			0.904***	(0.026)	0.891***	(0.27)	0.893***	(0.027)	0.893***	(0.028)
Any AP courses taken			0.75***	(0.464)	0.799***	(0.05)	0.799***	(0.051)	0.783***	(0.052)
College exam preparation			0.966	(0.053)	0.967	(0.052)	0.971	(0.052)	0.962	(0.052)
PSAT/PLAN tests taken			0.906**	(0.029)	0.910**	(0.032)	0.911**	(0.323)	0.93*	(0.034)

See notes at end of table.

Table 2.3: Ordered Logistic Regression Results for Variables Associated with the Preference to Remain Close to Home, by Model—Continued

	Ordered logistic regression model									
	Model 1 (Background only)		Model 2		Model 3		Model 4		Model 5 (Full model)	
Covariates	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**
Academic Characteristics										
SAT/ACT tests taken			1.043	(0.038)	1.043	(0.038)	1.044	(0.038)	1.052	(0.375)
Math self-efficacy			0.941*	(0.029)	0.93*	(0.029)	0.929*	(0.029)	0.932*	(0.029)
Science self-efficacy			0.994	(0.028)	1.002	(0.029)	1.002	(0.029)	0.995	(0.028)
Social/Cultural Capital										
Talked to parents about college: 9th grade					1.085	(0.09)	1.089	(0.091)	1.102	(0.092)
Talked to friends about college: 9th grade					0.853**	(0.052)	0.851**	(0.052)	0.852**	(0.052)
Talked to teachers about college: 9th grade					0.974	(0.071)	0.979	(0.072)	0.957	(0.07)
Talked to counselors about college: 9th grade					1.033	(0.083)	1.025	(0.082)	1.001	(0.081)
School motivation scale					1.244***	(0.043)	1.245***	(0.043)	1.243***	(0.044)
Educational expectations scale					0.902***	(0.019)	0.904***	(0.02)	0.906***	(0.019)
Family Characteristics										
Household size							1.031	(0.02)	1.027	(0.019)
Family income (categorical)							0.974	(0.016)	0.983	(0.017)
Parental expectations: most important for student to attend college in fall 2013							0.958	(0.109)	0.967	(0.108)
School Characteristics										
Private school									0.889	(0.088)
School locale (ref = City)										
Suburb									1.094	(0.092)
Town									1.195	(0.121)
Rural									1.037	(0.082)

See notes at end of table.

Table 2.3: Ordered Logistic Regression Results for Variables Associated with the Preference to Remain Close to Home, by Model—Continued

	Ordered logistic regression model									
	Model 1 (Background only)		Model 2		Model 3		Model 4		Model 5 (Full model)	
Covariates	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**
School Characteristics										
Percent free lunch eligible students									1.048**	(0.017)
School had college fairs									0.776	(0.102)
School had college information sessions									1.138	(0.111)
School assisted with college applications									0.947	(0.11)
School provided access to college information									0.7225	(0.173)
School assisted with selecting colleges									0.99	(0.175)
State (fixed effect)									1.003	(0.002)
/cut 1	2.652	(0.964)	0.789	(0.903)	0.191	(0.951)	0.184	(0.984)	-0.211	(0.977)
/cut 2	4.846	(0.965)	3.013	(0.901)	2.443	(0.947)	2.438	(0.982)	2.053	(0.973)
F-Statistic	*F*(20,480) = 12.20***		*F*(20, 470) = 11.94***		*F*(30, 460) = 12.73***		*F*(30, 460) = 11.83***		*F*(40, 450) = 10.00***	

* $p < .05$, ** $p < .01$, *** $p < .001$

NOTE: Degrees of freedom rounded to the nearest 10. OR = Odds ratio. Ref = Reference group. SD = Standard deviation.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

When including the full battery of predictors drawn from college choice conceptual frameworks and the extant literature, several characteristics were significantly associated with the preference. Hispanic or Latinx ethnicity was significantly associated with the preference, where Hispanidad net of other characteristics decreased the odds of a higher preference (t = -2.03, p < .05). Using White students as a reference group, White Hispanic students were less likely to have a higher preference for living close to home for college than White non-Hispanic students. This alone contradicts some of the evidence in literature Hispanic or Latinx ethnicity interacted with race in a meaningful way, wherein race categories were not alone associated with the preference to remain close to home, but several were when interacted with Hispanic or Latinx ethnicity. African-American Hispanic students were 2.02 times more likely to have a stronger preference to remain close to home that African-American non-Hispanic students (t = 2.06, p < .05), while Native Hawaiian/Pacific Islander Hispanic students were .18 times more likely to indicate a strong preference than Native Hawaiian/Pacific Islander non-Hispanic students (t = -2.85, p < .01). The results suggest that identifying as Hispanic or Latinx impacts the preference to remain close to home. Surprisingly, both race and ethnicity did not impact the preference as hypothesized; no race category was significantly associated on its own, although Native Hawaiian/Pacific Islander approached significance (t = 1.96, p = .051). Hispanic or Latinx status on its own was significant but was negatively associated with the preference which contradicts the original hypothesis. Additionally, the interactions between the racial categories and Hispanidad either decreased or increased the odds of higher preferences depending on the racial category—the complexity of how Hispanidad impacts the importance of living close to home for prospective college students is one worth further study, but beyond the scope of this paper.

Other significant demographic characteristics included SES (t = -2.42, p < .05), which decreased the odds of having a stronger preference as SES goes up, and age (t = 2.2, p < .001), where older students had higher odds of having a stronger preference. The evidence suggests that in terms of demographics, older students and students from lower SES families are more likely to find it important to stay near home when considering college—it may be the case that these students are more likely to have obligations that require them to remain near home, like supporting dependents or family members. It is also worth noting that the

variable included that is not traditionally in college choice frameworks or the literature, number of high schools attended, was not significantly associated with the preference (t = -1.77, p = .077).

Several academic characteristics were also significantly associated with the preference. Of these, the strongest association was between the preference and AP coursetaking, where taking AP courses was associated with reduced odds of a strong preference (t = -3.75, p < .001). More specifically, the odds of a higher preference for AP coursetakers was .78 times as high as non-AP coursetakers. 11th grade GPA was also associated with the preference, where higher GPAs reduced the odds of indicating a stronger preference (t = -3.64, p < .001). Two other academic variables, PSAT and PLAN testtaking (t = -2.03, p < .05) and math self-efficacy (t = -2.29, p < .05) were negatively associated with the preference, although the odds ratios reported were not as impactful as the other academic characteristics—the odds of a stronger preference were .93 for both variables. Academic characteristics seem to have mixed impacts on the preference, though the directionality of the impacts themselves are uniform: more academically successful and college ready students (i.e. those with higher GPAs and higher college-readiness indicators) are less inclined to consider living close to home important when thinking about college.

Of the four college-relevant social capital indicators—parents, friends, teachers, and counselors—only one was significantly associated with the preference to live close to home. Students who talked about college with their friends in 9th grade were .85 times as likely to have stronger preferences than those who did not (t = -2.5, p < .05). In other words, students who discussed college with their friends early in high school were less likely to consider living close to home important when attending college. The finding is important in that college discussions with peers seems important to whether they find living close to home important even when compared to discussions with parents and school personnel, but the analyses is not sufficient for explaining why or how this association takes place. Both the cultural capital measures, educational expectations and the school motivation scale, were significantly associated with the preference. Higher educational expectations reduced the odds of having a strong preference (t = -4.71, p < .001), while higher scores on the school motivation scale increased the odds of having a strong preference (t = 5.87, p < .001). Recall that the school motivation scale in HSLS is a construct of how motivated students are to attend school

and was developed based on measures such as absenteeism and tardiness; higher scores on the scale indicate higher school motivation as defined by the survey. Therefore, the results suggest that higher school motivation is associated with 1.2 times higher odds of having a strong preference.

Neither of the three family characteristics—household size, family income, or the parents' college expectations for the student in the fall of 2013, were significantly associated with the preference. Most notable of these is that parental expectations were not associated in one way or another of a preference to live close to home—it is possible that when expressing their expectation for college going to their students, it is presented agnostic of their expectation of where the student is going to live. While family income was not significantly associated either, it may be due to the inclusion of SES as part of the model; both family income and SES were strongly correlated with one another for the sample ($R^2 = .7$, $p < .001$). Despite this, family income was included partly to avoid manipulating the model after formulating it based on the literature and college choice frameworks, and partly due to SES being a more comprehensive indicator of social class than family income alone.

Lastly, of the school-level variables, only percent free lunch eligible students ($t = 2.92$, $p < .01$) were significantly associated with the preference. The results indicate that higher rates of free lunch eligible students correspond with 1.05 times higher odds of a stronger preference. Interestingly, the college fairs indicator approached significance in this model ($t = -1.92$, $p = .056$); it is possible that other models may find that conducting college fairs at the high school are significantly associated with the preference—specifically, should those models find similar results to this one, the presence of college fairs at the high school would correlate with lower odds of a strong preference.

Preferring to Live Close to Home and College Enrollment

Research question 2 examined the correlation between preferring to live close to home for college and college enrollment by the fall of 2013. Table 2.4 contains the results of the logistic regression models for research question 2. These models measured the association between the preference to remain close to home and college enrollment when controlling for other characteristics identified in the college choice literature. These variables mirror those in research question 1 as well as research question 3. The full model significantly

predicted the variance in college enrollment, $F(40, 450) = 26.22$, $p < .001$. The covariate of interest was the preference to remain close to home, as research question 2 sought to explain the association between the preference and college enrollment. As table 2.4 demonstrates, every model was significantly predictive of the variance in college enrollment, and the impact of the preference on college enrollment changed across the models as more covariates were introduced. In the first model, where the preference was the only covariate, stronger preferences for living close to home was negatively associated with college enrollment in the fall of 2013, with the not at all important category serving as the reference group. Both the somewhat important ($t = -2.63$, $p < .01$) and very important ($t = -7.59$, $p < .001$) categories were significantly associated with college enrollment, with odds ratios of .79 and .47, respectively. However, as more covariates were introduced, both the degree to which they impacted college enrollment and their statistical significance diminished. In the full model, neither the somewhat important category ($t = 0.21$, $p = 0.837$) nor the very important category ($t = -1.33$, $p = 0.184$) were significantly associated with college enrollment when controlling for other background and demographic characteristics, academics, social and cultural capital, family characteristics, and school characteristics.

Table 2.4: Logistic Regression Results for Variables Associated with College Enrollment, by Model

	Logistic Regression Model											
	Model 1 (Preference only)		Model 2		Model 3		Model 4		Model 5		Model 6 (Full model)	
Covariates	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**
Preference to remain close to home (ref = not at all important)												
Somewhat important	0.788**	(0.071)	0.918	(0.084)	1.018	(0.098)	1.013	(0.098)	0.995	(0.099)	1.02	(0.1)
Very important	0.471***	(0.047)	0.639***	(0.063)	0.806*	(0.085)	0.817	(0.087)	0.82	(0.088)	0.866	(0.094)
Background Characteristics												
Race (ref = White)												
Black or African American			1.0	(0.137)	1.204	(0.189)	1.089	(0.17)	1.124	(0.181)	1.248	(0.212)
Asian			2.48***	(0.617)	1.745*	(0.433)	1.872*	(0.464)	1.914**	(0.468)	1.76*	(0.409)
Native Hawaiian/Pacific Islander			1.413	(1.268)	1.181	(0.809)	1.178	(0.814)	1.31	(0.799)	1.147	(0.852)
American Indian/Alaska Native			0.734	(0.361)	0.824	(0.392)	0.723	(0.299)	0.807	(0.307)	0.859	(0.326)
Multiracial			0.923	(0.132)	0.998	(0.163)	1.017	(0.166)	1.08	(0.186)	1.096	(0.183)
Hispanic or Latinx status			1.268	(0.206)	1.397*	(0.221)	1.457*	(0.254)	1.346	(0.254)	1.407	(0.251)
Race and Hispanic Interaction Effects (ref = non-Hispanic)												
Black or African American, Hispanic			2.052	(1.173)	2.224	(1.342)	2.792	(1.832)	3.954	(3.138)	3.379	(2.531)
Asian, Hispanic			0.134**	(0.084)	0.174**	(0.115)	0.168**	(0.112)	0.177**	(0.118)	0.19*	(0.126)
Native Hawaiian/Pacific Islander, Hispanic			0.7	(0.702)	0.817	(0.65)	0.693	(0.562)	0.629	(0.476)	0.79	(0.681)
American Indian/Alaska Native, Hispanic			0.949	(0.558)	0.932	(0.54)	0.99	(0.518)	0.854	(0.43)	0.822	(0.41)
Multiracial, Hispanic			1.399	(0.508)	1.358	(0.523)	1.252	(0.511)	1.212	(0.491)	1.077	(0.404)

See notes at end of table.

Table 2.4: Logistic Regression Results for Variables Associated with College Enrollment, by Model—Continued

	Logistic regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariate	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**
Background Characteristics												
First-generation college student			0.832	(0.093)	0.784*	(0.09)	0.784*	(0.095)	0.705**	(0.087)	0.718**	(0.088)
Number of high schools attended			0.428***	(0.098)	0.638	(0.163)	0.637	(0.163)	0.625	(0.16)	0.666	(0.168)
Socioeconomic status			2.739***	(0.222)	2.084***	(0.177)	1.955***	(0.179)	1.5**	(0.179)	1.388**	(0.167)
Age			0.579***	(0.348)	0.705***	(0.043)	0.723***	(0.044)	0.738***	(0.046)	0.758***	(0.047)
Academic Characteristics												
11th grade GPA					1.599***	(0.062)	1.462***	(0.054)	1.428	(0.05)	1.435***	(0.05)
Any AP courses taken					2.301***	(0.22)	1.944***	(0.193)	1.89***	(0.188)	1.955***	(0.199)
College exam preparation					1.399***	(0.134)	1.274*	(0.122)	1.221*	(0.117)	1.265*	(0.122)
PSAT/PLAN tests taken					1.258***	(0.061)	1.205***	(0.058)	1.196***	(0.06)	1.173**	(0.06)
SAT/ACT tests taken					1.121	(0.745)	1.093	(0.07)	1.115	(0.072)	1.092	(0.073)
Math self-efficacy					1.083	(0.045)	1.012	(0.046)	1.023	(0.048)	1.011	(0.047)
Science self-efficacy					1.044	(0.047)	0.975	(0.046)	0.961	(0.049)	0.97	(0.47)
Social/Cultural Capital												
Talked to parents about college: 9th grade							1.295*	(0.136)	1.233*	(0.129)	1.205	(0.127)
Talked to friends about college: 9th grade							1.064	(0.095)	1.065	(0.095)	1.049	(0.094)
Talked to teachers about college: 9th grade							0.845	(0.101)	0.832	(0.101)	0.864	(0.108)
Talked to counselors about college: 9th grade							1.102	(0.111)	1.134	(0.121)	1.173	(0.123)
School motivation scale							1.245***	(0.588)	1.226***	(0.058)	1.238***	(0.06)
Educational expectations scale							1.314***	(0.037)	1.273***	(0.036)	1.273***	(0.037)

See notes at end of table.

Table 2.4: Logistic Regression Results for Variables Associated with College Enrollment, by Model—Continued

	Logistic regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**
Family Characteristics									0.993	(0.031)	0.995	(0.032)
Household size									1.091**	(0.03)	1.078**	(0.03)
Family income (categorical)									1.091**	(0.03)	1.078**	(0.03)
Parental expectations: most important for student to attend college in fall 2013									2.546***	(0.312)	2.502***	(0.305)
School Characteristics												
Private school											1.579*	(0.35)
School locale (ref = City)												
Suburb											0.865	(0.112)
Town											0.956	(0.153)
Rural											0.797	(0.11)
Percent free lunch eligible students											0.907***	(0.205)
School had college fairs											1.125	(0.19)
School had college information sessions											0.854	(0.144)
School assisted with college applications											0.766	(0.156)
School provided access to college information											1.748	(0.635)
School assisted with selecting colleges											1.626	(0.6)
State (fixed effect)											0.994*	(0.003)

See notes at end of table.

Table 2.4: Logistic Regression Results for Variables Associated with College Enrollment, by Model—Continued

	Logistic regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**	**OR**	**SD**
Constant	4.576***	(0.34)	254845.5***	(291998.7)	668.56***	(790.91)	241.67***	(228.24)	76.51***	(93.45)	43.115**	(55.91)
F-Statistic	*F*(0, 490) = 31.79***		*F*(20, 480) = 34.31***		*F*(20, 470) = 39.36***		*F*(30, 460) = 39.17***		*F*(30, 460) = 34.49***		*F*(40, 450) = 26.22***	

* $p < .05$, ** $p < .01$, *** $p < .001$

NOTE: Degrees of freedom rounded to the nearest 10. OR = Odds ratio. Ref = Reference group. SD = Standard deviation.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

In terms of the other covariates, several variables were significantly associated with college enrollment. For example, Asian students ($t = 2.44$, $p < .05$) were more likely to have enrolled in college than White students net of other characteristics. First generation college students were less likely to have attended college ($t = -2.71$, $p < .01$). SES was positively associated with college enrollment ($t = 2.72$, $p < .01$) while age was negatively associated with college enrollment ($t = -4.49$, $p < .001$). several academic variables were positively associated with college enrollment, including 11th grade GPA ($t = 10.31$, $p < .001$), AP coursetaking ($t = 6.59$, $p < .001$), college exam preparation ($t = 2.44$, $p < .05$) and PSAT/PLAN testtaking ($t = 3.14$, $p < .01$). Of these, AP coursetaking was the most impactful, increasing the odds of college enrollment by about double. While no social capital variable was significant in the full model, both school motivation ($t = 4.41$, $p < .001$) and educational expectations ($t = 8.4$, $p < .001$) were strongly and positively correlated with college enrollment, increasing the odds of enrollment 1.24 and 1.27 times, respectively. Family income ($t = 2.66$, $p < .01$) and parental expectations for college-going ($t = 7.53$, $p < .001$) were both positively associated with college enrollment. Having parents who identified college-going as the most important thing their student could do by the fall of 2013 was associated with 2.5 times increase in the odds of college enrollment, suggesting that parental expectations is largely important for college-going behaviors. At the school level, the only variables that were significantly associated with college enrollment was school control ($t = 2.06$, $p < .05$) and percent free lunch ($t = -4.34$, $p < .001$).

The Association Between Preference and Proximity

The last research question explored the connection between the preference to remain close to home for college and the distance of the college students ultimately enrolled in. Table 2.5 consists the results of the OLS regression models that examined these associations. As in the previous two research questions, all of the models tested significantly predicted the variance of the dependent variable, including the full model, $F(50, 440) = 16.00$, $p < .001$. In each model, the preference to remain close to home was significantly and negatively associated with the log distance of enrolled colleges. As more controls were added to subsequent models, both the somewhat important category and the very important category significantly impacted how far students decided to enroll in postsecondary education. Net of all other characteristics considered for this

study, considering it somewhat important to live close to home for college was associated with a .551 reduction in the log of distance—this is the equivalent of a 55 percent decrease in distance to enrolled college compared to those who considered it not at all important ($t = -6.58$, $p < .001$). More dramatically, considering living close to home very important was associated with a 78 percent decrease in distance compared to the distance when considering it not at all important ($t = -9.73$, $p < .001$).

Table 2.5. Ordinary Least Squares Regression Results for Variables Associated with the Log-Transformed Distance between High School and College, by Model

	OLS regression model											
	Model 1 (Preference only)		Model 2		Model 3		Model 4		Model 5		Model 6 (Full Model)	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Preference to remain close to home (ref = not at all important)												
Somewhat important	-0.71***	(0.104)	-0.62***	(0.092)	-0.58***	(0.086)	-0.56***	(0.084)	-0.56***	(0.084)	-0.55***	(0.084)
Very important	-1.11***	(0.089)	-0.94***	(0.081)	-0.83***	(0.789)	-0.79***	(0.079)	-0.79***	(0.079)	-0.78***	(0.08)
Background Characteristics												
Race (ref = White)												
Black or African American			-0.931	(0.169)	0.005	(0.156)	-0.015	(0.157)	0.032	(0.152)	0.158	(0.138)
Asian			-0.321	(0.175)	-0.48**	(0.171)	-0.47**	(0.179)	-0.457*	(0.185)	-0.199	(0.167)
Native Hawaiian/Pacific Islander			-0.493	(0.335)	-0.589	(0.315)	-0.64*	(0.322)	-0.616	(0.33)	-0.597	(0.315)
American Indian/Alaska Native			0.553	(0.492)	0.659	(0.501)	0.613	(0.483)	0.647	(0.488)	0.721	(0.477)
Multiracial			-0.269	(0.199)	-0.188	(0.192)	-0.17	(0.179)	-0.156	(0.173)	0.005	(0.139)
Hispanic or Latinx status			-0.282*	(0.118)	-0.249*	(0.117)	-0.263*	(0.117)	-0.249*	(0.117)	-0.066	(0.126)
Race and Hispanic Interaction Effects (ref = non-Hispanic)												
Black or African American, Hispanic			0.417	(0.518)	0.609	(0.486)	0.693	(0.521)	0.647	(0.503)	0.453	(0.404)
Asian, Hispanic			-1.019	(0.576)	-0.685	(0.554)	-0.635	(0.544)	-0.658	(0.536)	-0.949	(0.497)
Native Hawaiian/Pacific Islander, Hispanic			-0.488	(0.499)	-0.352	(0.48)	-0.284	(0.477)	-0.297	(0.482)	-0.215	(0.455)

Table 2.5. Ordinary Least Squares Regression Results for Variables Associated with the Log-Transformed Distance between High School and College, by Model—Continued

	OLS regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Background Characteristics												
American Indian/Alaska Native, Hispanic			-0.874	(0.551)	-0.865	(0.551)	-0.809	(0.535)	-0.842	(0.538)	-0.951	(0.528)
Multiracial, Hispanic			-0.167	(0.323)	-0.183	(0.307)	-0.22	(0.341)	-0.217	(0.322)	-0.225	(0.314)
First-generation college student			0.068	(0.078)	0.045	(0.074)	0.032	(0.073)	-0.015	(0.073)	-0.028	(0.074)
Number of high schools attended			-0.076	(0.191)	0.169	(0.19)	0.138	(0.186)	0.152	(0.184)	0.134	(0.173)
Socioeconomic status			0.609***	(0.06)	0.454***	(0.059)	0.437***	(0.057)	0.306***	(0.067)	0.289***	(0.07)
Age			-0.095	(0.064)	-0.042	(0.065)	-0.036	(0.064)	-0.035	(0.065)	-0.048	(0.062)
Academic Characteristics												
11th grade GPA					0.211***	(0.033)	0.197***	(0.034)	0.198***	(0.034)	0.164***	(0.036)
Any AP courses taken					0.402***	(0.06)	0.359***	(0.061)	0.361***	(0.061)	0.441***	(0.061)
College exam preparation					0.182*	(0.08)	0.176*	(0.078)	0.17*	(0.074)	0.155*	(0.066)
PSAT/PLAN tests taken					0.109*	(0.045)	0.102*	(0.043)	0.098*	(0.043)	0.11**	(0.036)
SAT/ACT tests taken					0.116**	(0.037)	0.112**	(0.036)	0.109**	(0.036)	0.102**	(0.035)
Math self-efficacy					0	(0.037)	-0.004	(0.037)	-0.003	(0.036)	0	(0.034)
Science self-efficacy					0.042	(0.041)	0.031	(0.038)	0.033	(0.036)	0.029	(0.034)
Social/Cultural Capital												
Talked to parents about college: 9th grade							-0.059	(0.084)	-0.065	(0.087)	-0.063	(0.082)
Talked to friends about college: 9th grade							0.124*	(0.055)	0.126*	(0.055)	0.111*	(0.055)
Talked to teachers about college: 9th grade							0.148	(0.095)	0.139	(0.093)	0.151	(0.087)

Table 2.5. Ordinary Least Squares Regression Results for Variables Associated with the Log-Transformed Distance between High School and College, by Model—Continued

	OLS regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Social/Cultural Capital												
Talked to counselors about college: 9th grade							-0.17*	(0.084)	-0.158	(0.082)	-0.141	(0.08)
School motivation scale							-0.028	(0.034)	-0.027	(0.034)	-0.04	(0.034)
Educational expectations scale							0.073***	(0.021)	0.073***	(0.02)	0.075***	(0.02)
Family Characteristics												
Household size									-0.02	(0.017)	-0.02	(0.016)
Family income (categorical)									0.055**	(0.018)	0.052**	(0.018)
Parental expectations: most important for student to attend college in fall 2013									-0.041	(0.191)	-0.08	(0.161)
School Characteristics												
Private school											0.444**	(0.137)
School locale (ref = City)												
Suburb											0.417**	(0.132)
Town											0.684***	(0.184)
Rural											0.81***	(0.137)
Percent free lunch eligible students											-0.014	(0.021)
School had college fairs											-0.234	(0.126)
School had college information sessions											0.052	(0.119)
School assisted with college applications											-0.02	(0.25)
School provided access to college information											0.221	(0.449)

Table 2.5. Ordinary Least Squares Regression Results for Variables Associated with the Log-Transformed Distance between High School and College, by Model—Continued

	OLS regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
School characteristics												
School assisted with selecting colleges											0.036	(0.345)
Number of colleges in same commuting zone											0.0002	(0.001)
State (fixed effect)											0.002	(0.003)
Constant	4.027***	(0.083)	5.757***	(1.19)	3.449**	(1.208)	3.172**	(1.189)	3.107*	(1.27)	2.834*	(1.304)
R^2	0.053		0.133		0.18		0.187		0.189		0.221	
F-Statistic	*F*(#, 490) = 78.68***		*F*(20, 470) = 22.32***		*F*(20, 460) = 21.43***		*F*(30, 460) = 18.28***		*F*(30, 460) = 16.68***		*F*(50, 440) = 16.00***	

* $p < .05$, ** $p < .01$, *** $p < .001$

Rounds to zero.

NOTE: Degrees of freedom rounded to the nearest 10. Coeff. = Coefficient. Ref = Reference group. SD = Standard deviation.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

Of the background characteristics, only SES was significantly associated with distance ($t = 4.13$, $p < .001$). SES was positively correlated with distance, indicating that students from higher SES families are more likely to attend schools at farther distances than those from lower SES families. None of the racial and ethnic categories, nor the interaction terms between race and Hispanidad, were significant in the model. On the other hand, all of the academic characteristic variables other than the math and science self-efficacy measures were significantly and positively associated with distance. With a one unit increase of GPA, there was a 16.4 percent increase in distance ($t = 4.56$, $p < .001$). Students who took at least one AP course were associated with a 44 percent increase in distance over students who did not ($t = 7.27$, $p < .001$). Partaking in college exam prep was associated with a 15.5 percent increase in distance ($t = 2.35$, $p < .05$), while both PSAT/PLAN ($t = 3.08$, $p < .01$) and SAT/ACT testtaking ($t = 2.92$, $p < .01$) were associated with a 11 percent and 10 percent increase in distance, respectively.

Only one social capital and one cultural capital measure significantly correlated with distance to enrolled college. Surprisingly, discussing college with parents was not significantly correlated ($t = -0.76$, $p = .445$). Instead, discussing college with friends was positively associated with distance, contributing an 11.1 percent increase to distance over those who did not, net of other characteristics ($t = 2.01$, $p < .05$). Postsecondary education expectations was also significant in the model, where every unit increase was associated with a 7.5 percent increase in distance ($t = 3.81$, $p < .001$). In other words, students who expected to receive at least some 4-year college education—three units above the reference category—would be associated with a 22.5 percent increase over students in the reference category, who expected a certificate or less or did not know how far they expected to get in postsecondary education. Of the three family income variables, only family income ($t = 2.80$, $p < .01$) was significantly associated with distance to enrolled college. Neither parental expectations nor household size significantly predicted distance.

Lastly, school control and school locale were significantly associated with distance to enrolled college. Attending a private school was associated with a 44.4 percent increase in distance over public school attendance ($t = 3.24$, $p < .01$). Compared city schools, attending suburban schools ($t = 3.15$, $p < .01$), small town schools ($t = 3.72$, $p < .001$), and rural schools ($t = 5.89$, $p < .001$) were all significantly associated with

increased distance to enrolled college—with rural school attendance contributing a 80.9 percent increase in distance over city schools. Considering that cities often have more postsecondary options available and students from areas with fewer options would have to travel to attend college, this finding is consistent with the literature and what would be expected of college choice behaviors. None of the other school characteristics, including the number of postsecondary options within the same commuting zone ($t = 0.36$, $p = .716$), were significantly associated with distance to enrolled college.

Limitations

There are several limitations to this study. First among them is that the study was done using secondary data—as such, the research questions and variables chosen were beholden to the structure and contents of the survey and data sets available for use. Qualitative studies, or quantitative studies using primary data sources, could have provided more targeted data and research foci related to habitus and student preferences for where they will live while in college. Second, due to the structure of the data and data cleaning procedures, there are several characteristics that could be considered important that were not included in the models for these research questions. For example, the preference to live close to home while attending college could be associated with whether their parents suggested they should remain close to home (Turley, 2009). However, that variable was not collected in HSLS:09, and therefore could not be used to model the preference.

Additionally, another limitation is the inability for this study to capture immediate postsecondary outcomes. Retention into the second year of college, for example, could be tied to the preference to remain close to home, with implications on retention for students who hold that preference and ultimately enrolled farther from home. College quality and level of postsecondary enrollment are also of interest, but were not captured here—even though the findings suggest that there is no association between the preference to remain close to home and college enrollment, it may be the case that college quality or type of college enrolled is associated. Lastly, this study focused on an analytic population of traditional-aged students. While this population is important to study, these particular research questions are also poignant for older students

and students with dependents. Future studies should consider examining these questions with older student populations or prospective college students who have been out of the educational system for a time.

Discussion

This study focused on whether the preference to remain close to home while attending college, a component of habitus within the college choice process. The research questions were tailored to examine the factors that may be tied to the importance students placed on living close to home, as well as the importance of the preference both for college enrollment decisions and for how far students choose to travel for a college education. Regarding the factors associated with the preference, there were several important characteristics that were significant in the models. For example, Hispanidad was associated with stronger preferences to remain close to home and suggested a complex relationship between Hispanic or Latinx ethnicities and several racial categories. While being a Hispanic or Latinx student alone was predictive of the strength of the preference, several racial subgroups interacted with Hispanidad in a meaningful way. For example, Black Hispanic students were almost twice as likely to have a strong preference to remain close to home than their Black non-Hispanic peers—who themselves were not significantly different than their other non-Latinx peers from other racial groups. Additionally, Native Hawaiian and Pacific Islander non-Hispanic students were 2.7 times more likely to have stronger preferences than White non-Hispanic students, but Native Hawaiian and Pacific Islander Hispanic students were 5.59 times less likely to hold a stronger preference than their Native Hawaiian and Pacific Islander non-Hispanic peers. The evidence suggests that there are dynamic and intricate racial and ethnic factors at play when students develop a preference to live close to home when considering college—there are likely cultural characteristics that intermingle and intersect in ways that are beyond the scope of this paper to examine, but are worthy of future study. Interestingly, there were few factors tied to the preference that are within the means of school administrators or families to change. Beyond the traditional college-readiness metrics tied to college enrollment—GPA, rigorous coursetaking, and postsecondary testtaking—schools interested in impacting student preferences related to where they live for college would do well to focus on building up math self-efficacy, postsecondary educational expectations, and

facilitating peer networks for discussing college (which were all negatively correlated with the preference), as well as driving up school motivation (which was positively associated with the preference).

However, there is evidence from the second research question to suggest that it may not be all that important for schools or families to consider impacting the preference if the primary goal for these stakeholders is college enrollment in general. The analyses indicate that there was no statistical association between college enrollment and the preference to remain close to home when controlling for demographics, academics, and social and cultural capital. In the full model, which included family and school characteristics as well, the preference remained nonsignificant. While this signifies that the hypothesis for research question 2 was not supported by the evidence, it is a promising finding when considering the implications on college choice. These findings suggest that, for individuals or organizations interested in driving up college enrollment, the preference to remain close to home is not an important factor to consider when developing interventions or college-relevant curriculum

Where the preference does matter, as expected, is in how far college enrolled students choose to travel for their college education. When controlling for a battery of factors related to college choice, the preference to remain close to home continued to be associated with distance of enrolled college. However, it is important to recall that the distance measure used for this study was log transformed in order to normalize the data when interpreting the results. While students who considered it very important to live closer to home had a 78 percent reduction in the distance they traveled to attend college, most students overall attended college very close to where they attended high school. Therefore, that 78 percent reduction may not constitute much of a meaningful change over students who had low levels of the preference.

REFERENCES

Alm, J., & Winters, J. V. (2009). Distance and intrastate college student migration. *Economics of Education Review*, *28*(6), 728–738. https://doi.org/10.1016/j.econedurev.2009.06.008

Attewell, P., & Domina, T. (2008). Raising the bar: Curricular intensity and academic performance. *Educational Evaluation and Policy Analysis*, *30*(1), 51-71.

Bailey, M., & Dynarski, S. (2011). *Gains and gaps: Changing inequality in U.S. college entry and completion* (No. w17633). National Bureau of Economic Research. https://doi.org/10.3386/w17633

Bleemer, Z., Brown, M., Lee, D., & Van der Klaauw, W. (2014). Debt, jobs, or housing: What's keeping millennials at home? *FRB of New York Staff Report*, *700*.

Byun, S., Meece, J. L., Irvin, M. J., & Hutchins, B. C. (2012). The role of social capital in educational aspirations of rural youth. *Rural Sociology*, *77*(3), 355–379. https://doi.org/10.1111/j.1549-0831.2012.00086.x

Cabrera, A. F., & La Nasa, S. M. (2000). Understanding the college-choice process. *New Directions for Institutional Research*, *2000*(107), 5–22. https://doi.org/10.1002/ir.10701

Castleman, B. L., & Page, L. C. (2014). A trickle or a torrent? Understanding the extent of summer "melt" among college-intending high school graduates. *Social Science Quarterly*, *95*(1), 202–220. https://doi.org/10.1111/ssqu.12032

Cerna, O. S., Pérez, P. A., & Sáenz, V. (2009). Examining the precollege attributes and values of Latina/o bachelor's degree attainers. *Journal of Hispanic Higher Education*, *8*(2), 130–157. https://doi.org/10.1177/1538192708330239

Cho, S.-J., Hudley, C., Lee, S., Barry, L., & Kelly, M. (2008). Roles of gender, race, and SES in the college choice process among first-generation and nonfirst-generation students. *Journal of Diversity in Higher Education*, *1*(2), 95–107. https://doi.org/10.1037/1938-8926.1.2.95

Conley, D. T. (2008). Rethinking college readiness. *New directions for higher education*, *2008*(144), 3-13.

Desmond, M., & Turley, R. N. L. (2009). The role of familism in explaining the Hispanic-White college application gap. *Social Problems*, *56*(2), 311–334. https://doi.org/10.1525/sp.2009.56.2.311

Duprey, M. A., Pratt, D. J., Jewell, D. M., Cominole, M. B., Fritch, L. B., Ritchie, E. A., ... & Wilson, D. H. (2018). *High School Longitudinal Study of 2009 (HSLS: 09) Base-Year to Second Follow-Up Data File Documentation (NCES 2018-140)*. National Center for Education Statistics, Institute of Education Sciences, US Department of Education. Washington, DC.

Engberg, M. E., & Wolniak, G. C. (2010). Examining the effects of high school contexts on postsecondary enrollment. *Research in Higher Education*, *51*(2), 132–153. https://doi.org/10.1007/s11162-009-9150-y

Frenette, M. (2004). Access to college and university: Does distance to school matter? *Canadian Public Policy/Analyse de Politiques*, *30*(4), 427. https://doi.org/10.2307/3552523

Fullerton, A. S. (2009). A conceptual framework for ordered logistic regression models. *Sociological methods & research*, *38*(2), 306-347.

Goldrick-Rab, S., & Pfeffer, F. T. (2009). Beyond access: Explaining socioeconomic differences in college transfer. *Sociology of Education*, *82*(2), 101–125. https://doi.org/10.1177/003804070908200201

Hamrick, F. A., Florence A., & Stage, F. K. (2004). College predisposition at high-minority enrollment, low-income schools. *The Review of Higher Education*, *27*(2), 151–168. https://doi.org/10.1353/rhe.2003.0058

Hillman, N. W. (2016). Geography of college opportunity: The case of education deserts. *American Educational Research Journal*, *53*(4), 987–1021. https://doi.org/10.3102/0002831216653204

Hillman, N. W. (2017). Geospatial analysis in higher education research. In M. B. Paulsen (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 32, pp. 529–575). Springer International Publishing. https://doi.org/10.1007/978-3-319-48983-4_11

Hillman, N., & Weichman, T. (2016). Education deserts: The continued significance of "place" in the twenty-first century. *Viewpoints: Voices from the Field.*

Horn, L., Kojaku, L. K., & Carroll, C. D. (2001). High school academic curriculum and the persistence path through college: Persistence and transfer behavior of undergraduates 3 years after entering 4-year institutions (NCES 2001-163). *National Center for Education Statistics.*

Hossler, D., Braxton, J., & Coopersmith, G. (1989). Understanding student choice. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 5, pp. 231–281). Agathion Press.

Ingels, S. J., Pratt, D. J., Herget, D. R., Bryan, M., Fritch, L. B., Ottem, R., Rogers, J. E., & Wilson, D. (2015). *High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript Data File Documentation* (NCES 2015-036). National Center for Education Statistics.

Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., Rogers, J. E., Jin, Y., & Leinwand, S. (2011). *High School Longitudinal Study of 2009 (HSLS: 09): Base-Year Data File Documentation.* (NCES 2011-328). National Center for Education Statistics.

Ingels, S. J., Pratt, D. J., Herget, D. R., Dever, J. A., Fritch, L. B., Ottem, R., Rogers, J. E., Kitmitto, S., & Leinwand, S. (2013). *High School Longitudinal Study of 2009 (HSLS:09) Base Year to First Follow-Up Data File Documentation.*

Keefe, S. (1984). Real and ideal extended familism among Mexican Americans and Anglo Americans: On the meaning of "close" family ties. *Human Organization*, *43*(1), 65–70. https://doi.org/10.17730/humo.43.1.y5546831728vn6kp

Klasik, D., Blagg, K., & Pekor, Z. (2018). Out of the education desert: How limited local college options are associated with inequity in postsecondary opportunities. *Social Sciences*, *7*(9), 165.

Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, *9*(8). https://doi.org/10.18637/jss.v009.i08

Massé, J. C., Perez, R. J., & Posselt, J. R. (2010). Revisiting college predisposition: Integrating sociological and psychological perspectives on inequality. *Equity & Excellence in Education*, *43*(3), 279-293.

Mayhew, M. J., Rockenbach, A. N., Bowman, N. A., Seifert, T. A. D., Wolniak, G. C., Pascarella, E. T., & Terenzini, P. T. (2016). *How college affects students: 21st century evidence that higher education works* (Vol. 3). Jossey-Bass. http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118462688.html

Mixon, F. G. (1992). Factors affecting college student migration across states. *International Journal of Manpower*, *13*(1), 25–32. https://doi.org/10.1108/EUM0000000000900

Niu, S. X. (2015). Leaving home state for college: Differences by race/ethnicity and parental education. *Research in Higher Education*, *56*(4), 325–359. https://doi.org/10.1007/s11162-014-9350-y

Nora, A. (2004). The role of habitus and cultural capital in choosing a college, transitioning from high school to higher education, and persisting in college among minority and nonminority Students. *Journal of Hispanic Higher Education*, *3*(2), 180–208. https://doi.org/10.1177/1538192704263189

Paulsen, M. B., & St. John, E. P. (2002). Social class and college costs: Examining the financial nexus between college choice and persistence. *Journal of Higher Education*, *73*(2), 189–236. https://doi.org/10.1080/00221546.2002.11777141

Perez, P. A., & McDonough, P. M. (2008). Understanding Latina and Latino college choice: A social capital and chain migration analysis. *Journal of Hispanic Higher Education*, *7*(3), 249–265. https://doi.org/10.1177/1538192708317620

Perna, L. W. (2000). Differences in the decision to attend college among African Americans, Hispanics, and Whites. *Journal of Higher Education*, *71*(2), 117–141. https://doi.org/10.1080/00221546.2000.11778831

Perna, L. W. (2006). Studying college access and choice: A proposed conceptual model. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 21, pp. 99–157). Springer Netherlands. https://doi.org/10.1007/1-4020-4512-3_3

Perna, L. W., & Titus, M. A. (2005). The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences. *Journal of Higher Education*, *76*(5), 485–518. https://doi.org/10.1080/00221546.2005.11772296

Ramirez, N. A., Lacy, A., Duprey, M., & Jones, A. (2019). NCES sample surveys: A practical primer for new users, practitioners, and policymakers. *New Directions for Institutional Research*, *2019*(181), 73-90.

Roderick, M., Coca, V., & Nagaoka, J. (2011). Potholes on the road to college: High school effects in shaping urban students' participation in college application, four-year college enrollment, and college match. *Sociology of Education*, *84*(3), 178–211. https://doi.org/10.1177/0038040711411280

Sá, C., Florax, R. J., & Rietveld, P. (2004). Determinants of the regional demand for higher education in the Netherlands: A gravity model approach. *Regional Studies*, *38*(4), 375–392. https://doi.org/10.1080/03434002000213905

Schwartz, S. J. (2007). The applicability of familism to diverse ethnic groups: A preliminary study. *The Journal of Social Psychology*, *147*(2), 101–118. https://doi.org/10.3200/SOCP.147.2.101-118

Smith, J., Pender, M., & Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review*, *32*, 247–261. https://doi.org/10.1016/j.econedurev.2012.11.001

Tierney, W. G., & Sablan, J. R. (2014). Examining college readiness. *American Behavioral Scientist*, *58*(8), 943–946. https://doi.org/10.1177/0002764213515228

Tuckman, H. P. (1970). Determinants of college student migration. *Southern Economic Journal*, 184-189.

Turley, R. N. L. (2006). When parents want children to stay home for college. *Research in Higher Education*, *47*(7), 823–846. https://doi.org/10.1007/s11162-006-9017-4

Turley, R. N. L. (2009). College proximity: Mapping access to opportunity. *Sociology of Education*, *82*(2), 126–146. https://doi.org/10.1177/003804070908200202

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, *40*(1), 7-20.

CHAPTER 3: GOING THE DISTANCE: DESCRIBING THE GEOGRAPHIC MOBILITY OF FIRST-TIME BEGINNING COLLEGE STUDENTS

Introduction

It is largely accepted in the study of college access that most students make decisions on where to attend college based in part on the costs and benefits of postsecondary education. Taking a traditional human capital perspective on college choice, it is assumed that potential students make "investment decisions" implicitly by weighing factors like availability of financial aid and financial resources, foregone earnings, and the earning differential between being college educated and not—or when deciding on specific colleges, the differential between education at one college over another (Paulsen, 2001). Conceptual extensions to the traditional theoretical frameworks incorporate multidisciplinary factors into this decision-making process, such as social capital and parental expectations (Cabrera & La Nasa, 2000; Perna, 2000, 2006). The current body of work in the area of college choice, however, treats this decision-making process as a-geographical; the importance of geography in the decision-making process, whether in terms of proximity to nearby colleges, region of the country, or a desire to relocate or migrate for college, has not been the focus of much research. It is the latter of these topics that is likely salient in the minds of all college students considering whether to go to college and where, as the available options for postsecondary education rely heavily upon whether there are options within a close enough proximity given their preferences to relocate. It is therefore imperative to understand more about the geographic mobility patterns of college students, and what characteristics are tied to the spectrum of geographic mobility for college.

The literature that does exist on geographic mobility for college tends to focus on migration patterns across the country (Cooke & Boyle, 2011), interstate migration (Cooke & Boyle, 2011; González Canché, 2017, 2018; Mixon, 1992), or on the importance of distance to college options on attendance patterns (Frenette, 2004; Hillman, 2016; Turley, 2009). This body of research addresses geographic concerns at a

macro level, seeking to understand or describe the migration patterns of large swaths of students. For example, researchers have gleamed that educational outcomes for a given state, like college attendance, are spatially dependent on what other states share a border (Cooke & Boyle, 2011; González Canché, 2018). Similarly, we know that proximity to college options matter in the sense that students are more likely to attend college if they have postsecondary options within a close proximity (Frenette, 2004; Hillman, 2016). Less attention has been paid, however, on describing individual student characteristics and their associations with migration patterns. This study contributes to the literature by integrating the various research findings on geography and college access in order to better understand geographically mobile and immobile first-time college students.

Literature Review

Mobility Patterns of College Students

The extant literature as a whole suggests that geographic location matters (Cullinan & Duggan, 2016; Frenette, 2004; González Canché, 2017; Hillman, 2016; Hillman & Weichman, 2016; Klasik, Blagg, & Pekor, 2018; Mixon, 1992; Mixon & Hsing, 1994; Turley, 2009). Early interstate mobility research operating through a human capital lens suggested that higher in-state tuition (Mixon, 1992; Tuckman, 1970), lower in-state college selectivity, and lower per capita income within state (Mixon, 1992) are associated with increased rates of interstate outmigration, meaning that these metrics were related to whether students chose to attend college out of their home state. Later research used more and better parameters to test migration patterns associated with state-level conditions. A study of college choice sets measured by where students sent their SAT scores found different rates of in-state and out-of-state college considerations by race and ethnicity—White students send scores to and enroll out-of-state more than Black and Latinx students, but Black students end up more likely to send scores and enroll out-of-state after controlling for academic preparation, state fixed effects, and parental education than White students (Niu, 2015). The same study also found that parental education levels were positively associated with outmigration, and that Black, Latinx, and academically high-achieving students were more likely to enter more selective institutions if they left the state than if they remained in-state (Niu, 2015).

However, there are differential effects based on the state you come from and the state you are going to. The distance between your home state and destination state, whether the states are adjacent to one another, urbanicity of the destination, and per capita income at both the home state and destination state are all associated with whether students flow from one state to another (Cooke & Boyle, 2011). Lastly, studies of intrastate mobility suggest that increased population numbers at the destination (Tuckman, 1970), closer proximity to state colleges, and higher high school degree levels of Hispanic students in the home school district (Alm & Winters, 2009) increased the odds of attending an in-state school, while increased levels of educational expenditures in the home district decreased those odds (Alm & Winters, 2009). Interestingly, there is evidence that institutions try to attract students from specific states as a function of different tuition rates—private institutions tend to pull students from wealthier states than public institutions (González Canché, 2017).

Nontraditional Students and College Frameworks

The three-stage model of college choice (Cabrera & La Nasa, 2000), which serves as a foundation for many other conceptual frameworks, was conceptualized in accordance with the college choice process of traditional students. The college choice process of nontraditional students—those who are adult learners or those in nontraditional modes of study—is thus insufficiently captured common theoretical and conceptual approaches to the study of access, transition, and match (Paulsen & St. John, 2002; Perna, 2006). Some frameworks, such as Perna's (2006) multilayered college choice framework, are constructed particularly to contextualize group differences in the decision-making process; however, many of these frameworks do not account for the fact that nontraditional students often seek nontraditional modes of study, like nondegree-seeking vocational programs and distance-learning programs, or lack sufficient financial support for traditional enrollment, and thus face barriers to access above and beyond that which traditional students and even other marginalized communities face (Schuetze & Slowey, 2002). The U.S. higher education system is fortunately a leader in incorporating institutional factors friendly to nontraditional students compared to other nations (Schuetze & Slowey, 2002), but these students have not been sufficiently explored under common frameworks.

Defining Local Geographies

One of the first steps to studying geography of opportunity is defining the geographic unit of analysis that will be the focus of the research task. One of the simplest ways of operationalizing a geographic unit of analysis in a U.S-centric setting is by examining large political boundaries like states and regions. Analysis of college access related to states often look at interstate student mobility (Cooke & Boyle, 2011; Mixon, 1992; Mixon & Hsing, 1994; Niu, 2015), though there have been studies on intrastate mobility as well (Alm & Winters, 2009) and mobility across regions as well (Turley, 2009). Regions or other geographic units similar to U.S. states have been used in international studies of student mobility (Sá, Florax, & Rietveld,2004). States, regions, and similar large political areas are simple to use for special analyses typically because of the (relatively) small number of geographic units to analyze. In the U.S. context, analyses of states are useful to researchers beyond their simplicity due to the differences in policy contexts between each state—for example, analyses of state merit aid programs can examine which types of state level policies are more effective at retaining or attracting students than others (Cooke & Boyle, 2011). State-level analyses can also be easy to execute without diving into network analyses or GIS. However, there is plenty of nuance that cannot be easily captured at a macro level like state—for example, local economic conditions and proximal postsecondary institutions.

Smaller and more granular geographic units of analysis include core based statistical areas (CBSA), commuting zones (CZ), and census tracts. These are geographic areas defined by government agencies—usually the Census Bureau. CBSAs are defined as geographic areas with a populous "core" and adjacent communities with a "high degree of economic and social integration with that core. (U.S. Census Bureau, 2012b)" Commuting zones are clusters of counties that "share similar labor markets and economic activities…where people live, work, and commute. (Hillman, 2016, p. 996)" Census tracts are small subdivisions with a population size between 1,200 and 8,000 that is updated every decennial census (U.S. Census Bureau, 2012a). These smaller units of analysis, while more difficult to use than states or regions, allow for very granular data analysis that allows for diversity within state.

Research Questions

This research study serves to extend the literature beyond questions of access related to traditional students; by using a sample of nationally representative first-time beginning (FTB) college students, this exploratory study examines the factors that are related to attending colleges closer or farther away from home. The two guiding research questions are:

1) What demographic, familial, academic, financial, geographic, and institutional factors are associated with the distance between a student's enrolled college and their home? and;
2) What notable geographic characteristics related to the college choices of traditional and nontraditional students are present in the data?

Research question one is a more traditional research question in that it is an empirical question that is examined using traditional quantitative methods—ordinary least squares (OLS) regression. Using college choice factors identified in the literature, the analyses that address research question one seeks to illuminate what characteristics are associated with the distance between home and college. Given the more diverse set of beginning college students in the sample compared to a sample of students matriculating to college at a traditional age, the findings will be useful for the college choice literature to see what specifically is connected with traveling away from home for college. On the other hand, the second research question is more exploratory—after processing the data and deriving new variables related to geography, traditional and nontraditional students will be compared to identify patterns and distributions in where they attend to college. This study will identify nontraditional students as students who began college at age 20 or older, students who began college at least a year after graduating high school, and students who attended a private, for-profit college. These criteria were chosen to coincide with definitions of nontraditional related to age and mode of education. Due to the structure of the variables included in these analyses, the sample will include FTB students who attended only one institution in 2011-2012; as such, all findings from this report is only generalizable to that population of students.

Data

Data Source and Description

This study utilized Beginning Postsecondary Students study of 2017 (BPS:12/17), a longitudinal postsecondary study looking at first-time beginning students in 2012 (Bryan, Cooney, & Elliot, 2019). Unlike HSLS:09, BPS:12/17 tracks students who have already enrolled in college and follows them over time. NCES's postsecondary student studies are complimentary in a way that is not the case with its secondary studies—there is one cross-sectional study named the National Postsecondary Student Aid Study (NPSAS), and two longitudinal studies: BPS and Baccalaureate and Beyond (B&B). After a NPSAS data collection, a subsample of the cross-sectional cohort is followed over several years; in the case of BPS, there are follow-ups 2- and 5-years after the NPSAS data collection (Hill et al., 2016). In other words, BPS:12/17 follows a cohort of FTB students sampled for the 2011-12 National Postsecondary Student Aid Study (NPSAS:12).

BPS is focused first and foremost with postsecondary academic experiences, persistence and attainment (Hill et al., 2016). However, BPS:12/17 contains a diverse array of data covering many useful domains of postsecondary research. For example, because it is a follow-up to a subsample of NPSAS:12, which is focused on how students finance college, there are useful financial aid and college financing variables included from the base year data collection. Each iteration of BPS is concerned with the college experiences of students who enter college for the first time, their retention and persistence, educational attainment, and some early college outcomes related to employment, wages, and family formation—as such, it is less suited for studies of college access and choice. However, what it does offer is a nationally representative sample of FTBs, which includes nontraditional-aged students and students who delay college after high school; I will provide more information on the analytic population in a subsection below.

As mentioned, this iteration of BPS began with NPSAS:12, which samples a wider range of students in college during the 2011-2012 academic year. Because it is a longitudinal study, there are three data collection periods for BPS: the base-year NPSAS data collection (NPSAS conducted in 2011-2012, the first follow-up (BPS:12/14) conducted in 2013-2014, and the second follow-up conducted in 2016-2017 (Bryan et

al., 2019). BPS:12/14 is the most current data release of this iteration of BPS; as of this proposal, data from the second follow up, BPS:12/17, has yet to be released.

The core component of NPSAS:12, and by extension BPS:12/14, is the student survey. In the base year, students are surveyed about their high school characteristics, major, current labor characteristics like employment and salaries, financial aid sources, and other topics (Wine et al., 2014). There are seven topics of data collection in the survey: enrollment, education experiences, financial aid, current employment, income and expenses, background, and locating (only administered to FTBs so they could be located for follow up studies) (Wine et al., 2014). The survey is administered via web and telephone, with an option of an abbreviated interview for refusals. In the base year, the FTB response rate was 60% (Wine et al., 2014). Data collected from the survey is then matched to administrative data from a variety of sources, including the Central Processing System (CPS) for federal financial aid data, the National Student Loan Data System (NSLDS), National Student Clearinghouse (NSC), ACT, and College Board for SAT scores (Wine et al., 2014). BPS:12/17 followed a similar structure of data collection as NPSAS:12, containing the same administrative data sources and survey topics (Bryan et al., 2019)

Several other sources of data were used alongside BPS:12/17. Shapefiles containing data from the U.S. Census were used to combine students and their institutions with geographic identifiers. Shapefiles for census tracts, CZs, counties, and states were all used to generate geographic variables for analyses. These data sources were combined in order to utilize different units of geography to examine student mobility. Data from IPEDS that were included in the BPS:12/17 RUF as part of the source files were also included to get additional institutional characteristics and longitude and latitude data. Lastly, a file of IPEDS institutions from 2014 was used to calculate the number of institutions contained within each commuting zone and census tract.

Sample

NPSAS:12 and BPS:12/17 are generalizable to the national population of students at Title IV eligible postsecondary institutions during the 2011-2012 academic year, with BPS:12/17 being generalizable to students who enrolled in college at these institutions for the first time in 2011-2012 (Bryan et al., 2019; Wine

et al., 2014). The base year sample was generated with a two-stage stratified random sample with probability proportional to size (PPS), with the first stage of sampling consisting of an institutional sampling frame, and the second consisting of the student sampling frame (Wine et al., 2014). In the base year, there were about 85,000 (69% unweighted) respondents out of an eligible respondent sample of about 123,600—of these, about 37,170 were sampled as FTBs, about 35,540 were eligible to complete the study, and about 24,770 responded to the survey, generating a weighted response rate of 67.6% (Hill et al., 2016). Given the sampling design, as well as data processing procedures explained later in the article, this study is only generalizable to FTBs who entered postsecondary education in the 2011-2012 academic year and who only attended one institution in that year. The analytic sample used for this study included BPS:12/17 respondents that had complete geographic data—that is, had complete data in the GIS source file included in the BPS:12/17 RUF—as well as complete institutional data from the NPSAS:12 Institution source file. Using these criteria, there were approximately 18,440 students included in the analytic sample.

Research Design

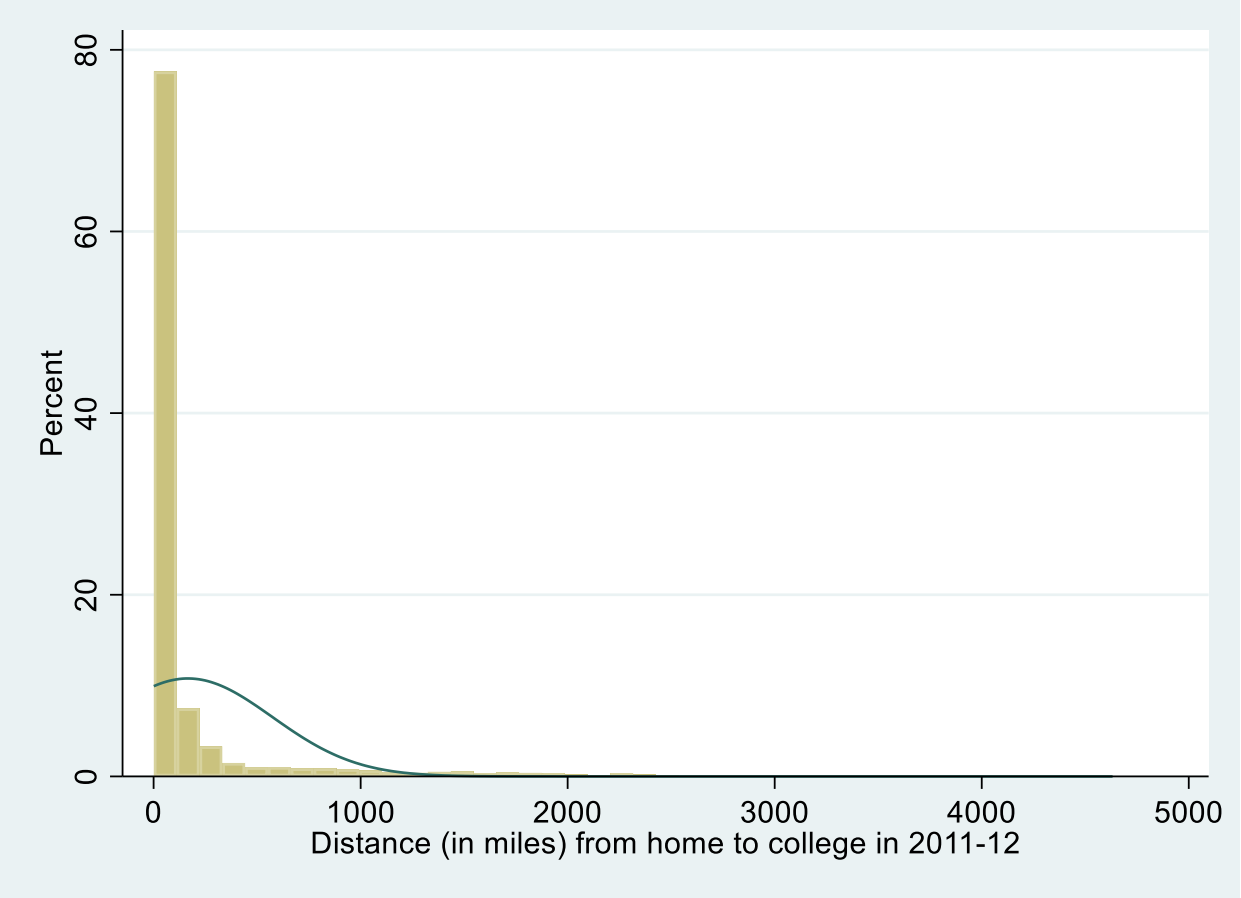
Methods

This study relied on a mixture of methodologies to answer its research questions. The first research question, which examined the various factors associated with the distance between a student's home and their first college, incorporated OLS regression models. The full model, which included all of the chosen predictors drawn from the research literature, is as follows:

$$log(DISTANCE_i) = \beta_0 + \beta_1 D_i + \beta_2 FAM_i + \beta_3 FIN_i + \beta_4 A_i + \beta_5 I_i + \beta_6 G_i + \varepsilon_i \quad (4)$$

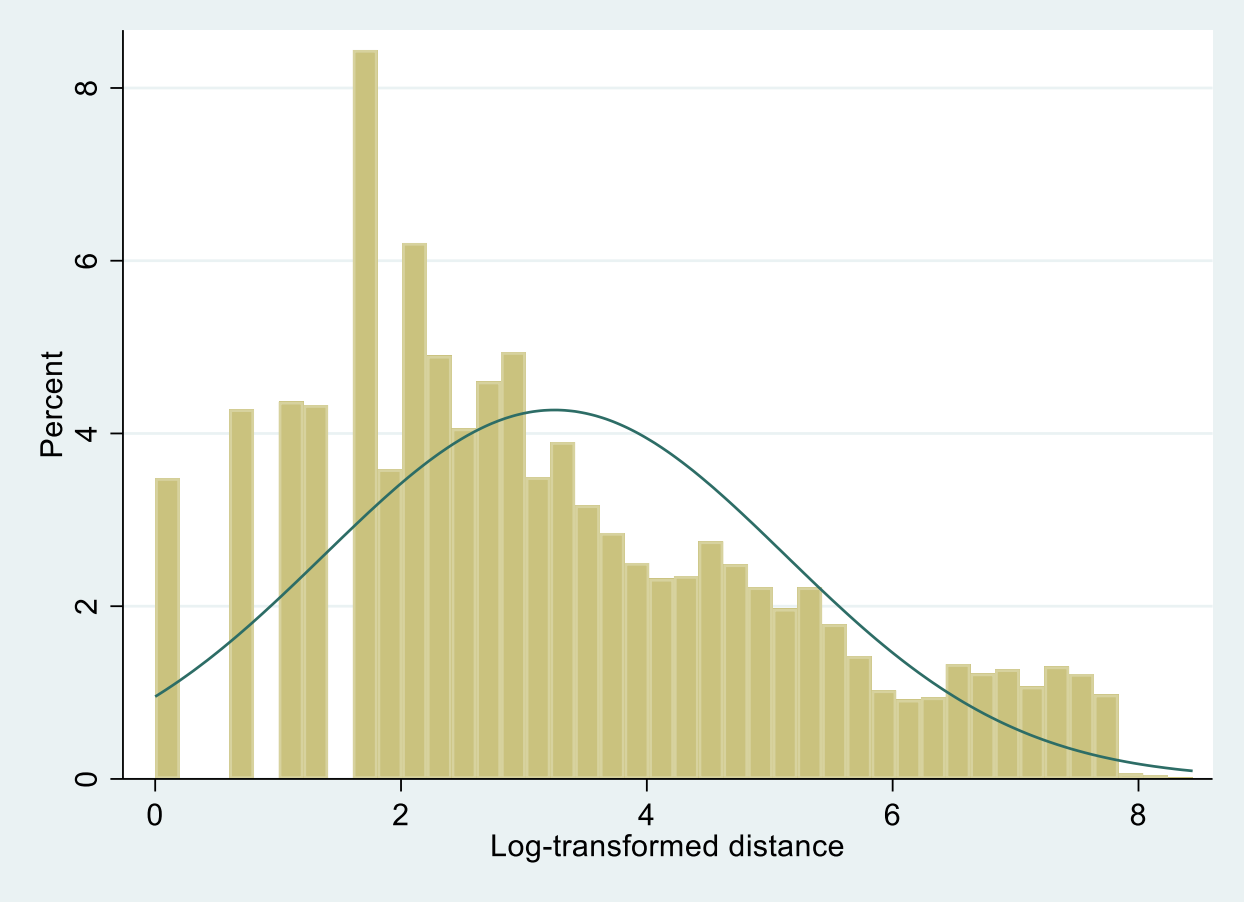
The dependent variable, *DISTANCE*, refers to the geodesic distance in miles between the permanent residence of student *i* and the institution they attended in academic year 2011-2012. The distance was included as a variable in BPS:12/17 and permanent residence in the base year was determined by the survey. Due to the skewed nature of the distance (see figure 3.1) measure, distance was log-transformed to force a normal distribution. The distribution of the log-transformed distance measure was more normalized, as shown in figure 3.2.

Figure 3.1: Distribution of Distance from Home to College in 2011-12, in Miles



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

Figure 3.2: Distribution of Log-Transformed Distance from Home to College in 2011-12, in Miles



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

The model predictors are as follows: β_0 represents the constant; *D* is a vector of demographic characteristics; *FAM* is a vector of family characteristics; *FIN* is a vector of income, finances, and financial aid variables; *A* is a vector of academic and programmatic variables; *I* is a vector of institution-level variables; and *G* is a vector of geographic characteristics. ε_i signifies observational error.

Research question two involved comparing traditional and nontraditional students regarding the geographic characteristics of their college choices. To examine this, several statistical methods were used. First, the analytic sample members were categorized among three criteria: age in which they entered college, whether they entered college immediately after graduating high school, and whether they attended a private for-profit college. Students were compared within criteria after defining groupings within-criteria as

traditional or nontraditional. In the first, students who entered college between the ages of 17 and 19 were considered traditional, while older or younger students were considered nontraditional. In the second, students who attended college immediately after graduating high school were considered traditional while those who did not were considered nontraditional. Finally, students who attended colleges other than private for-profit colleges were considered traditional, and those who attended private for-profit colleges were considered nontraditional.

Descriptive statistics, *t*-tests and chi-squared analyses were utilized to compare each traditional-nontraditional pairing. Descriptive statistics outlined the differences between both groups for most of the college choice factors examined for research question one. *T*-tests were utilized to compare mean distances between home and college and log-transformed distance for each traditional-nontraditional pair. Lastly, chi-squared analyses were used to test the significant differences between each traditional-nontraditional pair and the mobility patterns related to their college choices—the distribution of those who attended college in the same commuting zone, a neighboring commuting zone within the same state, a non-neighboring commuting zone in the same state, a neighboring state, or a non-neighboring state. These analyses allow for a deeper exploration into the differences between traditional and nontraditional student college choice besides interstate-instrastate differences and differences in distance alone.

Variables

All of the variables used for this study are related to the base year enrollment in the 2011-12 academic year. Although data concerning postsecondary education in later years was used, BPS:12/17 was chosen over NPSAS:12 for several reasons First, the data set includes updated data for variables that may have been missing or previously imputed, but administrative data matches after NPSAS:12 or interview data in BPS:12/17 provided valid responses. Second, the data are already subsetted to FTB students in 2011-2012, requiring less data cleaning efforts to have the sample represent the population I wish to analyze. Third, some derived variables used in the model (e.g. DISABLE and ENINPT1, discussed below) were not a part of the original NPSAS:12 derived variables file. Lastly, BPS:12/17 is unique in providing a GIS source file, matching

sampled students to their census tract. This allows for the inclusion of geographic data more specific to each students' home communities than using states or counties.

Dependent Variables

The dependent variable for both research questions is the distance between students' permanent home prior to attending college and first institution attended. I intend on deriving this variable from location data on the BPS:12/17 RUF and IPEDS. BPS:12/17 contains a derived variable called DISTANCE that is the distance in miles between students' permanent residence and their first institution at the base year; this variable is useful in that the distance was calculated using their home address, which is inaccessible through the RUF. Additionally, as mentioned previously, the distance variable was log-transformed in order to make the dependent variable skew closer to normality. To supplement the analysis of distance categories for research question 2, indicator variables were derived flagging cases where the institution attended was within the same or a neighboring geographic unit of analysis—CZ or state.

Predictor Variables

There are six groups of predictor variables. The first group, demographic characteristics, contains seven variables identified in the literature as important to college choice. These variables include sex (GENDER), Census-defined race (RACECEN), Hispanic or Latinx ethnicity status (HISPANIC), first-generation college student status (derived from PAREDUC), an indicator for whether a sibling attended college before the respondent (SIBCOLFT), immigration generation status (IMMIGEN), an indicator for a collection of long-lasting disabilities (DISABLE), and respondent age in 2011-12 (AGE). The last three variables are ones that are not typically included in college choice studies. Immigration generation status was included to examine whether it might be negatively correlated with distance; as an example of why immigration generation status might be related to college choice, consider that familism as a cultural value is associated with both college enrollment and immigrant generational status (Desmond & Turley, 2009; Tseng, 2004). Similarly, a disability as derived in BPS:12/17—defined as a long-lasting condition like hearing impairments, vision impairments, difficulty with attention, and others—could be negatively associated with distance. Lastly, age is included as a predictor primarily because the diverse age population of BPS:12/17 is

well suited to examine the impact of being a traditional- or nontraditional-age college student on the distance of enrolled college.

The second variable group contain variables related to family characteristics. These variables include students' marital status (SMARITAL), dependency status (DEPEND), the number of dependents under the student (DEPNUM), and family size (HSIZE). These variables are included in the model to predict the role that familial context has on the distance of the chosen institution, as well as the magnitude of the association between the context and distance.

The third group contains variables related to financial aid and economic capital. While other groups may contain financial variables or variables related to financial resources (for example, average in-state tuition and fees for institutions), this group is for financial aid and capital specifically tied to the respondent. This group includes the amount of student loan debt through Direct Subsidized and Unsubsidized loans (STFY12), adjusted gross income in the base year (CAGI), total federal Title IV aid received (TITIVAMT), an indicator for whether the student applied for any financial aid (AIDAPP), and the amount of veterans' benefits received (VETBENY12). It also includes financial variables related to the institution, including the amount of institutional need-based grants received (INSTNEED), institutional non-need-based grants received (INSTNOND), institutional tuition and fee waivers received (INSWAIV), and net price after grants as a percentage of income (EFFORT3). All of the continuous variables in this group—essentially every variable except for any aid applications and net price as a percentage of income—were transformed to represent the dollar amount in thousands in order to better interpret regression results during analysis.

The fourth group contains academic and programmatic variables. These variables are associated with college coursework, attendance patterns, and high school graduation. The four variables in this group include attendance intensity in the base year (ENINPT1), an indicator for any online, night, or weekend classes taken (ALTANY), an indicator for whether the student took remedial courses (REMETOOK), and an indicator for whether the student graduated high school in 2011-2012 (RCNTGRAD). These variables, like age, serve to introduce predictors related to nontraditional students—specifically, students from nontraditional programs, attendance patterns, and those that do not enroll in college immediately after high school.

The next group of variables contained institutional characteristics. The variables in this group corresponded to the institution they attended in 2011-12. These variables include a categorical variable identifying the institution level (LEVEL), an identifier for whether the institution was public or private (CONTROL), a measure of institutional selectivity (SELECTV2), and the degree of urbanicity of the institution (LOCALE). The group also contained two financial variables related to the average in-state tuition (TUITION2) and fees (FEES2) at the institution. Like the variables in the financial aid and economic capital group, these two variables were transformed to represent the dollar amount in thousands.

The final group of variables were geographic variables. These variables serve both as controls for other predictors, descriptors of the local geography, and to signify the number of postsecondary options within a close distance. These variables included an indicator variable for the urbanicity of the student's permanent address (LOCALEST), the number of 4-year institutions in the student's CZ, the number of 2-year institutions in the student's CZ, and a vector of variables related to the student's census tract. These variables included the median household income (MED_HH_INC) transformed to represent the dollar amount in the thousands, the percentage of White residents (PER_WHITE), the percentage of Black residents (PER_BLACK), the percentage of Asian residents (PER_ASIAN), the percentage of American Indian or Alaska Native residents (PER_AIAN), the percentage of Bachelor's degree holders or higher levels of education (PER_BACHELORS_PLUS) and the percentage of unemployed residents (PER_UNEMPLOYED). Further details about these variables and how they were derived are listed in the following section.

Data Processing

The analytic data used for this study was processed in order to incorporate all of the necessary variables, either through merging with external data sources or to derive new variables using existing data. Several data files from the BPS:12/17 RUF were combined—the Derived file, the School Information file, the NPSAS:12 Institution file, and the GIS file. The Derived file contains the majority of variables needed for modeling and statistical analysis. The School Information file was used to combine the students in the file with a unique institution identifier for the first institution attended in 2011-12. The NPSAS:12 Institution

uses the unique institution identifier to attach the student rows with institution-level variables related to the first institution attended. Lastly, the GIS file was used to merge a geographic identifier tying students to the census tracts their home address was in during the base year. The GIS file also contains basic descriptive information about the census tracts.

As the smallest geographic unit used for this analysis, the census tracts from the GIS file were used to pull CZ, county, and state identifiers and merge them onto the student file. A shapefile containing all of the census tracts was generated using the U.S. Census TIGER/Line database. Fifty-two shapefiles containing census tracts for all 50 states, Washington D.C. and Puerto Rico were consolidated into a single shapefile, which was then merged with shapefiles containing CZs, counties, and states based on the location of the census tract polygons using the WGS 84 Coordinate Reference System (EPSG:4326). Census tract polygons located within the CZ, county, and state polygons were considered matches and given the identifiers for their polygons. Once all of the geographic units were tied together, the identifiers were merged into the student file. Lastly, the IPEDS institution file was used to calculate the number of postsecondary institutions overall and from each level—four-year, two-year, and less-than-two-year—in each CZ.

Weighting and Subpopulation Considerations

Unlike HSLS:09, BPS:12/17 has only one cross-sectional analytic weight (WTA000). The reason for a lack of analytic weights to choose is that, unlike HSLS:09, missing data due to item and unit nonresponse are imputed using weighted sequential hot deck imputation (Hill et al., 2016). Weighted sequential hot deck is a method of imputation that finds respondents similar to nonrespondents along a set of variables, selects a donor among those respondents, and replaces the missing value with the donor's value (Hill et al., 2016). In other words, there is less need for weights to account for nonresponse in BPS:12/17.

To properly estimate variance given the complex sampling design, balance repeated replication (BRR) was used alongside the svyset command in Stata 16. BPS:12/17 includes replicate weights and sampling design variables in order to properly estimate variance. The data contain an analytic, cross-sectional weight (WTA000) used for this study as well as the corresponding replicate weights (WTA001-WTA200). BRR is

one of two approaches the BPS:12/17 data file documentation report suggests for variance estimation (Bryan, Cooney, & Elliot, 2019).

Lastly, the subpop option of svyset was used to subset the analytic sample based on the criteria listed herein without dropping respondents from the data file and, consequently, their weights. As mentioned previously, students were included in the study if they attended one institution in 2011-12 (STUDMULT = 1) and had a nonmissing census tract ID that allowed for matching to U.S. Census TIGER/Line shapefiles for geographic variable derivation. Subsetting onto respondents who attended one institution is necessary due to many of the financial aid and institutional variables that required for the model only applied to students who attended one institution. The subpopulation used for this study included 16,150 respondents (rounded to the nearest ten), or 71.7% of the total BPS:12/17 cross-section.

Results

Descriptive Statistics

The unweighted and weighted frequency statistics results for the analytic variables used in research question one are presented in table 3.1. The statistics described in this section will refer to the weighted percentages, means, and standard deviations unless otherwise specified.

Table 3.1: Frequency Statistics for Analytic Variables

Variable	N	Unweighted percent	Weighted percent
Race (Census Categories)			
White	10,730	66.47	68.9
Black or African American	3,020	18.72	16.38
Asian	800	4.97	5.89
American Indian or Alaska Native	480	2.97	2.57
Native Hawaiian/other Pacific Islander	280	1.73	1.49
More than one Race	830	5.14	4.77
Hispanic or Latino origin			
Non-Hispanic or Latino	12,820	79.37	81.06
Hispanic or Latino	3,330	20.63	18.94
Gender			
Male	7,050	43.65	44.79
Female	9,100	56.35	55.21

See notes at end of table.

Table 3.1: Frequency Statistics for Analytic Variables—Continued

Variable	N	Unweighted percent	Weighted percent
Disability Status: Has long-lasting disability or condition in 2011-12			
No	14,260	88.31	88.79
Yes	1,890	11.69	11.21
First-generation college student			
No	9,920	61.43	66.4
Yes	6,230	38.57	33.6
Had sibling who attended college first			
No	8,760	54.24	52.28
Yes	7,390	45.76	47.72
Immigrant generation status			
First generation immigrant	1,170	7.24	8.01
Second generation immigrant	3,200	19.8	19.46
Third generation immigrant or higher	11,780	72.96	72.53
Marital status in 2011-12			
Single, divorced, separated, or widowed	15,120	93.63	94.74
Married	1,030	6.37	5.26
Dependency status in 2011-12			
Dependent student	11,950	74.03	80.63
Independent student	4,190	25.97	19.37
Applied for any financial aid in 2011-12			
No	940	5.8	10.7
Yes	15,210	94.2	89.3
Attendance intensity pattern in 2011-12			
Full-time	11,500	71.2	65.31
Part-time or mixed	4,650	28.8	34.69
Took online, night, or weekend classes at first institution			
No	9,740	60.3	63.26
Yes	6,410	39.7	36.74
Took remedial courses in 2011-12			
No	12,310	76.2	76.22
Yes	3,840	23.8	23.78
Recent (2011) high school graduate			
No	5,410	33.5	26.06
Yes	10,740	66.5	73.94
Urbanicity of student's permanent address			
City	5,880	36.39	35.56
Suburb	7,120	44.07	44.87
Town	1,240	7.67	7.57

See notes at end of table.

Table 3.1: Frequency Statistics for Analytic Variables—Continued

Variable	N	Unweighted percent	Weighted percent
Rural	1,920	11.87	12.01
Institution level			
4-year	10,190	63.09	56.24
At least 2-year but less than 4-year	5,960	36.89	43.73
Less-than-2-year	#	0.02	0.03
Institution control			
Public	8,400	52.02	74.84
Private, not-for-profit	3,110	19.26	17.4
Private, for-profit	4,640	28.72	7.76
Institution selectivity			
Open admission or not 4-year institution	8,200	50.79	53.65
Minimally selective	3,090	19.12	7.8
Moderately selective	2,750	17.03	23.3
Very selective	2,110	13.05	15.24
Institution urbanicity			
City	8,990	55.7	54.93
Suburb	4,110	25.44	24.7
Town	1,500	9.26	11.57
Rural	1,550	9.6	8.8

Rounds to zero.
NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

As the dependent variable, distance, was not categorical, the variables included in table 6 were all predictors in the model used for research question one. As noted in table 6, the majority of the sample were White non-Hispanic female students. About 11.2 percent of the sample identified themselves as having a long-lasting disability or condition in the 2011-12 academic year, 33.6 percent were first-generation college students, and a combined 27.47 percent were either first- or second-generation immigrant. It is notable that few members of the sample were married—roughly 5.3 percent of the analytic sample—while the rest were single, divorced, separated, or widowed. Similarly, few were independent students—about 19.4 percent. There were several other characteristics that stood out from the frequency distributions; The most numerous

category of student urbanicity was suburban with 44.87 percent, less than ten students attended less-than-2-year institutions in the analytic sample (.03 percent), and the majority of students attended institutions located in cities.

Table 3.2 shows the summary statistics for the continuous variables examined for research question one. As mentioned in prior sections, the distance variable was heavily skewed, which is evident from table 3.2. The mean distance between home and school for the analytic sample was 145.48 miles; however, the minimum distance between a student's home and their school was one mile, and the max was 4,390 miles. The log-transformation of the distance measure forced the variable to approach normality and allowed for more interpretable analyses later in the study.

Table 3.2: Summary Statistics for Continuous Analytic Variables

Variable	Min	Max	Unweighted Mean	Unweighted SD	Weighted Mean	Weighted SD
Dependent Variable						
Distance from permanent home to college attended in 2011-12	1.00	4,390.00	158.90	401.14	145.48	5.06
Log-transformed distance	0.00	8.39	3.15	1.88	3.13	0.03
Independent Variables						
Age in 2011-12	15.00	75.00	21.06	6.47	20.33	0.15
Number of Student's Dependents	0.00	10.00	0.31	0.87	0.21	0.01
Household Size	1.00	13.00	3.69	1.60	3.81	0.03
Direct subsidized and Unsubsidized loans (in thousands)	0.00	23.54	3.69	4.41	2.27	0.07
Adjusted gross income in 2011-12 (in thousands)	0.00	1,000.00	55.24	66.22	64.58	1.05
Total federal Title IV aid received (in thousands)	0.00	69.50	7.59	7.76	5.60	0.17
Amount of institutional need-based grants received (in thousands)	0.00	49.46	1.13	4.38	1.13	0.05
Amount of institutional non-need-based grants received (in thousands)	0.00	50.00	1.39	4.41	1.37	0.06
Amount of institutional tuition and fee waivers received (in thousands)	0.00	40.00	0.12	1.65	0.13	0.02
Net price after grants as percentage of income	0.00	100.00	47.78	35.80	37.53	0.45
Veterans' benefits amount received (in thousands)	0.00	54.18	0.28	2.72	0.13	0.02
In-state tuition for full-time undergraduates (in thousands)	0.00	44.22	11.92	10.46	9.35	0.14
In-state fees for full-time undergraduates (in thousands)	0.00	10.90	0.68	0.97	0.81	0.03
Census tract: median household income (in thousands)	4.51	250.00	59.73	29.07	62.83	0.45
Census tract: percent White	0.00	100.00	70.35	25.95	72.28	0.42
Census tract: percent Black	0.00	100.00	15.24	23.09	13.41	0.34
Census tract: percent Asian	0.00	92.05	5.25	9.22	5.57	0.16
Census tract: percent American Indian/Alaska Native	0.00	97.19	0.65	2.87	0.70	0.05
Census tract: percent with a Bachelor's degree or higher	0.00	91.57	28.03	17.82	30.27	0.30

See notes at end of table.

Table 3.2: Summary Statistics for Continuous Analytic Variables—Continued

Variable	Min	Max	Unweighted Mean	Unweighted SD	Weighted Mean	Weighted SD
Census tract: percent unemployed	0.00	35.77	6.45	3.41	6.11	0.05
Commuting zone: number of postsecondary institutions	1.00	380.00	93.13	104.70	90.13	2.79
Commuting zone: number of 4-year institutions	0.00	141.00	38.90	41.73	37.61	1.08
Commuting zone: number of 2-year institutions	0.00	108.00	24.09	28.57	23.25	0.76
Commuting zone: number of less-than-2-year institutions	0.00	131.00	30.14	36.59	29.28	0.98

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

There are several other notable characteristics contained in table 3.2. For example, the youngest sample member was 15 years old in 2011-12, while the oldest was 75. The mean age was roughly 20 years old. Financially, the adjusted gross income for the sample members was highly variable, with $64,580 as the mean adjusted gross income with a max of $1,000,000—it is likely that the max was forced to a ceiling of $1,000,000 in the survey. About 78 percent of students in the sample received federal Title IV funds (not displayed in table 3.2), with the maximum funds received totaling $69,500.

The final 11 rows in table 3.2 include geographic characteristics pulled from the GIS source file or derived using a combination of BPS:12/17 data, IPEDS data, and U.S. Census TIGER/Line shapefiles. The average median household income at the census tract level for sample members was approximately $62,830; based on the unweighted mean and standard deviation values, this figure was much less variable than the adjusted gross income measure. The average student in the sample lived in census tracts that were majority White; the lowest racial proportion at the census tract level was American Indian/Alaska Native, with a mean of .7 percent among sample members' census tracts. Roughly 30 percent of the residents of each census tract had at least a Bachelor's degree, while a mean of about 6 percent were unemployed. Lastly, the distribution of postsecondary institutional options across CZs was highly variable. The minimum number of postsecondary institutions across all sectors was one, while the maximum was 380—this was for a CZ identified as being located in California. As shown in table 7, there were students who lived in commuting zones that had no options in specific sectors. 1.8 percent of students lived in CZs that had no 4-year institutions, 1.5 percent lived in CZs with no 2-year institutions, and 6.3 percent lived in CZs with no less-than-2-year institutions (percentages not included in table 3.2).

Regression Outcomes

The results of the OLS regression models are displayed in table 3.3. Six models were analyzed, each progressively adding more variables into the model. The first model included only the dependent variable (log-transformed distance) and the background characteristics. Each successive model added family characteristics, financial and economic capital variables, academic and programmatic characteristics,

institutional characteristics, and geographic characteristics, respectively. The overview of the results below will cover the first and the final model to conserve space.

Table 3.3: Ordinary Least Squares Regression Results Predicting Log-Transformed Distance from Permanent Home to Enrolled College for FTB Students, by Model

	OLS regression model											
	Model 1 (Background only)		Model 2		Model 3		Model 4		Model 5		Model 6 (Full Model)	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Background Characteristics												
Race (ref. = White)												
Black or African American	-0.217**	(0.074)	-0.206**	(0.074)	-0.240***	(0.064)	-0.171**	(0.064)	-0.11	(0.062)	0.002	(0.068)
Asian	-0.005	(0.108)	0.003	(0.106)	-0.076	(0.102)	-0.074	(0.1)	-0.207*	(0.093)	-0.061	(0.094)
American Indian or Alaska Native	-0.294**	(0.096)	-0.297**	(0.099)	-0.165	(0.095)	-0.113	(0.094)	-0.089	(0.096)	#	(0.096)
Native Hawaiian/ other Pacific Islander	-0.348*	(0.171)	-0.363*	(0.169)	-0.328*	(0.165)	-0.264	(0.164)	-0.27	(0.176)	-0.13	(0.167)
More than one race	-0.109	(0.089)	-0.104	(0.088)	-0.165	(0.086)	-0.122	(0.087)	-0.134	(0.075)	-0.069	(0.075)
Hispanic or Latino Origin	-0.466***	(0.068)	-0.468***	(0.067)	-0.308***	(0.061)	-0.234***	(0.064)	-0.198**	(0.059)	-0.094	(0.061)
Gender (ref. = male)	0.125**	(0.044)	0.119**	(0.044)	0.041	(0.04)	0.056	(0.04)	0.033	(0.039)	0.032	(0.039)
Age	-0.013**	(0.004)	-0.001	(0.004)	0.001	(0.004)	0.005	(0.004)	-0.001	(0.004)	#	(0.004)
Disability Status	-0.166*	(0.066)	-0.143*	(0.068)	-0.106	(0.062)	-0.069	(0.059)	-0.044	(0.057)	-0.035	(0.058)
First-generation college student	-0.406***	(0.051)	-0.384***	(0.053)	-0.204***	(0.048)	-0.163***	(0.047)	-0.06	(0.046)	-0.037	(0.044)

See notes at end of table.

Table 3.3: Ordinary Least Squares Regression Results Predicting Log-Transformed Distance from Permanent Home to Enrolled College for FTB Students, by Model—Continued

	OLS regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Had sibling who attended college first	-0.084*	(0.04)	-0.088*	(0.04)	-0.068	(0.038)	-0.069	(0.038)	-0.062	(0.036)	-0.057	(0.036)
Immigration status (ref. = first generation immigrant)												
Second generation immigrant	0.448***	(0.084)	0.439***	(0.084)	0.302***	(0.081)	0.258**	(0.082)	0.220**	(0.077)	0.201**	(0.075)
Third generation immigrant or higher	0.768***	(0.079)	0.762***	(0.079)	0.575***	(0.081)	0.540***	(0.081)	0.455***	(0.073)	0.374***	(0.073)
Family Characteristics												
Marital status: Married			0.116	(0.128)	0.354**	(0.124)	0.382**	(0.121)	0.284*	(0.114)	0.252*	(0.11)
Dependency status: Independent			-0.470***	(0.104)	-0.399***	(0.113)	-0.313**	(0.111)	-0.052	(0.095)	-0.066	(0.091)
Number of student's dependents			0.129**	(0.049)	0.115*	(0.046)	0.104*	(0.047)	0.06	(0.036)	0.06	(0.034)
Household size			0.007	(0.017)	-0.015	(0.016)	-0.018	(0.015)	-0.014	(0.012)	-0.021	(0.012)

See notes at end of table.

Table 3.3: Ordinary Least Squares Regression Results Predicting Log-Transformed Distance from Permanent Home to Enrolled College for FTB Students, by Model—Continued

	OLS regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Financial and Economic Capital												
Direct Subsidized and Unsubsidized loans (in thousands)					0.048***	(0.008)	0.039***	(0.009)	-0.014	(0.008)	-0.016*	(0.008)
Adjusted gross income in 2011-12 (in thousands)					0.006***	(#)	0.006***	(#)	0.003***	(#)	0.003***	(#)
Total federal Title IV aid received (in thousands)					0.035***	(0.004)	0.031***	(0.004)	0.023***	(0.004)	0.025***	(0.004)
Applied for any financial aid in 2011-12					-0.055	(0.083)	-0.127	(0.084)	-0.176*	(0.068)	-0.105	(0.068)
Amount of institutional need-based grants received (in thousands)					0.062***	(0.004)	0.057***	(0.004)	0.024***	(0.005)	0.028***	(0.005)

See notes at end of table.

Table 3.3: Ordinary Least Squares Regression Results Predicting Log-Transformed Distance from Permanent Home to Enrolled College for FTB Students, by Model—Continued

	OLS regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Amount of institutional non-need-based grants received (in thousands)					0.059***	(0.005)	0.051***	(0.005)	0.012	(0.006)	0.015*	(0.007)
Amount of institutional tuition and fee waivers received (in thousands)					-0.052***	(0.012)	-0.050***	(0.011)	-0.034**	(0.012)	-0.034**	(0.011)
Net price after grants as percentage of income					0.008***	(0.001)	0.006***	(0.001)	0.001	(0.001)	0.001	(0.001)
Veterans' benefits amount received (in thousands)					-0.003	(0.011)	-0.015	(0.011)	-0.024*	(0.011)	-0.024*	(0.01)
Academic and Programmatic Characteristics												
Attendance intensity: Part-time or mixed							-0.512***	(0.046)	-0.172***	(0.046)	-0.164***	(0.044)

See notes at end of table.

Table 3.3: Ordinary Least Squares Regression Results Predicting Log-Transformed Distance from Permanent Home to Enrolled College for FTB Students, by Model—Continued

	OLS regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Took online, night, or weekend classes at first institution							0.055	(0.044)	0.073	(0.04)	0.062	(0.041)
Took remedial classes in 2011-12							-0.375***	(0.048)	-0.126**	(0.046)	-0.112*	(0.046)
Recent (2011) high school graduate							0.036	(0.067)	-0.011	(0.062)	-0.05	(0.059)
Institutional Characteristics												
Level (ref. = 4-year)												
2-year									-0.957***	(0.113)	-0.928***	(0.116)
Less-than-2-year									-2.678***	(0.29)	-2.515***	(0.417)
Control: Private									0.598***	(0.071)	0.617***	(0.069)
Selectivity (ref. = Open admission or not 4-year)												
Minimally selective									-0.357*	(0.155)	-0.322*	(0.154)
Moderately selective									0.24	(0.125)	0.228	(0.124)
Very selective									0.611***	(0.146)	0.570***	(0.141)

See notes at end of table.

Table 3.3: Ordinary Least Squares Regression Results Predicting Log-Transformed Distance from Permanent Home to Enrolled College for FTB Students, by Model—Continued

	OLS regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Institution locale (urbanicity)									0.157***	(0.027)	0.119***	(0.027)
In-state tuition (in thousands)									0.001	(0.005)	-0.003	(0.005)
In-state fees (in thousands)									0.054	(0.027)	0.047	(0.029)
Geographic Characteristics												
Home locale (urbanicity)											0.226***	(0.026)
Census tract: Median household income (in thousands)											0.008***	(0.001)
Census tract: Percent White											-0.005*	(0.002)
Census tract: Percent Black											-0.001	(0.003)
Census tract: Percent Asian											-0.010**	(0.003)
Census tract: Percent American Indian/ Alaska Native											-0.002	(0.009)

See notes at end of table.

Table 3.3: Ordinary Least Squares Regression Results Predicting Log-Transformed Distance from Permanent Home to Enrolled College for FTB Students, by Model—Continued

	OLS regression model											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Census tract: Percent of Bachelor's or higher											0.002	(0.002)
Census tract: Percent unemployed											-0.004	(0.007)
Commuting zone: Number of 4-year institutions											-0.006***	(0.001)
Commuting zone: Number of 2-year institutions											0.006**	(0.002)
Constant	3.012***	(0.12)	2.811***	(0.13)	1.834***	(0.147)	2.191***	(0.177)	2.726***	(0.19)	2.539***	(0.31)
R^2	0.06		0.065		0.209		0.232		0.336		0.363	
F-Statistic	*F*(10, 190) = 41.86***		*F*(20, 180) = 38.63***		*F*(30, 170) = 86.67***		*F*(30, 170) = 76.59***		*F*(40, 100) = 107.01***		*F*(50, 90) = 87.4***	

* $p < .05$, ** $p < .01$, *** $p < .001$

Rounds to zero.

NOTE: Degrees of freedom rounded to the nearest 10. Coeff. = Coefficient. Ref. = Reference group. SD = Standard deviation.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

In the first model, many of the background characteristics were significantly associated with log-transformed distance. Compared to White students, every racial category other than Asian and multiracial were negatively associated with distance, with Native Hawaiian/Pacific Islander students having the largest discrepancy—the distance for Native Hawaiian/Pacific Islander students was about 34.8% lower than White students (t = -2.04, p < .05). Hispanic or Latinx status was also negatively associated with distance, with Hispanic or Latinx students showing a 46.6 percent decrease in distance compared to non-Hispanic or Latinx students (t = -6.86, p < .001). Female students were associated with an increase in distance (t = 2.83, p < .01), while older students (t = -3.14, p < .01), disabled students (t = -2.5, p < .05), first-generation college students (t = 7.88, p < .001), and students who had a sibling attend college first (t = -2.11, p < .05) all had significant, negative associations with distance. For the last background characteristic, students who were second generation (t = 5.36, p < .001) or third generation or higher (t = 9.78, p < .001) were associated with higher distances than first generation immigrants—they were associated with a 44.8 percent and 76.8 percent increase in distance over first generation immigrants, respectively.

In the final model, the only background variable to remain significant after controlling for the spectrum of variables described herein was immigration status, with second generation immigrants being associated with a 20.1 percent increase in distance (t = 2.67, p < .01) and third generation immigrants or higher being associated with a 37.4 percent increase in distance (t = 5.11, p < .001) over first generation immigrants. Similarly, the only family characteristic significantly associated with distance was marital status, with married students showing a 25.2 percent increase in distance over nonmarried students (t = 2.29, p < .05). Many of the financial and economic capital variables were significantly associated with distance; the financial and economic capital variables not tied to institution aid with significant associations included federal Direct subsidized and unsubsidized loans (t = -2.11, p < .05), adjusted gross income (t = 6.32, p < .001), total federal Title IV aid received (t = 6.31, p < .001), and veterans benefits received (t = -2.3, p < .05). However, the associated change in distance was relatively low for all of the financial and economic capital variables—the largest difference was in the amount of institutional tuition and fee waivers received, where every thousand dollar increase in waivers was associated with a 3.4 percent decrease in distance (t = -3.02, p <

.01). It is worth noting that the directionality of the associations between the institutional aid received variables and distance are difficult to interpret with these analyses; because we only have data on the institutions they attended and not all of the institutions they considered for enrollment, it is difficult to parse out what the impact is of these financial characteristics in terms of whether it is associated with increases or decreases in distance. What we can safely interpret with these results, however, is that institutional aid factors into the distance between students and their colleges.

For the academic and programmatic variables, two of four were significantly associated with distance. Students who attended college either part-time or mixed intensity were associated with a 16.4 percent decrease in distance compared to only full-time students (t = -3.69, p < .001). Similarly, students who attended remedial classes in their first year of college were associated with an 11.2 percent decrease in distance over those who did not (t = -2.47, p < .05). Institutional characteristics appeared to be highly associated with distance—six of nine variables were significantly tied to distance. The strongest impact on distance came from attending less-than-2-yr institutions, which was associated with a 251.5 percent decrease in distance over those who attended 4-year institutions (t = -6.03, p < .001). Less dramatically, attending a two-year institution was associated with a 92.8 percent decrease (t = -8.01, p < .001). the two variables associated with the strongest increase in distance were attending a private school (t = 8.90, p < .001) and attending a very selective institution (t = 4.03, p < .001), which showed a 61.7 percent and 57 percent increase in distance to college, respectively.

Lastly, over half of the geographic characteristics were significantly associated with distance. The urbanicity of the student's home address was significantly tied to distance, with every unit increase in urban locale (city vs. suburb, town, and rural locales) being associated with a 22.6 percent increase in distance. The number of 4-year and 2-year institutions in the CZ both had conflicting relationships with distance; each unit increase in 4-year college opportunities in the CZ was associated with a .6 percent decrease in distance (t = -4.16, p < .001) while each unit increase in 2-year college opportunities was associated with a .6 percent increase in distance (t = 2.66, p < .01). Median household income at the CZ level was positively associated with distance, where every thousand dollars increase in median income showed a .8 percent increase to

distance (t =7.66, p < .001). The final two significant variables were percentage of White (t = -2.08, p < .05) and percentage of Asian residents (t = -2.9, p < .01) in the CZ, where both negatively impacted distance travelled.

Traditional and Nontraditional Student Comparisons

Descriptive Statistics

To examine differences between traditional and nontraditional students, the student sample was categorized into dichotomous groups based on three criteria, which are referred to here as "traditional-nontraditional pairs." The distribution of the sample based on the three criteria is found in table 3.4. To conserve space, all descriptive statistics reported in this section beyond frequency distributions in table 3.4 are presented as appendices.

Table 3.4: Frequency Distributions for Traditional-Nontraditional Student Pairs, by Traditional-Nontraditional Criteria

Criteria	N	Unweighted percent	Weighted percent
Age when entering college: Age 17–19			
Traditional students	11,570	71.66	79.08
Nontraditional students	4,580	28.34	20.92
College enrollment timing: Recent high school graduates			
Traditional students	10,740	66.5	73.94
Nontraditional students	5,410	33.5	26.06
Institutional control: Private, nonprofit and public			
Traditional students	11,510	71.28	92.24
Nontraditional students	4,640	28.72	7.76

NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

The first criterium involved the age of the student when entering college. Traditional aged students were considered those who were between 17 and 19 when entering college, while those who were not among those ages were considered nontraditional. About 11,570 students in the sample were traditional (79.1 percent) while 4,580 were considered nontraditional (20.9 percent) based on the first criterium. It is worth noting that about 30 students in the sample were younger than 17 years old and considered part of the

nontraditional group—given the vast majority of students in the nontraditional group were older than 19, it is understandable to consider the nontraditional group "older" than the traditional group. The second criterium was college enrollment timing related to high school completion. Those that entered college immediately after graduating high school were referred to as traditional, while those who did not were considered nontraditional. For this traditional-nontraditional pair, about 10,740 students were traditional (73.9 percent) while 5,410 were nontraditional (26.1 percent). The final criterium was institutional control. Students who attended college at an institution that was either public or a private, nonprofit college were considered traditional, while those who attended private, for-profit colleges were considered nontraditional. About 11,510 students were traditional under this pairing, while about 4,640 students were nontraditional. Interestingly, while the distribution of traditional-nontraditional students was roughly similar to those from the first two pairings, the weighted percentage was 92.2 percent traditional and 7.8 percent nontraditional—this is compared to the unweighted 71.3 and 28.7 percent distribution, respectively.

Appendix 3.1 and appendix 3.2 contain the frequency statistic and summary statistic tables for the first traditional-nontraditional pair criterium. As expected, the first traditional-nontraditional pair differed in characteristics related to age—dependency status, marital status, number of dependents, and adjusted gross income. Interestingly, the nontraditional group—who are almost exclusively older than the traditional group—had a lower adjusted gross income than the traditional group. This is due to adjusted gross income being calculated at the household level and may be caused by traditional aged college students having on average wealthier families than older, nontraditional students. Based on the descriptive statistics, it seems that as a group, nontraditional students had lower means for postsecondary opportunities in their CZs. They also lived in CZs with lower rates of Bachelor's or higher degree holders and higher rates of unemployed individuals.

Appendix 3.3 and appendix 3.4 contain the frequency statistic and summary statistic tables for the second traditional-nontraditional pair criterium. Based on the frequency distributions, nontraditional students in the college enrollment timing criterium were more likely to be Black and Hispanic or Latinx. Higher percentages of nontraditional students were also first-generation college students and more likely to have

part-time or mixed enrollment or online, night, or weekend classes. Interestingly, 21.2 percent of nontraditional students in this grouping attended private, for-profit institutions compared to 3.02 percent of traditional students. Nontraditional students, as expected, had a higher mean age and higher mean number of dependents than traditional students (25.79 compared to 18.41, respectively). Similar to the first traditional-nontraditional pair, nontraditional students in this group had lower mean adjusted gross income, higher net price as percentage of income, lower rates of Bachelor's or higher degree holders in their CZ, fewer postsecondary education institutions in their CZ, and higher unemployment rates than traditional students.

Lastly, appendix 3.5 and appendix 3.6 contain the frequency statistic and summary statistic tables for the third traditional-nontraditional pair criterium. The nontraditional group showed larger representation of Black, American Indian or Alaska Native, Native Hawaiian/other Pacific Islander, Multiracial, and Hispanic or Latinx students than the traditional group. Over half of the nontraditional students in this group were first-generation college students (57.4 percent), which was higher than the traditional student group (31.6 percent). On average, nontraditional students were older (about 25 years old). Almost all of the nontraditional students in this group applied for financial aid—99.2 percent of nontraditional students applied for financial aid compared to 88.47 percent of traditional students. These nontraditional students—those who attended a private, for-profit college, were substantially more likely to have not been recent high school graduates when starting college—71.2 percent—compared to traditional students, of whom 22.3 percent were not recent high school graduates. Geographically, this traditional-nontraditional pair was similar to prior ones, having lower proportions of White residents, higher proportions of residents of color, lower Bachelor's or higher degree holders, and higher rates of unemployed residents in their CZs. Additionally, while the rates of 4-year, 2-year and less-than-2-year college options in the CZs of nontraditional students were relatively comparable to those of traditional students, the overall mean of postsecondary options in their CZ was higher (92.1 compared to 89.97).

Traditional-Nontraditional Distance Comparisons

Using the log-transformed distance measure, the three traditional-nontraditional pairs were compared to see how nontraditional status across the three criteria associated with the distance of enrolled college.

Table 3.5 shows the outcome of the *t*-tests comparing the means of traditional and nontraditional students on log-transformed distance.

Table 3.5. *T*-tests for Traditional-Nontraditional Student Differences, by Traditional-Nontraditional Criteria

	Student Type						
	Traditional			Nontraditional			
Criteria	**N**	**M**	**SD**	**N**	**M**	**SD**	***t***
Age when entering college: Age 17-19	11570	3.23	0.03	4580	2.77	0.06	7.12***
College enrollment timing: Recent high school graduates	10740	3.24	0.03	5410	2.83	0.06	6.98***
Institution control: Private, nonprofit and public	11510	3.05	0.03	4640	4.10	0.09	-10.16***

* $p < .05$, ** $p < .01$, *** $p < .001$

M = Mean, SD = Standard deviation

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

The two groups in the first pair, age when entering college, were significantly different from each other, $t(200) = 7.12$, $p < .001$. Students who entered college between the ages of 17 and 19 showed a higher mean distance than those who entered college at younger or older ages. In other words, traditional students using this criterium attended college at farther distances than nontraditional students. The two groups in the second pair, timing of college enrollment, were also significantly different from one another, $t(200) = 6.98$, $p < .001$. Similar to the first pair, students who attended college straight after graduating high school had a higher mean distance than those who waited more than one year after graduating—traditional students once again traveled farther distances to attend college than nontraditional students. The two groups in the third pair, institutional control, differed from the first two in that the difference in means was inverted, $t(200) = 10.16$, $p < .001$. Students who attended private, nonprofit or public colleges showed a lower mean distance than students who attended private, for-profit colleges. As such, there was a significant difference in means between traditional and nontraditional students using this criterium, where traditional students attended college closer to home than nontraditional students.

Traditional-Nontraditional Mobility Patterns

As mentioned in the methods section, the sample was coded along five "mobility pattern categories" corresponding to the relationship between the home and institution location: attended a college in the same CZ; attended a college in a neighboring, in-state CZ; attended a college in a nonneighboring, in-state CZ; attended a college in a neighboring state; and, attended a college in a nonneighboring state. These five categories were chosen arbitrarily based on data exploration. Then, crosstabulations were generated for the three traditional-nontraditional pairs with these geographic pattern categories. These crosstabulations can be found in table 3.6. Lastly, Pearson chi-square tests were conducted on the crosstabulations to see if both variables were significantly associated with each other. Since the Pearson chi-square test, when applied to complex survey data, is automatically corrected to a design-based *F*-statistic in Stata 16, the corrected design-based *F* statistic will be reported below alongside the chi-square coefficient.

Table 3.6: Chi-square coefficient and *F*-statistic for Traditional-Nontraditional Student Differences, by Criteria and Student Type

	Student Type					
	Traditional			Nontraditional		
Criteria	**N**	**Unweighted percent**	**Weighted percent**	**N**	**Unweighted percent**	**Weighted percent**
Age when entering college: Age 17–19						
Same commuting zone	6,830	42.27	57.51	3,170	19.63	72.73
Neighbor commuting zone, same state	1,090	6.76	9.92	250	1.57	5.41
Nonneighbor commuting zone, same state	1,600	9.93	16.04	320	1.97	7.42
Neighbor state	910	5.64	6.95	150	0.9	1.93
Nonneighbor state	1,140	7.07	9.58	690	4.28	12.51
$\chi^2(4) = 611.9$	***F*(#, 640) = 31.73*** **					
College enrollment timing: Recent high school graduates						
Same commuting zone	6,290	38.98	57.28	3,700	22.92	70.38
Neighbor commuting zone, same state	1,030	6.37	9.99	320	1.96	6.1
Nonneighbor commuting zone, same state	1,510	9.33	16.1	420	2.58	8.95
Neighbor state	850	5.25	6.93	210	1.28	2.98
Nonneighbor state	1,060	6.58	9.7	770	4.77	11.6
$\chi^2(4) = 482.25$	***F*(#, 690) = 25.32*** **					
Institutional control: Private, nonprofit and public						
Same commuting zone	7,060	43.69	61.55	2,940	18.21	50.5
Neighbor commuting zone, same state	1,110	6.89	9.42	230	1.43	3.72
Nonneighbor commuting zone, same state	1,530	9.49	14.88	390	2.41	6.58
Neighbor state	780	4.84	5.85	270	1.69	6.44
Nonneighbor state	1,030	6.36	8.3	810	4.99	32.75
$\chi^2(4) = 1115.375$	***F*(#, 720) = 138.72*** **					

* $p < .05$, ** $p < .01$, *** $p < .001$

Rounds to zero.

NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

The first traditional-nontraditional pair was significantly associated with the mobility patterns, $\chi^2(4) = 611.9$, $F(\#, 640) = 31.73$, $p < .001$. The majority of students in both groups lived in the same CZ as their college—however, more nontraditional students lived in the same CZ (72.7 percent) than traditional students (57.5 percent). A cumulative 12.8 percent of nontraditional students attended college in CZs in the same state excluding their own CZ, compared to a cumulative 26 percent of traditional students. When comparing the out-of-state enrollment of these groups, a higher proportion of traditional students attended college in a neighboring state (7 percent) than nontraditional students (1.9 percent), while the inverse was true of nonneighbor states—nontraditional students attended college in a nonneighbor state at higher proportions (12.5 percent versus 9.6 percent).

The second traditional-nontraditional pair was significantly associated as well, $\chi^2(4) = 482.25$, $F(\#, 690) = 25.32$, $p < .001$. The distribution of traditional and nontraditional students in the five mobility pattern categories was similar in directionality and proportions to the distribution of the first pair. Higher proportions of nontraditional students attended college in same commuting zone than traditional students (70.4 percent versus 57.3 percent), as well as in nonneighbor states (11.6 percent versus 9.7 percent). Cumulatively, more traditional students attended out-of-state colleges than nontraditional students as well—16.6 percent and 14.6 percent, respectively.

Lastly, the third traditional-nontraditional pair was significantly associated with the mobility pattern categories, $\chi^2(4) = 1115.375$, $F(\#, 720) = 138.72$, $p < .001$. The distribution for these variables was slightly different than prior ones; a narrow majority of nontraditional students lived in the same CZ than any other category at 50.5 percent. Unlike the first two pairs, 32.8 percent of nontraditional students lived in nonneighbor states—compared to the 8.3 percent of traditional students in the same mobility pattern category, the differences seem substantial. Overall, more nontraditional students in this category were out-of-state students (39.2 percent) compared to traditional students (14.15 percent), signifying that students attending private, for-profit colleges attended out-of-state schools at higher rates than those who enrolled in other types of institutions.

Limitations

The findings from this study are limited by several factors. First, as a secondary data analysis study, the research agenda is limited by the available variables—or possible derivations from the available variables—and thus more specific questions may not be answerable given lack of data. For example, as the student's home address is not available in the data set, different measures for distance between home and college are not possible without approximating based on other geographic units such as CZ, census tract, or county. An important factor that was not included in this analysis was pre-college academic achievement. This factor is typically included in college choice studies but was unavailable for this paper. Second, the methods used for this study were correlational—the findings herein should not be interpreted as causal. Moreover, the methods used for research question two in particular were meant for data exploration; the findings are surface-level and serve only as a stepping-stone to more intricate and meaningful analyses of mobility patterns in the future. Mobility patterns in these data, as well as other data sources, could be spliced in different ways to capture different forms or categories of student mobility. The categories described here were chosen as a negotiation between minimizing the number of categories while still effectively describing more nuanced differences in mobility compared to interstate-intrastate mobility. Other researchers examining similar patterns may choose other, potentially better, mobility categories.

Discussion

The traditional approach to studying college choice often treats geography as an afterthought or does not utilize geography at all. There is a dearth of literature that has examined the factors that impact how far students are willing to travel to attend college, despite distance playing an important role in the types of college students choose to attend (Frenette, 2004; Hillman, 2016; Turley, 2009). Further, few studies have examined or described the mobility patterns of nontraditional college students. This study utilized a complex survey sample of FTB students from 2011-12 to answer two research questions: first, what demographic, familial, academic, financial, institutional, and geographic characteristics are associated with the distance between students' homes and schools? And second, are there notable descriptive differences in mobility patterns for traditional and nontraditional FTB college students?

Analyses from the first research question offered several interesting findings. First, demographic and background characteristics were significantly associated with distance, became nonsignificant once the full vector of controls was introduced into the regression models. So, while early models suggested that characteristics like race or Hispanic and Latinx ethnicity impacted the distance which students chose to attend college, buttressing the literature (Alm & Winters, 2009; Mixon, 1992; Niu, 2015; Tuckman, 1970; Turley, 2006, 2009), the full model's output provides evidence that economic capital, and financial, institutional, and geographic characteristics are more impactful on distance. The full model also showed small, but significant associations between the number of institutions found in students' CZs and the distance of their enrolled colleges. Students who lived in areas with more 4-year college options were more likely to remain closer to home when attending college, while those who had more 2-year college options were more likely to travel farther distances for college. Students from more rural areas, as well as those from areas with high levels of household income, were likely to travel farther distances. Taken together, the evidence suggests that not only are finances, institutional characteristics, and home geography important when choosing how far to travel for college, but that individual and familial traits tend not to be impactful on student mobility when accounting for these and other factors for FTB students. Further research and replication are needed to see if other samples of mixed-age, traditional and nontraditional FTB students yield similar findings.

This study also explored the mobility patterns of traditional and nontraditional students defined by three different criteria—age when entering college, timing of college enrollment respective of high school graduation, and enrollment in a private, for-profit college. These exploratory analyses showed that nontraditional students who are older or begin college more than a year after graduating high school attended college closer to home than traditional students. Using this definition, higher proportions of nontraditional students tended to remain in-state for college, but more specifically, higher proportions remained in the same CZ as the one they live; even among in-state mobility patterns, nontraditional students differed from traditional students. And among those that traveled out of state, nontraditional students attending neighboring states than nonneighboring states while traditional students did the inverse. While this study does not offer explanations for why these differences take place—future studies can explore the reasons why these

patterns exist—some of the descriptive differences between both groups ,such as the number of dependents, household income, disabilities and marital status, may help explain differences in mobility. However, the third nontraditional definition, which was based on whether students attended private, for-profit institutions, makes pinpointing differences in distance more difficult. Students at these institutions differed from traditional students in the aforementioned traits (dependents, number of dependents, income, and marital status) in similar ways as the prior two definitions, yet these nontraditional students traveled farther for college than traditional students did. Further, larger proportions of these nontraditional students attended out-of-state schools—more specifically, nonneighboring states—than their peers. Future research interested in mobility patterns for nontraditional students should consider exploring why students attending private, for-profit colleges buck the mobility and distance trends of other nontraditional students and whether these differences have any implications for college retention, college spending, and graduation rates.

REFERENCES

Alm, J., & Winters, J. V. (2009). Distance and intrastate college student migration. *Economics of Education Review*, *28*(6), 728–738. https://doi.org/10.1016/j.econedurev.2009.06.008

Bryan, M., Cooney, D., and Elliott, B. (2019). *2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17) Data File Documentation* (No. NCES 2020-522). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Cabrera, A. F., & La Nasa, S. M. (2000). Understanding the college-choice process. *New Directions for Institutional Research*, *2000*(107), 5–22. https://doi.org/10.1002/ir.10701

Cooke, T. J., & Boyle, P. (2011). The migration of high school graduates to college. *Educational Evaluation and Policy Analysis*, *33*(2), 202–213. https://doi.org/10.3102/0162373711399092

Desmond, M., & Turley, R. N. L. (2009). The role of familism in explaining the Hispanic-White college application gap. *Social Problems*, *56*(2), 311–334. https://doi.org/10.1525/sp.2009.56.2.311

Frenette, M. (2004). Access to college and university: Does distance to school matter? *Canadian Public Policy / Analyse de Politiques*, *30*(4), 427. https://doi.org/10.2307/3552523

González Canché, M. S. (2017). The heterogeneous non-resident student body: Measuring the effect of out-of-state students' home-state wealth on tuition and fee price variations. *Research in Higher Education*, *58*(2), 141–183. https://doi.org/10.1007/s11162-016-9422-2

González Canché, M. S. (2018). Geographical network analysis and spatial econometrics as tools to enhance our understanding of student migration patterns and benefits in the U.S. higher education network. *The Review of Higher Education*, *41*(2), 169–216. https://doi.org/10.1353/rhe.2018.0001

Hill, J., Smith, N., Wilson, D., & Wine, J. (2016). *2012/14 Beginning Postsecondary Students Longitudinal Study (BPS:12/14): Data File Documentation* (No. NCES 2016-062). Washington, DC: National Center for Education Statistics.

Hillman, N. W. (2016). Geography of college opportunity: The case of education deserts. *American Educational Research Journal*, *53*(4), 987–1021. https://doi.org/10.3102/0002831216653204

Hillman, N., & Weichman, T. (2016). Education deserts: The continued significance of "place" in the twenty-first century. *Viewpoints: Voices from the Field*.

Klasik, D., Blagg, K., & Pekor, Z. (2018). Out of the education desert: How limited local college options are associated with inequity in postsecondary opportunities. *Social Sciences*, *7*(9), 165.

Mixon, F. G. (1992). Factors affecting college student migration across states. *International Journal of Manpower*, *13*(1), 25–32. https://doi.org/10.1108/EUM0000000000900

Mixon, F. G., & Hsing, Y. (1994). College student migration and human capital theory: A research note. *Education Economics*, *2*(1), 65–73. https://doi.org/10.1080/09645299400000005

Niu, S. X. (2015). Leaving home state for college: Differences by race/ethnicity and parental education. *Research in Higher Education*, *56*(4), 325–359. https://doi.org/10.1007/s11162-014-9350-y

Paulsen, M. B. (2001). The economics of human capital and investment in higher education. In M. B. Paulsen & J. C. Smart (Eds.), *The Finance of Higher Education: Theory, Research, Policy, and Practice*. Algora.

Paulsen, M. B., & St. John, E. P. (2002). Social class and college costs: Examining the financial nexus between college choice and persistence. *Journal of Higher Education*, *73*(2), 189–236. https://doi.org/10.1080/00221546.2002.11777141

Perna, L. W. (2000). Differences in the decision to attend college among African Americans, Hispanics, and Whites. *Journal of Higher Education*, *71*(2), 117–141. https://doi.org/10.1080/00221546.2000.11778831

Perna, L. W. (2006). Studying college access and choice: A proposed conceptual model. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 21, pp. 99–157). Springer Netherlands. https://doi.org/10.1007/1-4020-4512-3_3

Tuckman, H. P. (1970). Determinants of college student migration. *Southern Economic Journal*, 184-189.

Sá, C., Florax, R. J., & Rietveld, P. (2004). Determinants of the regional demand for higher education in the Netherlands: A gravity model approach. *Regional Studies*, *38*(4), 375–392. https://doi.org/10.1080/03434002000213905

Schuetze, H. G., & Slowey, M. (2002). Participation and exclusion: A comparative analysis of non-traditional students and lifelong learners in higher education. *Higher Education*, *44*(3), 309–327. https://doi.org/10.1023/A:1019898114335

Tseng, V. (2004). Family interdependence and academic adjustment in college: Youth from immigrant and U.S.-born families. *Child Development*, *75*(3), 966–983. https://doi.org/10.1111/j.1467-8624.2004.00717.x

Turley, R. N. L. (2009). College proximity: Mapping access to opportunity. *Sociology of Education*, *82*(2), 126–146. https://doi.org/10.1177/003804070908200202

U.S. Census Bureau. (2012a). 2010 Geographic terms and concepts: Census tract. Retrieved March 16, 2019, from https://www.census.gov/geo/reference/gtc/gtc_ct.html

U.S. Census Bureau. (2012b). 2010 Geographic terms and concepts: Core based statistical areas and related statistical areas. Retrieved March 16, 2019, from https://www.census.gov/geo/reference/gtc/gtc_cbsa.html

Wine, J., Bryan, M., & Siegel, P. (2014). *2011-12 National Postsecondary Student Aid Study (NPSAS:12) Data File Documentation* (No. NCES 2014-182). Washington, DC: National Center for Education Statistics.

CHAPTER 4: PREDICTING STUDENT MOBILITY: AN EXAMINATION OF COLLEGE CHOICE SETS

Introduction

The process through which students choose a college to attend has been formulated in multiple ways, but a necessary component of it is the construction of a college choice set. The college choice set is defined as the list of postsecondary institutions—whether an explicit, written list or a mental one—created during a "search" (Cabrera & La Nasa, 2000; Hossler & Gallagher, 1987) and from which the student chooses one (or several) to attend. The models used to define the process through which students search for, and ultimately choose, postsecondary institutions have until recently barely considered geography in their formulations. Many interventions designed to aid students during their searches and choices often focus on improving the transmission of college-relevant information or behavioral nudges (Castleman & Page, 2015; Hoxby & Turner, 2015; Sherwin, 2012). While the research has shown that many of these interventions have had success in increasing college applications and enrollment, particularly for low-income or first-generation college students, they rarely if ever grapple with the geospatial realities that students often face. Familial obligations, community ties, or other factors may make students more reluctant to travel farther for college and consider distal postsecondary opportunities, net of economic realities. In these situations, increasing the college-relevant information students receive or improving the methods for assisting college-going behaviors may not be as important to growing their college choice sets—especially if their local communities offer few postsecondary opportunities. It is imperative to understand more about how students develop choice sets, particularly regarding college choice sets that mix proximate and distal institutions.

This study includes analyses that elaborate on college choice sets and explore what factors play in to whether colleges within a choice set are closer or farther away from home. Using secondary data from HSLS:09/12 and the U13 follow-up, as well as public data from CCD, PSS, IPEDS, and the U.S.

Census/TIGERline geographic data, the study explores the college choice sets of a cohort of college-bound, traditional age students. The data comprises, among many variables, up to three colleges that each student considered before enrolled in college in the fall of 2013. OLS regression and multilevel modeling (MLM) will be used to explore the colleges nested within each student's choice set. Furthermore, this study will argue for the importance of considering geography when constructing models of college decision-making, and the need for conceptual frameworks that situate proximity and geographic characteristics at the forefront of the decision-making process.

Literature Review

The literature reviews for chapters 2 and 3 highlight the research behind an abundance of factors that play into how students decide whether to attend college as well as what postsecondary options to consider for enrollment. There are numerous demographic, academic, institutional, and geographical considerations that factor into the college decision-making process for students and families. Similarly, as these reviews have shown, part of the choice process involves grappling with the distance students wish to travel when attending college—whether they prefer to live close to home or are willing to go farther away from their community. Reasons for differential mobility patterns range from financial, to cultural, to convenience of college enrollment. In Paulsen and St. John's (2002) overview of the assumptions behind the student-choice construct and the financial nexus model, they mention a lack of mobility tied to financial constraints and cultural habitus as a reality for college-intending students. In the case of Latinx students, researchers suggest that *familismo*—the process through which individuals are taught to understand that familial values, interests, desires, and choices take precedent over their own, normally associated with Latinx populations—can manifest in the form of a preference to live close to home (Desmond & Turley, 2009; Ovink & Kalogrides, 2015). This preference is associated with significantly reduced likelihoods of applying to college, and even erases the application gap between Latinx and White students to selective institutions when included as a control, suggesting that mobility has strong implications for college match (Desmond & Turley, 2009). Other work by Ruth López Turley (2009) has provided evidence of geographic spillover effects, as each additional college within commuting proximity increases the likelihood of applying to and enrolling in colleges,

particularly for low-income students—she argues that this is due to the convenience factor related to having colleges close by, making the transition to college easier. All told, finances, culture, institutional characteristics and dispositions (that is, preferences and desires related to college-going) are all weighed when making college choices.

The college search and choice components of the college choice process comprise the second and third pieces of the process as formulated by Hossler and others (Cabrera & La Nasa, 2000; Hossler, Braxton, Coopersmith, & Smart, 1989). During the search phase, students create their choice sets—the group of colleges they are considering. Research on choice sets often define them differently, such as the list of colleges that they have been admitted to (Avery & Hoxby, 2004; Chapman, 1986; Niu, Tienda, & Cortes, 2006), a list of colleges constrained by academic characteristics signifying likely admission (Niu & Tienda, 2008), or a list of every college in the United States (Long, 2004). At the student level, academic achievement, demographic characteristics, financial resources, and their locations play a role in the types of colleges they seek to include in their choice sets (Shaw, Kobrin, Packman, & Schmidt, 2009). Once the students are faced with a group of colleges to which they've been admitted (regardless of how one defines college choice sets), the student makes a choice decision (Chapman, 1986). That decision relates to which college to attend, whether to attend college at all, and a spectrum of other admissions-related decisions—for instance, taking a gap year, deferring enrollment, or enrolling in multiple institutions simultaneously. Like the search phase, this choice is made based on a number of factors that the students must weigh. From a human capital perspective, this means deciding whether the costs of attending a college are worth the benefits to their future income, skills, and growth.

It is important to note that not all factors affect college choices for all students the same way. If one considers the college choice process to be made up of decisions about whether to invest in human capital, it would make sense to view college choice as a weighing of costs—both actual financial costs to attend but also opportunity costs related to not attending college—and benefits of a degree. However, students react to and are influenced by certain costs or benefits in differential ways and as such invest in human capital at differential rates. For example, Avery and Hoxby (2004) highlight several examples in which students are

influenced by financial aid so much that they make suboptimal choices in regards to their human capital investiture by choosing a college that would not give them the best return on their investment. Similarly, some students are not enticed enough by financial aid to attend colleges that are of higher quality or more selective. Other studies have noted the differential impacts of financial aid for students across income levels (Long, 2004). Similarly, there is experimental evidence that suggests that interventions that provide college-relevant information and funding may affect the choice of college they seek, but not whether they develop predispositions toward postsecondary education in the first place (Bergin, Cooks, & Bergin, 2007).

While there is plenty of research exploring how students construct choice sets or how they choose an institution, there's little regarding proximity to institutions in their choice sets. As noted in both the literature reviews and analyses of chapters 2 and 3, most students choose to attend college near their home. In some cases, this is an explicit decision, as when students attend a local community college in order to remain close to home (Reyes et al., 2019). There is also evidence that labor markets in the home community can drive enrollment decisions in specific programs tied to high-growth job markets near home (Reyes et al., 2019). Studies on the geography of opportunity (Turley, 2009) have examined racial/ethnic and economic disadvantages to local college options and the need for development of more postsecondary opportunities in areas that are disproportionally low-income and minority-majority (Dache-Gerbino, 2018). While the focus on the geography of opportunity is thankfully increasing in the literature, there is still much to be learned about how geography and distance play into what institutions are considered in students' choice sets.

Research Questions

This study examines how far students are willing to travel for college by exploring the colleges they identified as being part of their choice sets. To drive inquiry, two research questions have been developed:

1) What is the association between students' demographic, academic, and other factors and the distance between them and the colleges in their choice set?
2) What institutional factors can predict the distance between the high school attended and institutions in each student's choice set?

Using both the literature and the two prior studies, this research project seeks to identify whether the distance between the high school attended and colleges that students consider can be predicted in the same way that college enrollment as a whole is predicted. Answering these research question requires further exploration than simply running a predictive model, however; this study will also contain descriptive statistics regarding the distribution of distance within college choice sets, whether there is a correlation between distance and number of colleges applied to, and other diagnostic and exploratory analyses tangential but related to the main research question. It is important to note that, given the way the survey instrument was developed, not all respondents in the sample received the question related to the main dependent variables—therefore, this study looks specifically at students who ultimately registered in at least one college in the fall of 2013 and had nonmissing data for the variables used in the models.

Data

Data Source and Description

The primary data source used for this study was HSLS:09/12 and the U13 follow-up. The data set comprised data on 9th grade students in 2009, and its follow-ups used for this study consisted of data two years and three years after the base year, respectively. Further detail on HSLS:09/12 and U13 can be found in the data section of chapter 2. Data from HSLS:09/12 and U13 were supplemented with data from CCD, PSS, IPEDS, and U.S. Census TIGER/line geographic data files. CCD and PSS were utilized to pull longitude and latitude information about the secondary schools that students attended as that was the most specific geographic data available on where the students resided. IPEDS was used both to pull longitude and latitude of the colleges in each student's choice sets as well as supplementary institutional characteristics; the U13 update provides some institutional characteristics for the colleges in their choice set derived from IPEDS data as part of the follow-up, though the number of characteristics are limited. IPEDS identifiers are provided to merge other IPEDS data to these institutions. Lastly, shapefiles provided by U.S. Census TIGER/line were used to count institutions within CZs and add CZ identifiers to students based on where their high school was located.

Data Processing

The data sources were prepared for analysis through a process of subsetting and reshaping the data sets. Data were processed in R, Stata, and QGIS. Data processing mirrored the procedures described in the data processing section of chapter 2: HSLS:09/12 and U13 were merged with CCD, PSS, and IPEDS based on the appropriate ID variables found in each data set. The data were then subset to students who attended college in the fall of 2013 prior to calculating the distance between them and their college(s). However, instead of utilizing GEOTOOLS to draw lines between students and institutions and calculating distances in QGIS using a Python script as was done in chapter 2, this study instead relied on the GEOSPHERE package in R version 3.6.1 to calculate distances between students and institutions. GEOSPHERE provides a toolkit for creating distance matrices between a set of coordinates—that is, pairs of longitudes and latitudes—along either ellipsoids or spheres. This package negates the need to draw lines within shapefiles to then calculate the length of the lines for distance, as the GEOTOOLS Stata package does, and instead calculates the distance matrices with the coordinates themselves. The distance was calculated using the Vincenty ellipsoid method with the WGS 84/Pseudo-Mercador CRS (EPSG:3857).

As mentioned in chapter 2, this is preferred to get a more accurate measure of distance in miles rather than degrees. The merged data set—which has observations at the student level—was reshaped prior to running the distance calculation so that each observation referred to a student-college pair. There were between one and three observations per student, as the U13 follow-up only allowed for up to three colleges reported as part of their choice sets including the one they ultimately attended. Once distances were calculated at the student-college observation level, final data cleaning was done to remove missing flags and further subset the data set to students without missing data; subpopulation information is reported later in the article. Finally, the data were cloned and reshaped to the original student observation level—both cleaned data sets were used for descriptive statistics and modeling purposes.

Sample

Two analytic samples were used for each research question. First, given the nature of the data set, the analytic samples reflected a population of high school graduates who began high school in 2009 and attended

a postsecondary institution in the fall of 2013. The sample was further constrained to include complete data on both geographic characteristics in order to calculate distances, as well as complete data for covariates (described in a subsequent section). The first research question concerned the distance of colleges inside each student's choice set and was analyzed at the student level. The analytic sample used for this question after data processing comprised 8,220 students. Each student identified between one and three colleges that were a part of their college choice set, with one of these being the college they attended in the fall of 2013. The analytic sample used for research question 2 was constituted of each individual college that comprised the choice sets of the 8,220 students. In order to be included in the analytic sample for research question 2, the colleges needed to have location data for calculating distances and complete institution data taken from IPEDS. In total, there were 17,160 colleges across the 8,220 students.

Research Design

Methods

This study relies on a combination of OLS and MLM to perform the analysis. MLM allows for analyses within a regression framework that accounts for clustering of observations (Rabe-Hesketh & Skrondal, 2012). This methodology avoids the assumption that all residual error is independent, and instead separates cluster- or group-level error from the observation-level residual error (Rabe-Hesketh & Skrondal, 2012). First, summary statistics will be provided on the mean distance of colleges in students' choice sets across demographic, academic, and other characteristics. Additionally, the colleges in the students' choice sets will be ordered based on which is closer to the students (that is, rank them as closest, second closest, and third closest, respective of the number of colleges that respondents identified) and provide proportions for the sample regarding how many attended the closest in their choice set and which did not.

Then, OLS and MLM will be conducted to answer the research questions. During OLS regressions, four different versions of the model presented below will be analyzed based on different formulations of the dependent variable. MLM will be conducted in a stepwise manner—First, an unconditional model will be tested to examine the dependent variable while accounting for clustering without any predictive variables

(Rabe-Hesketh & Skrondal, 2012). Then, a single random intercept model will be tested. The full models are as follows:

$$Log(AVGDIST)_i = \beta_0 + \beta_1 B_i + \beta_2 A_i + \beta_3 S_i + \beta_4 F_i + \beta_5 SC_i + \beta_6 G_i + \varepsilon_i \quad (5)$$

$$Log(DISTANCE)_{ij} = \gamma_{00} + \gamma_{01} I_{ij} + u_{0j} + r_{ij} \quad (6)$$

Equations 5 and 6 answer research questions 1 and 2, respectively. Equation 5 represents an OLS model, while equation 6 is a multilevel model. The two dependent variables are the log-transformed variables $AVGDIST_i$ and $DISTANCE_{ij}$. $AVGDIST_i$ refers to the geodesic distance between the high school attended by each student *i* and the colleges in their choice set. Four different versions of equation 5 were analyzed—one in which the mean distance of their choice sets were taken, one in which the minimum distance was examined, one in which the maximum distance was examined, and one in which the range between minimum and maximum distance were examined. $DISTANCE_{ij}$ is a continuous variable representing the distance between college *i* from student *j*. In other words, the MLM considered colleges in the choice set the level-1 component of the model, while students were level-2 components of the model.

There are many predictive variables in the models, consolidated here within groups to conserve space. β_0 and γ_{00} are the constants; *B* is a vector of variables related to demographic and background characteristics; *A* is a vector of academic variables; *C* is a vector of social and cultural capital variables, which include preferences related to college choice and components of habitus as referenced in chapter 2; *F* is a vector of family characteristics; *S* is a vector of school-level variables; and *G* is a vector of geographic variables. In the OLS model, ε_i represents the observation error. In the multilevel model, *I* is a vector of institution-level characteristics for each college in the choice set. Finally, in the multilevel model, u_{0j} is the school-level random intercept term, and r_{ij} is observation error.

Variables

Dependent Variables

The dependent measure for research question 1 was the distance between students and the colleges in their choice set. This distance measure was operationalized in four different ways, yielding four different dependent variables that were modeled. The four dependent variables for this research question were the

mean distance from the choice set, the minimum distance from the choice set, the maximum distance from the choice set, and the distance range from the closest and farthest college in the choice set. In order to normalize the variables, each one was log-transformed prior to regression analysis. The second research question, which used an analytic sample at the college choice level, explored the distance between the institution and the corresponding student's high school. As with the distance measures used for research question one, this value was log-transformed to normalize the outcome prior to analysis.

Predictor Variables

As shown previously, the first model involved six vectors of predictor variables, while the second model included one vector of predictor variables at level-1 (the college choice level). The first vector of predictor variables pertained to background characteristics. These included demographic variables and consisted of race, Hispanic or Latinx identity, first-generation college student status, the number of high schools attended, a composite score of socioeconomic status, and student age in the fall of 2013. These predictors are important factors in college choice, although the findings from chapter 2 did not yield significant associations between distance and these measures. The next group of characteristics pertained to academic experiences, academic success, and college applications. These variables included 11th grade GPA, an indicator variable for AP level coursetaking, an indicator for college exam prep coursetaking, continuous variables for PSAT/PLAN and SAT/ACT testtaking, a categorical variable controlling for the number of colleges in each student's choice set, a continuous variable for the number of college applications submitted, and the degree program level the student enrolled in for the fall of 2013.

The next vector of characteristics modeled included social and cultural capital variables. These variables comprised measures of college-relevant social capital as well as dispositions relevant to habitus and preferences noted by the student both prior to and during the U13 data collection. The variables in this vector included indicator variables for whether the student talked to their parents, peers, teachers, and counselors about college in 2009, a school motivation scale (described in chapter 2) measured in 2011, a categorical variable highlighting the student's educational expectations, the student's preference for living close to home when deciding on a college, and the stated importance of distance when choosing their

enrolled college in 2013. This variable was not included in chapter 2, and as measured did not identify whether the importance of distance meant that being more proximate or more distant was important; rather, the item as measured asked whether distance in general was important when making their college choice. The next group of variables consisted of family variables. This group included four characteristics: the number of household members, a categorical measure of family income, an indicator variable identifying whether the student's parents considered postsecondary education the most important thing for them to do in 2013, and a categorical variable identifying the student's perception of whether their parents could afford to send them to college. The item used to derive this variable asked whether, their parents could afford to send them to college. They could choose "strongly disagree," "disagree," "agree," or "strongly agree" in response.

The final two groups used in the first model pertained to high school characteristics and geographic characteristics. The three high school characteristics used as predictors were school control—that is, whether the school was public or private—the school's urban locale, and the percentage of free lunch eligible students. The geographic variables included characteristics related to the student's CZs. CZs, as described in prior chapters, refer to a unit of geography made up of several counties that are considered coherent labor markets. Included as predictors were the number of 4-year institutions in the CZ, the number of 2-year institutions in the CZ, the median population of the CZ in 2010 in thousands, and the median wage of residents of the CZ in 2010 in thousands of dollars. Lastly, a state fixed effect was included to control for the state in which a student resided. The multilevel model for research question 2 only included one group of variables at level-1. This group comprised institution characteristics extracted from IPEDS data in the 2013-14 academic year. These variables included institution level and control, institution selectivity, HBCU status, the 12-month enrollment count in thousands, the proportion of White students in the 12-month enrollment count, the proportion of Hispanic or Latinx students in the 12-month enrollment count, in-state tuition and fees in thousands of dollars, and institution urban locale.

Weighting and Subpopulation Considerations

This study relies on similar weighting considerations as chapter two does; for detailed information about how the complex survey design was accounted for in this study, see the weighting and subpopulation

considerations section of chapter two. There are two main differences between this study and the one conducted in chapter two: first, weights were not used for the modeling analysis of research question 2—MLM—due to an incompatibility with the weights provided and the requirements for MLM using survey weights in Stata 16. In order to conduct multilevel analyses in Stata 16, the data need to have weights for the level-one and level-two characteristics; in this case, as level-1 was defined as colleges within the choice set and level-2 was the students, the weights needed were not available. Instead of using the SVY command, the student weight (W3W1W2STUTR) was included as part of the random effects syntax so that the higher level components of the multilevel models were weighted with the available student weight. Second, the subpopulation used for this study differed from the study in chapter two given the difference in variables used and research questions asked. Given the differences in variables used for both studies, the present study had a different number of analytic sample members compared to the study in chapter two. As mentioned in the sample section above, the analytic sample for this study contained about 8,220 students, with about 17,160 student-college pairs.

Results

Descriptive Statistics

Table 4.1 shows the frequency distribution of the variables of interest for the student sample. The majority of students in the analytic sample identified three colleges in their choice set—the maximum number allowed in the U13 follow up for HSLS. About 43.9 percent of students identified three colleges as having been considered, compared to 20 percent and 36 percent for two colleges and one college, respectively. Additionally, the majority—66.8 percent—of the sample ultimately enrolled in the closest college in their choice set. This includes all of the students who only identified one school, as they by default attended the closest college in their choice set. About 44.6 percent of the sample that identified two or three colleges in their choice set enrolled in the closest college in their choice set.

Table 4.1: Frequency Statistics for Student Analytic Sample Variables

	Analytic Sample: Students		
Variable	**N**	**Unweighted percent**	**Weighted percent**
Dependent Variables			
Number of colleges in choice set			
One	2,680	32.62	36.04
Two	1,650	20.13	20.02

See notes at end of table.

Table 4.1: Frequency Statistics for Student Analytic Sample Variables—Continued

Variable	Analytic Sample: Students N	Unweighted percent	Weighted percent
Three	3,800	46.19	43.94
College chosen by relative distance in choice set			
Not enrolled in closest college	2,830	34.38	33.21
Enrolled in closest college	5,300	64.56	66.79
Independent Variables			
Preference to remain close to home for college			
Not at all important	2,500	30.46	29.9
Somewhat important	4,000	48.71	48.86
Very important	1,620	19.77	21.24
Race			
White	5,560	67.61	69.42
Black/African American	720	8.74	11.06
Asian American	830	10.08	5.03
Native Hawaiian/Pacific Islander	80	0.91	1.26
American Indian/Alaska Native	140	1.75	2.76
Multiracial	810	9.85	10.46
Hispanic or Latinx			
No	7,120	86.68	83.13
Yes	1,010	12.26	16.87
First-generation college student			
FGCS	5,850	71.2	65.31
Non-FGCS	2,280	27.74	34.69
Number of high schools attended			
One	7,430	90.42	92.1
Two	620	7.56	6.83
Three or more	80	0.96	1.07
Any AP courses taken			
No or not sure	4,200	51.17	53.15
Yes	3,930	47.77	46.85
Took a course to prepare for college admissions exam			
No	3,900	47.47	52.67
Yes	4,230	51.47	47.33
Number of times PSAT or PLAN taken			
Never or don't know	1,650	20.07	24.78
Once	3,440	41.82	41.95
Twice	2,220	27.04	25.4

See notes at end of table.

Table 4.1: Frequency Statistics for Student Analytic Sample Variables—Continued

Variable	Analytic Sample: Students N	Unweighted percent	Weighted percent
3 or more times	820	10	7.87
Number of times SAT or ACT taken			
Never or don't know	3,900	47.43	55.78
Once	2,900	35.3	31.43
Twice	910	11.08	8.91
3 or more times	420	5.14	3.88
Talked to parents about going to college: 9th grade			
No	860	10.46	11.42
Yes	7,270	88.49	88.58
Talked to friends about going to college: 9th grade			
No	3,300	40.21	41.69
Yes	4,830	58.73	58.31
Talked to teacher about going to college: 9th grade			
No	6,360	77.36	77.69
Yes	1,770	21.58	22.31
Talked to counselor about going to college: 9th grade			
No	6,620	80.57	81.54
Yes	1,510	18.37	18.46
Postsecondary education expectations			
Less than certificate or don't know	800	9.71	10.39
Complete a certificate	180	2.22	2.69
Some 2-year college or 2-year degree completion	540	6.56	7.53
Some 4-year college or 4-year degree completion	2,560	31.16	32.36
Some Master's education or Master's completion	2,380	28.99	28.87
Some doctoral/terminal education or doctoral/terminal degree completion	1,670	20.3	18.16
Importance of distance when choosing college			
Not at all important	2,090	25.44	25.98
Somewhat important	3,290	40.02	38.36
Very important	2,750	33.48	35.66
Family income (categorical)			
Less than or equal to $15,000	490	5.99	7.61
$15,001 - $35,000	1,010	12.28	15.3
$35,001 - $55,000	1,190	14.44	15.69
$55,001 - $75,000	1,180	14.33	14.97
$75,001 - $95,000	1,020	12.35	12.38
$95,001 - $115,000	870	10.61	10.28

See notes at end of table.

Table 4.1: Frequency Statistics for Student Analytic Sample Variables—Continued

Variable	Analytic Sample: Students N	Unweighted percent	Weighted percent
$115,001 - $135,000	650	7.87	6.47
Greater than $135,000	1,730	21.07	17.3
Parents think is most important for student to do in fall 2013			
Continuing education after high school	420	5.11	5.81
Something other than continuing education	7,710	93.83	94.19
Parents can't afford to send to college			
Strongly disagree	2,320	28.2	25.75
Disagree	4,220	51.4	51.56
Agree	1,290	15.75	18.49
Strongly agree	300	3.59	4.21
First follow-up school: control			
Public	6,420	78.19	91.32
Private	1,710	20.75	8.68
First follow-up school: locale (urbanicity)			
City	2,390	29.08	29.06
Suburb	2,530	30.78	30.59
Town	960	11.65	11.23
Rural	2,250	27.43	29.12
First follow-up school: percent free lunch (categorical)			
Zero	1,210	14.72	6.67
0 - 9%	850	10.33	8.31
10 - 19%	1,030	12.49	12.06
20 - 29%	1,070	13.01	14.82
30 - 39%	1,010	12.27	13.05
40 - 49%	900	10.89	14.1
50 - 59%	740	9.04	10.05
60 - 69%	580	7.1	7.95
Greater than or equal to 70%	750	9.09	13
Program level in college			
No degree or other	1100	13.39	16.11
Certificate or diploma program	160	1.94	2.09
Associate's degree program	2450	29.87	34.16
Bachelor's degree program	4420	53.75	47.64
Number of colleges in choice set			
One	2680	32.62	36.04
Two	1650	20.13	20.02

See notes at end of table.

Table 4.1: Frequency Statistics for Student Analytic Sample Variables—Continued

Variable	Analytic Sample: Students N	Unweighted percent	Weighted percent
Three	3800	46.19	43.94
Enrolled college was closest in choice set			
Didn't enroll in closest college	2830	34.38	33.21
Enrolled in closest college	5300	64.56	66.79

NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

Table 4.2 shows the distribution of continuous characteristics for the student sample. As was the case in chapters two and three, the distribution for distance was heavily skewed, necessitating log-transformation to normalize the distance measure. This was true for each version of distance for the choice sets—mean, minimum, maximum, and range. The mean minimum distance from colleges in the choice sets was 82.2 miles, with 248 miles as the mean maximum distance. On average, students maintained a difference of 165.9 miles between the closest and farthest colleges in their choice sets—however, as noted by the high standard deviation, this figure was largely skewed.

Table 4.2: Summary Statistics for Student Analytic Sample Variables

	Analytic Sample: Students					
Variable	**Min**	**Max**	**Unweighted Mean**	**Unweighted SD**	**Weighted Mean**	**Weighted SD**
Dependent Variables						
Mean distance to colleges in choice set, in miles	0.17	4,377.69	173.57	310.10	157.96	7.23
Log-transformed mean distance	-1.76	8.38	4.09	1.60	3.93	0.04
Minimum distance to colleges in choice set, in miles	0.01	3,971.32	87.70	229.32	82.18	4.98
Log-transformed minimum distance	-4.38	8.29	3.16	1.63	3.06	0.06
Maximum distance to college in choice set, in miles	0.17	4,975.20	273.39	478.93	248.04	10.88
Log-transformed maximum distance	-1.76	8.51	4.42	1.75	4.24	0.04
Range between minimum and maximum distance, in miles	0.00	4,973.67	185.69	400.17	165.86	8.11
Log-transformed range distance	-4.84	8.51	4.62	1.61	4.51	0.04
Independent Variables						
Socioeconomic status	-1.75	2.15	0.33	0.73	0.17	0.02
Age in 2013	16.00	23.00	18.29	0.52	18.29	0.01
11th grade GPA reported on transcript	0.00	4.00	3.01	0.87	2.92	0.02
School motivation scale	-5.64	1.21	0.22	0.80	0.18	0.01
Number of household members	1.00	13.00	4.20	1.38	4.22	0.02
Number of college applications	1.00	60.00	3.34	2.90	3.15	0.06
Commuting zone: Median population in 2010, in thousands	1.87	2,189.64	251.85	458.31	308.98	18.67
Commuting zone: Median wage in 2010, in thousands of dollars	22.54	57.92	37.60	6.51	38.54	0.34
Commuting zone: number of postsecondary institutions	1.00	380.00	62.29	75.31	67.12	3.16
Commuting zone: number of 4-year postsecondary institutions	0.00	141.00	26.08	30.66	27.77	1.28
Commuting zone: number of 2-year postsecondary institutions	0.00	108.00	16.22	20.07	17.44	0.79

NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

Tables 4.3 and 4.4 provide descriptive statistics for categorical and continuous characteristics at the college choice level, respectively. These tables explore an analytic sample comprising each institutional choice nested within students with a total sample size of 17,160 college choice observations. Given the structure of the data set, these statistics were not weighted as mentioned in the weighting section previously. The majority of institutions considered as part of the choice set for the 8,220 students in the data set were 4-year institutions, comprising 84.9 percent of the college choices in the sample. The majority of the sample were public institutions, and a total of 71.8 percent of the college choices were either moderately selective or very selective 4-year institutions. Almost none of the students in the analytic sample considered less-than-2-year college options, as they accounted for .01 percent of the college choice sample. Predictably, the distance from students to colleges in the choice set was highly variable as evidenced by the large standard deviation The majority of schools in the college choice sets were majority White, with a mean proportion of .63 of the 12-month enrollment in the 2013-14 academic year. The mean in-state tuition was $14,810 with a mean in-state fee of $1,220.

Table 4.3: Frequency Statistics for College Analytic Sample Variables

	Analytic sample: Colleges	
Variable	**N**	**Unweighted percent**
College selectivity		
Not classified, less than 4-year college	2,660	15.53
Inclusive 4-year college	2,180	12.68
Moderately selective 4-year college	6,460	37.67
Highly selective 4-year college	5,860	34.13
HBCU Status		
Not an HBCU	16,910	98.54
HBCU	250	1.46
Institution level		
Less-than-2-year	#	0.01
2-year	2,590	15.07
4-year	14,570	84.92
Institution control		
Public	11,950	69.63
Private, nonprofit	5,020	29.24
Private, for-profit	190	1.13

Rounds to zero.

NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

Table 4.4: Summary Statistics for College Analytic Sample Variables

	Analytic sample: Colleges			
Variable	**Min**	**Max**	**Unweighted Mean**	**Unweighted SD**
Dependent Variables	0.01	4,975.20	204.40	403.21
Distance to college, in miles	-4.38	8.51	4.08	1.72
Log-transformed distance				
Independent Variables				
Total 12-month enrollment, in thousands	0.12	180.46	20.90	17.58
12-month enrollment: proportion of White students	0.00	0.97	0.63	0.18
12-month enrollment: proportion of Hispanic students	0.00	0.95	0.09	0.10
Published in-state tuition, in thousands of dollars	0.00	49.68	14.81	13.63
Published in-state fees, in thousands of dollars	0.00	60.21	1.22	1.29

NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

Factors Associated with Distance from High School to Choice Set Colleges

Table 4.5 includes the coefficients for the variables included in the four OLS models pertaining to research question one. These models examined what the associations were between a set of factors traditionally associated with college choice and four conceptualizations of distance to choice set colleges—mean, minimum, maximum, and range—which help describe the mobility patterns that students are willing to consider for their prospective colleges. Model one pertained to the mean distance between the student and the choice set colleges. The model was significantly predictive of the variance in mean distance, $F(50, 450) = 67.4$, $p < .001$.

Table 4.5: Ordinary Least Squares Regression Results for Factors Associated with Distance from High School to Choice Set Colleges, by Model

	OLS regression model							
	Model 1 (Mean distance)		Model 2 (Minimum distance)		Model 3 (Maximum distance)		Model 4 (Distance range)	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
Background Characteristics								
Race (ref = White, non-Hispanic)								
Black or African-American	0.138	(0.087)	-0.024	(0.113)	0.164	(0.089)	0.161	(0.113)
Asian	-0.078	(0.101)	-0.296*	(0.140)	-0.03	(0.105)	0.044	(0.157)
Native Hawaiian/Pacific Islander	-0.3	(0.243)	-0.720**	(0.270)	-0.244	(0.256)	0.078	(0.285)
American Indian/Alaska Native	0.522	(0.316)	0.615	(0.450)	0.478	(0.284)	0.002	(0.262)
Multiracial	-0.013	(0.117)	-0.082	(0.128)	-0.004	(0.119)	0.105	(0.118)
Hispanic or Latinx status	0.097	(0.094)	-0.036	(0.116)	0.113	(0.095)	0.249	(0.128)
Race and Hispanic Interaction Effects (ref = non-Hispanic)								
Black or African-American, Hispanic	-0.367	(0.280)	0.052	(0.296)	-0.408	(0.282)	-0.705*	(0.279)
Asian, Hispanic	-1.045*	(0.453)	-0.646	(0.482)	-1.081*	(0.455)	-0.12	(0.449)
Native Hawaiian/Pacific Islander, Hispanic	-0.609	(0.340)	-0.167	(0.372)	-0.64	(0.358)	-0.867	(0.465)
American Indian/Alaska Native, Hispanic	-1.005**	(0.380)	-0.825	(0.488)	-1.009**	(0.356)	-0.591	(0.444)
Multiracial, Hispanic	-0.139	(0.205)	-0.033	(0.266)	-0.134	(0.206)	-0.951***	(0.275)
First-generation college student	0.018	(0.077)	0.032	(0.080)	0.011	(0.083)	-0.143	(0.105)
Number of high schools attended	0.157	(0.082)	0.113	(0.083)	0.161	(0.087)	0.144	(0.128)
Socioeconomic Status	0.083	(0.060)	0.095	(0.061)	0.062	(0.062)	-0.029	(0.074)
Age	0.012	(0.049)	0.003	(0.055)	0.012	(0.049)	-0.007	(0.066)
Academic Characteristics								
11th grade GPA	0.130***	(0.032)	0.135***	(0.034)	0.132***	(0.033)	0.125*	(0.061)
Any AP courses taken	0.245***	(0.048)	0.263***	(0.056)	0.248***	(0.048)	0.208**	(0.074)
College exam preparation	0.043	(0.045)	0.126	(0.064)	0.026	(0.045)	-0.166*	(0.067)
PSAT/PLAN tests taken	0.041	(0.052)	0.053	(0.042)	0.037	(0.054)	0.061	(0.044)

See notes at end of table.

Table 4.5. Ordinary Least Squares Regression Results for Factors Associated with Distance from High School to Choice Set Colleges, by Model—Continued

	OLS regression model							
	Model 1 (Mean distance)		Model 2 (Minimum distance)		Model 3 (Maximum distance)		Model 4 (Distance range)	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
SAT/ACT tests taken	-0.018	(0.027)	0.042	(0.034)	-0.03	(0.028)	-0.06	(0.054)
Number of colleges in choice set	0.566***	(0.034)	-0.203***	(0.046)	0.846***	(0.037)	1.007***	(0.081)
Number of college applications sent	0.082***	(0.009)	0.088***	(0.016)	0.080***	(0.009)	0.085***	(0.012)
Program level in college	0.167***	(0.027)	0.166***	(0.029)	0.170***	(0.028)	0.084*	(0.037)
Social/Cultural Capital								
Talked to parents about college: 9th grade	0.043	(0.070)	-0.002	(0.074)	0.043	(0.071)	0.044	(0.103)
Talked to friends about college: 9th grade	0.047	(0.050)	0.069	(0.049)	0.041	(0.053)	0.109	(0.074)
Talked to teachers about college: 9th grade	0.13	(0.073)	0.115	(0.086)	0.128	(0.074)	-0.008	(0.087)
Talked to counselors about college: 9th grade	-0.072	(0.057)	-0.07	(0.067)	-0.085	(0.058)	-0.075	(0.081)
School motivation scale	-0.064*	(0.031)	-0.022	(0.033)	-0.075*	(0.032)	-0.130**	(0.046)
Educational expectations scale	0.037*	(0.016)	0.013	(0.018)	0.043*	(0.017)	0.042	(0.024)
Preference to remain close to home	-0.282***	(0.035)	-0.270***	(0.038)	-0.275***	(0.036)	-0.275***	(0.046)
Importance of distance when choosing college	-0.233***	(0.030)	-0.260***	(0.034)	-0.222***	(0.031)	-0.115**	(0.040)
Family Characteristics								
Household size	-0.015	(0.014)	-0.018	(0.016)	-0.016	(0.014)	-0.017	(0.022)
Family income (categorical)	0.035*	(0.014)	0.041*	(0.018)	0.036**	(0.014)	0.046**	(0.017)
Parental expectations: most important for student to attend college in fall 2013	0.078	(0.133)	#	(0.127)	0.098	(0.136)	0.22	(0.173)
Parents can't afford to send to college	-0.071*	(0.033)	-0.087*	(0.039)	-0.074*	(0.033)	-0.092	(0.053)
School Characteristics								
Private school	0.248**	(0.085)	0.195	(0.130)	0.250**	(0.084)	0.041	(0.101)

See notes at end of table.

Table 4.5. Ordinary Least Squares Regression Results for Factors Associated with Distance from High School to Choice Set Colleges, by Model—Continued

	OLS regression model							
	Model 1 (Mean distance)		Model 2 (Minimum distance)		Model 3 (Maximum distance)		Model 4 (Distance range)	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**	**Coeff.**	**SD**
School locale (ref = City)								
Suburb	0.189	(0.097)	0.489***	(0.146)	0.147	(0.089)	-0.086	(0.091)
Town	0.307*	(0.129)	0.766***	(0.169)	0.236	(0.125)	-0.1	(0.115)
Rural	0.479***	(0.090)	0.959***	(0.135)	0.417***	(0.085)	-0.121	(0.087)
Percent free lunch eligible students	-0.038*	(0.016)	-0.02	(0.020)	-0.040*	(0.016)	-0.055**	(0.019)
Geographic Characteristics								
Commuting zone: Number of 4-year colleges	-0.004	(0.002)	#	(0.003)	-0.004	(0.002)	-0.010***	(0.003)
Commuting zone: Number of 2-year colleges	0.006	(0.003)	0.002	(0.004)	0.007*	(0.003)	0.016***	(0.004)
Median population in 2010, in thousands	#	#	#	#	#	#	#	#
Median wage in 2010, in thousands of dollars	-0.016	(0.009)	-0.01	(0.012)	-0.015	(0.008)	-0.014*	(0.007)
State fixed effect	0.001	(0.003)	#	(0.004)	0.001	(0.003)	#	(0.002)
Constant	2.657**	(0.917)	2.324*	(1.088)	2.628**	(0.936)	2.716*	(1.250)
R^2	0.427		0.224		0.499		0.274	
$F(50, 450)$	67.397		21.178		86.575		15.579	

* $p < .05$, ** $p < .01$, *** $p < .001$

Rounds to zero.

NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding. Coeff. = Coefficient. Ref = Reference group. SD = Standard deviation.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

Most of the background characteristics were not significantly associated with the mean distance between students and their choice set colleges net of other characteristics. The interaction terms between race and ethnicity did yield significant associations—Asian Hispanic students had significantly closer mean distances to their choice set than Asian non-Hispanic students (t = -2.30, p < .05), as well as American Indian/Alaska Native Hispanic students compared to their non-Hispanic peers (t = -2.64, p < .01). Instead, some of the most important factors related to mean distance to choice set were academic, related to social and cultural capital, or secondary school characteristics. Higher 11th grade GPA and rigorous high school coursetaking were associated with larger mean distances, suggesting that students who are high achieving in high school will consider a wider range of colleges. Similarly, students who applied to more colleges and who ultimately enrolled in higher program levels—associate's and bachelor's programs—had higher mean distances to their choice sets. Several key perceptions, expectations, and cultural capital variables were also significantly associated with the distance to choice set colleges. Surprisingly, no geographic characteristics related to commuting zones were significantly associated with mean distance to choice set colleges.

Models 2 and 3 pertained to minimum and maximum distances to choice sets; both models were significantly predictive of variance as well, $F(50, 450) = 21.18$, $p < .001$ and $F(50, 450) = 86.57$, $p < .001$. The model examining maximum distance in choice set predicted almost 50 percent of the variance in distance, which was the most variance explained out of any model. Both minimum and maximum distance were predicted by many of the same variables, as noted in table 16. The findings suggest that the closest and farthest colleges students consider are often predicted by the same characteristics—likely artificially inflated by students in the sample who only identified one college in their choice set. Interestingly, high school locale was significantly and positively associated with minimum distance considered across the three comparative groups (suburb, town, and rural) while only rural was significantly and positively associated with maximum distance; this suggests that students who attend less urban communities have to consider colleges farther from them overall, while rural students are also more likely to explore options very distal from themselves.

Model 4 pertained to the range between minimum and maximum distance in the choice set, and was significantly predictive of the variance in the outcome, $F(50, 450) = 15.58$, $p < .001$. Like prior models, few

demographic characteristics were significantly associated with the distance measure; in this case, Black Hispanic students were likely to consider smaller ranges of colleges (t = -2.53, p < .05) than their non-Hispanic counterparts. This was also true for multiracial Hispanic students (t = -3.46, p < .001). Again, academic characteristics were more predictive of the distance measure, with higher GPAs and AP coursetaking associated with wider ranges of colleges in the choice set. Surprisingly, college exam preparation was associated with smaller ranges (t = -2.50, p < .05). While none of the social capital measures were associated with range (as in the previous models), several cultural capital measures, including school motivation (t = -2.83, p < .01), the preference to remain close to home (t = -5.99, p < .001) and a stated importance of distance in choosing a college (t = -2.89, p < .01) were significantly negatively associated with range. Students who were more motivated and had predispositions for remaining near their homes were more likely to shrink the range band between their choice sets than those who did not. Several geographic variables, including the number of 4-year and 2-year colleges in the student's CZ and median wage were significantly associated with range.

Institutional Factors and Distance to Individual Choice Set Schools

Table 4.6 shows the results of the MLM analysis regarding the association between institutional characteristics and distance to choice set colleges The second research question examined a data set made up of colleges nested within individual student's choice sets and explored several institutional characteristics that could be predictive with the distance between them and the student they were nested under. Both the unconditional model (χ^2 = 1301.87, p < .001) and the random intercepts model (χ^2 = 776.68, p < .001) were predictive above the linear model, suggesting that there was utility in running a multilevel model that accounted for variance at the student level. Likelihood-ratio tests comparing the unconditional model with the random intercepts model suggest that both are significantly different, $\chi^2(20)$ = 3981.41, p < .001.

Table 4.6: Multilevel Modeling Results for Institution Factors Predicting Log-Transformed Distance from High School to Choice Set Colleges, by Model

	Multilevel model			
	Model 1 (Unconditional model)		Model 2 (Random intercepts)	
Covariates	**Coeff.**	**SD**	**Coeff.**	**SD**
Institutional Characteristics (Level-1)				
Level of institution (ref. = less-than-2-year)				
2-year			-1.815	(1.049)
4-year			-2.367*	(1.062)
Control of institution (ref. = Public)				
Private, nonprofit			.135*	(0.066)
Private, for-profit			.874***	(0.113)
Institution selectivity (ref. = Unclassified or less than 4-year)				
Inclusive 4-year institution			1.444***	(0.173)
Moderately selective 4-year institution			1.916***	(0.172)
Highly selective 4-year institution			2.372***	(0.174)
HBCU status			0.86***	(0.115)
Total 12-month enrollment, in thousands			.007***	(0.001)
Proportion of 12-month enrollment: White			-0.106	(0.101)
Proportion of 12-month enrollment: Hispanic			-0.557**	(0.161)
In-state tuition, in thousands of dollars			.016***	(0.002)
In-state fees, in thousands of dollars			.038***	(0.010)
Institution locale (urbanicity) (ref. = City)				
Suburb			-0.144***	(0.029)
Town			.385***	(0.035)
Rural			.292***	(0.056)
Constant	3.813***	(0.029)	4.242***	(1.050)
Level-2 random intercept	1.06	(0.036)	0.584	(0.025)
Observational error	1.192	(0.069)	1.689	(0.024)
Chi-square			4586.02***	
Intraclass correlation	0.355	(0.010)	0.257	(0.010)

* $p < .05$, ** $p < .01$, *** $p < .001$

NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding. Coeff. = Coefficient. Ref = Reference group. SD = Standard deviation.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript.

Of the institutional factors considered for the random intercepts model, all but two variables (2-year institution level and proportion of White students in the 12-month enrollment count) were significantly associated with the distance between the student and college. Four-year institutions in student choice sets were more likely to be closer than less-than-2-year institutions, which served as the reference group for institution level (t = -1.69, p < .05). Private schools overall were more distal than public schools, with private nonprofits (t = 2.04, p < .05) being closer than private for-profit colleges (t = 7.71, p < .001). College selectivity was also significantly associated with distance, where more selective colleges tended to be more distal from the students that considered them—in other words, more selective colleges appear to be farther overall than nonselective institutions. This makes sense if one considers nonselective or open access institutions to be the most numerous across the country, with most students being substantially more likely to have a community college or 2-year institution nearby compared to a highly selective, prestigious 4-year college. Surprisingly, higher in-state tuition and fees was associated with farther distances—one would expect that more expensive institutions would be inversely related with distance, where a pricier education would make it more difficult for students to consider including them in their choice set if they were also farther away. However, the evidence suggests this is not the case.

Limitations

This study suffered from several limitations that inhibit stronger interpretations of their results. First, as with the previous two chapters, the results from this study are correlational. There were no causal effects examined with these methodologies, and as such no claim can be made about the effects of these factors on the distances between students and their choice sets. Given the structure of the survey, students could only identify up to three institutions within their choice set in such a way that allows for merging in supplemental institutional and geographic data. Limiting the choice sets to three institutions was not ideal, especially when many students only identified a single institution as part of their choice set. For the findings to be stronger, choice sets would have to be made up of more institutions—this is evidenced by another variable examined in the descriptive statistics section of the findings, number of college applications, which ranged from one application to 60. If students were considering such a vast array of institutions by applying to a wider

spectrum than was captured here, future studies that use similar methodologies would do well to consider expanding the number of colleges within the explored choice sets. Additionally, the sample included in this study comprised traditional-age students who graduated high school in a traditional timeline. Unlike the findings of chapter 3, these findings can only be extrapolated to more traditional students who matriculated to college. Additionally, the use of geodesic distance over road network distance—where distance is calculated via traveling on roads instead of a straight line—makes the analyses less realistic, as students will not travel the distance as measured here. Lastly, the multilevel models used for the second research question did not incorporate the full survey design given the lack of appropriate weights for both levels in question. The use of MLM allowed for a more nuanced examination of institutional factors tied to distance between high school and college, and the findings are important—however, more robust methods that account for survey design should be considered by future studies.

Discussion

The literature on how students construct college choice sets is vast, but relatively little attention has been paid to how far the colleges within each student's choice sets are from one another and from the student. This study explored how willing students were to be mobile for college by examining how far the colleges they considered were from their high schools, as well as the characteristics that could predict the distance between them and their choice sets. The models analyzed in this study were fairly predictive, accounting between 22 and 50 percent of the variance. The most surprising finding from this study is how little demographic and background characteristics, including socioeconomic status, predicted the distance measures when controlling for academic characteristics, social and cultural capital, family factors, school characteristics, and geography. Instead, some of the more predictive aspects of a student's live were their academic achievement, dispositions towards college-going, preferences regarding mobility, and income. Students who described distance as being important to making college choices—without specifying whether being closer or farther was more important to them—tended to explore college choices closer to home. Similarly, students who believed their family could not afford to send them to college explored a shorter band of institutions both in terms of mean distance to choice set colleges as well as lower minimum and maximum

values of distance to choice set colleges. The models suggest that academics, predispositions, and perceptions are among the more important factors for students when constructing a wider net distance-wise of colleges for their choice set. Students are more likely to expand their considerations and explore distal options when they believe they can afford to go to college, when they have higher educational expectations of themselves, are more successful in their coursework, and participate in more challenging curriculum.

At the institution level, there are several institutional factors that were associated with the distance between them and the students who included them in their choice sets. Private institutions, both nonprofit and for-profit, were more distal than public colleges. This suggests that when students include private colleges in their choice sets, net of other characteristics, they tend to be farther away from them than the public colleges in their choice sets. As mentioned in a previous example, this could be due to the abundance of and likelihood of living near open access public institutions. Nevertheless, the findings suggest that students are willing to consider distal private options in their choice sets. Similarly, students are willing to consider more distal, higher priced colleges in their choice sets as well as evidenced by the significant and positive association between tuition, fees, and distance. Higher priced colleges that students included in their choice sets tended to also be farther away from them; it may be possible that students consider more expensive options due to a perception of prestige associated with colleges that charge more money, or a perception that price signals higher institutional quality. However, this is purely speculation—nothing in the data can support that assertion, and future research can explore this idea further than this work can.

The findings from this study suggest that students are willing to be more mobile for postsecondary education given increased academic achievement and predispositions towards their ability to afford college and achieve higher levels of education. College access advocates interested in expanding the options that students wish to include in their choice sets would do well to focus on changing those affordability perceptions and educational expectations; these results suggest that demographics are not nearly as important at predicting student mobility for traditional high school students than one would think would be the case.

REFERENCES

Avery, C., & Hoxby, C. M. (2004) Do and should financial aid packages affect students' college choices? In *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 239-302). University of Chicago Press.

Bergin, D.A., Cooks, H.C., & Bergin, C.C. (2007). Effects of a college access program for youth underrepresented in higher education: A randomized experiment. *Research in Higher Education, 48*, 727–750. doi:10.1007/s11162-006-9049-9

Cabrera, A. F., & La Nasa, S. M. (2000). Understanding the college-choice process. *New Directions for Institutional Research*, *2000*(107), 5–22. https://doi.org/10.1002/ir.10701

Castleman, B. L., & Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, *115*, 144–160. https://doi.org/10.1016/j.jebo.2014.12.008

Chapman, R. G. (1986). Towards a theory of college selection: A model of college search and choice behavior. In R. J. Lutz (Ed.) *Advances in Consumer Research* (Vol. 13, pp. 246-250). Association for Consumer Research.

Dache-Gerbino, A. (2018). College desert and oasis: A critical geographic analysis of local college access. *Journal of Diversity in Higher Education*, *11*(2), 97-166. doi: 10.1037/dhe0000050

Desmond, M., & Turley, R. N. L. (2009). The role of familism in explaining the Hispanic-White college application gap. *Social Problems*, *56*(2), 311–334. https://doi.org/10.1525/sp.2009.56.2.311

Hossler, D., Braxton, J., & Coopersmith, G. (1989). Understanding student choice. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 5, pp. 231–281). Agathion Press.

Hossler, D., & Gallagher, K. S. (1987). Studying student college choice: A three-phase model and the implications for policymakers. *College and University*, *62*(3), 207–221.

Hoxby, C. M., & Turner, S. (2015). What high-achieving low-income students know about college. *American Economic Review*, *105*(5), 514-17.

Long, B. T. (2004). How have college decisions changed over time? An application of the conditional logistic choice model. *Journal of Econometrics*, *121*(1-2), 271-296.

Niu, S. X., & Tienda, M. (2008). Choosing colleges: Identifying and modeling choice sets. *Social Science Research*, *37*(2), 416-433

Niu, S.X., Tienda, M., & Cortes, K. (2006). College selectivity and the Texas top 10% law. *Economics of Education Review, 25*(1), 259–272.

Ovink, S. M., & Kalogrides, D. (2015). No place like home? Familism and Latino/a–white differences in college pathways. *Social Science Research*, *52*, 219–235. https://doi.org/10.1016/j.ssresearch.2014.12.018

Paulsen, M. B., & St. John, E. P. (2002). Social class and college costs: Examining the financial nexus between college choice and persistence. *Journal of Higher Education*, *73*(2), 189–236. https://doi.org/10.1080/00221546.2002.11777141

Reyes, M., Dache-Gerbino, A., Rios-Aguilar, C., Gonzalez-Canche, M., & Deil-Amen, R. (2019). The "geography of opportunity" in community colleges: The role of the local labor market in students' decisions to persist and succeed. *Community College Review*, *47*(1), 31-52.

Shaw, E. J., Kobrin, J. L., Packman, S. F., & Schmidt, A. E. (2009). Describing students involved in the search phase of the college choice process: A cluster analysis study. *Journal of Advanced Academics*, *20*(4), 662-700.

Sherwin, J. (2012). *Make me a match: Helping low-income and first-generation students make good college choices*. MDRC.

Turley, R. N. L. (2009). College proximity: Mapping access to opportunity. *Sociology of Education*, *82*(2), 126–146. https://doi.org/10.1177/003804070908200202

CHAPTER 5: CONCLUSION: TOWARDS A GEOSPATIAL CONCEPTUAL FRAMEWORK OF COLLEGE CHOICE

The literature exploring the importance of geography, mobility, and geospatial considerations when making college choices continues to grow. Thanks to the work of numerous scholars, we continue to understand more about how the geography of opportunity constricts or expands the postsecondary options available to students. This three-article dissertation participates in the larger discussion on geography and college choice with three studies dedicated to different aspects of student mobility for college: the preference to remain close to home when attending college, mobility differences between traditional and nontraditional FTB students, and an exploration of institutional and student-level characteristics that predict student mobility. These studies, as mentioned in the introduction to this dissertation, are also meant to serve as the foundation for a geospatial framework of college choice—a conceptual framework that positions geography and proximity to postsecondary options at the forefront of college decision-making. To conclude the dissertation, this chapter will summarize major findings from the three articles herein, highlighting implications of the work in relation to both the literature and future research. The chapter, and the dissertation, will conclude with an exploration of the geospatial conceptual framework

Study Findings

The role of attitudes towards college-going—whether referred to as predispositions, habitus, or another term—is one that impacts both college enrollment in general and college choices in particular. The first study in the dissertation focused its attention on a predisposition relevant to geography and college choice: the desire to live near home when choosing to attend college. Framed around three research questions, the analyses in the first study explored the characteristics associated with differing levels of the preference, whether the preference had implications for college enrollment rates, and whether students who

stated they preferred to remain close to home ultimately chose colleges near their home communities. Very few demographic and background characteristics were associated with differences in the preference to remain close to home. Of those that were, some associations ran counter to traditional understanding of the factors associated with wanting to live near home; surprisingly, Hispanic or Latinx identity and SES were found to have negative associations with the preference to remain close to home, despite the research literature to suggest the opposite. However, there are interactions between race and ethnicity, and certain racial minority Latinx students—namely Black Latinx students—are more likely to have a stronger preference than their non-Latinx peers. Higher levels of academic achievement and math self-efficacy, indicators of college readiness, were associated with reduced levels of the preference. Of the social and cultural capital variables, higher levels of educational expectations were also found to be associated with lower levels of the preference, while college-relevant peer social capital and school motivation was associated with reduced levels of the preference.

Preferring to live close to home would likely have an impact with whether one chooses to attend college if there were no colleges near one's home community. However, the first study did not support this idea—while the preference to remain close to home seemed to be associated with college enrollment, this association became nonsignificant when controlling for other important college choice factors. This finding is important in that a desire to live near home is not in and of itself a hinderance to postsecondary engagement. On the other hand, the preference does seem to impact where one chooses to attend college; the preference remained significantly associated with the distance from home to the enrolled college after controlling for relevant college choice factors. Students who expressed a stronger desire to live near their communities were more likely to choose colleges that were in fact closer to home. The implications of these findings are important for students who are not proximate to high-quality college options. While students are not any less likely to enroll in college when they want to live near their homes, they are more likely to aim for colleges near home—which may be institutions that are lower quality or worse academic matches than those further from home. Future studies should explore whether students who prefer to remain close to home and choose

to enroll in college attend higher quality colleges, or colleges that are an appropriate academic match based on their precollege academic achievement.

The second study explored student mobility for FTB students, along with mobility differences between traditional and nontraditional students at their first institutions. Using a nationally representative data set of FTBs, the study examined the factors associated with the distance that students traveled to attend college for the first time. Across all FTBs nationwide, very few demographic characteristics, including race and ethnicity, gender, age, or FGCS status, were associated with student mobility once family characteristics, financial and economic capital, academic characteristics, local geography factors, and postsecondary institution characteristics were included in the predictive model. In other words, many of the background factors that are typically the focus of college access research was significantly tied to student mobility for college; instead, financial and economic capital, institution characteristics, and local geography were more important when predicting student mobility. For instance, students with higher financial resources, attending 4-year colleges, or living in more rural communities were more likely to attend distal colleges. The second study also explored differences between traditional and nontraditional students in their mobility patterns. Defining traditional students by age, recency of high school graduation, and private for-profit college attendance, traditional and nontraditional students under every definition significantly differed in the distance they traveled to attend college. Higher proportions of nontraditional students in the first two groups—age and high school graduation recency—attended college in the same CZ in which they lived, numbering at 72.7 and 70.4 percent, respectively. This was a large difference from nontraditional students from these same groups, of which 57.5 and 57.3 percent attended college in the same CZ, respectively. However, nontraditional students from the third group—private, for-profit institutional control—had high proportions of attending out-of-state colleges and fewer proportions of attending college in the same CZ as their traditional peers. Regardless of the definition used to examine traditional and nontraditional students, the analyses found different mobility patterns—both intrastate and interstate. Future studies should continue to explore how traditional and nontraditional students mobilize for college and whether they consider different institutional features when deciding whether to travel farther distances for college. These studies should also

focus on students attending for-profit colleges and examine how these institutions are able to attract students from distal states.

The final study examined college choice sets and whether the average distance to colleges in each student's choice set could be predicted by factors identified in the college access literature. Further, the study explored whether a vector of institutional characteristics was significantly associated with the distance between high school and each college when accounting for student-level variance. Concurring with the findings in the second study related to factors tied to student mobility, few demographic characteristics were significantly associated with the distance to colleges in students' choice sets. Academics, social and cultural capital, family characteristics, and high school characteristics were more impactful across the different choice set distance metrics—mean distance, minimum distance, maximum distance, and choice set distance range. Using multilevel models that placed choice set colleges at level-1 and students at level-2, analyses identified many significant factors tied to distance. While accounting for student-level variance in distance to colleges, institution level, control, selectivity, HBCU status, total enrollment, the proportion of student population that identified as Latinx, in-state tuition and fee rates, and urban locale were all significantly associated with the distance between each student's high school and the colleges in their choice set. Institutional factors play an important role in predicting the distance between colleges and students even when controlling for which students are considering them in their choice sets. Future studies should continue to explore how students construct their choice sets and how wide of a net distance-wise they are willing to consider for college. The present study was hindered in that choice sets could only consist of three colleges max, and therefore could not fully explore the breadth of choice sets, as some students will consider much more than three colleges. Further, future studies should explore whether students who only include one college in their choice set differ from those that consider more institutions and if they ultimately attend farther institutions.

Building a Geospatial Conceptual Framework

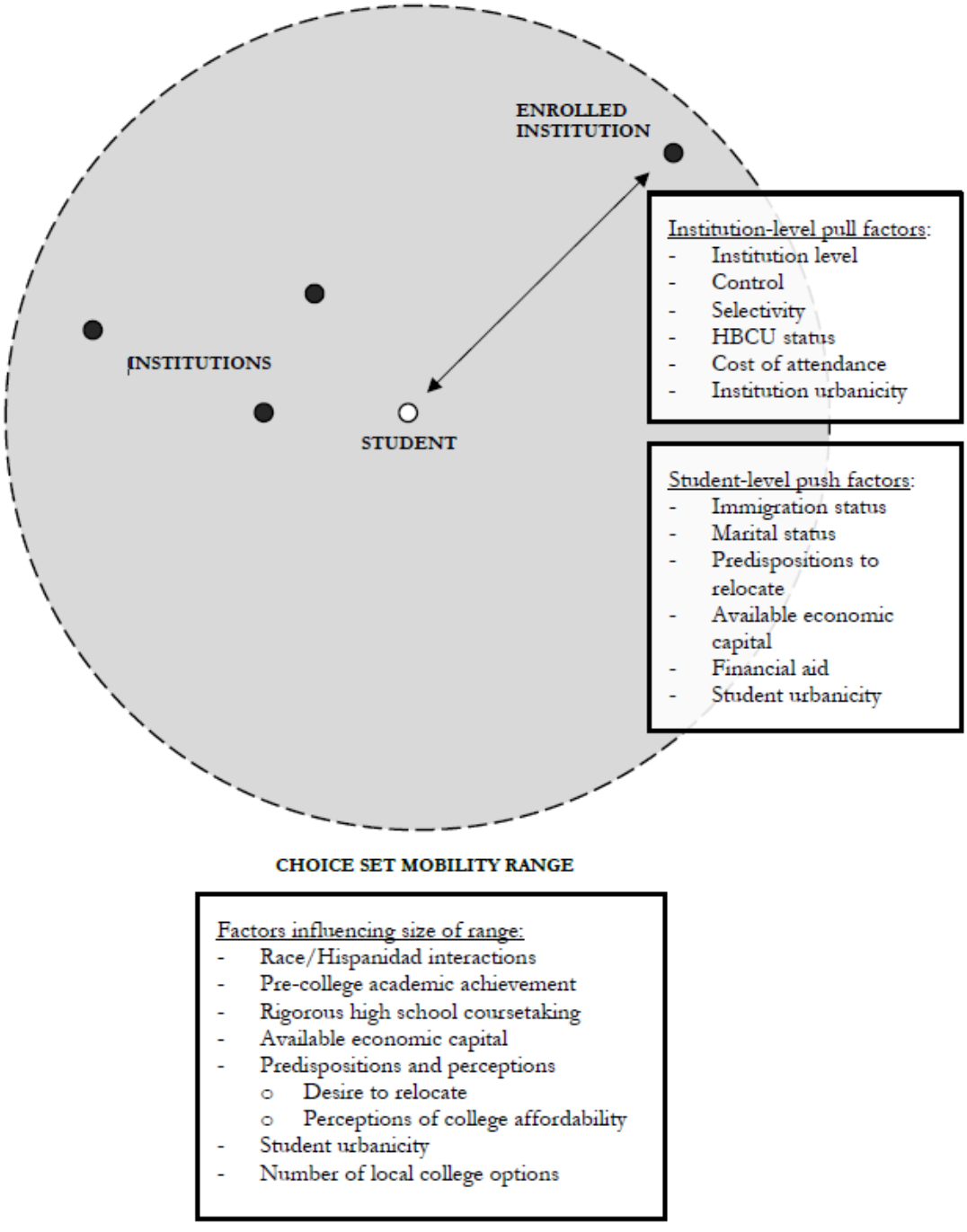
Finally, this dissertation will conclude by identifying the need and potential for a new conceptual framework that places mobility at the forefront of analysis. This dissertation—throughout the introduction and within the literature reviews of chapters 2, 3, and 4—has covered multiple topics: the research

foundations behind studies of college access and choice, descriptions of popular conceptual frameworks with a focus on process-oriented frameworks, research evidence behind the effects of different demographic, familial, community, and institutional contexts on college choice, and the research regarding geographic opportunities and the role of location on postsecondary enrollment patterns. What remains to be seen in the literature is an established conceptual framework that focuses the college decision-making process on geographic or geospatial factors. As many have argued, there is a need for geospatial factors to be the focus of decision-making (Hillman & Weichman, 2016; Turley, 2009). This is an important gap in the literature that needs to be filled with empirical work that mixes both the individual and contextual components of prominent process-based frameworks and the analyses on student migration patterns, "push" and "pull" factors from local and distal institutions, and improved measures of distance missing from the traditional college choice literature. The three studies included in this dissertation provide analyses that can serve as a foundation through which to design a conceptual framework in this manner.

Given the literature and empirical explorations described in this dissertation, the foundations for a geospatial framework of college choice can be built. Based on the findings contained herein, it is proposed that students construct college choice sets, and ultimately enroll in specific colleges, while relying in large part on the distance between them and their institutions. When constructing a college choice set, prospective college students generate a choice set mobility range—a theoretical or concrete distance range within which lie institutions that they are willing to relocate to or travel based on their proximity. Students determine this range based on a number of characteristics: the interaction between their race and Hispanidad, pre-college academic achievement, high school coursetaking, available economic capital, predispositions towards being mobile, perceptions of college affordability, urbanicity of their home community, and availability of local college options. High-achieving students, students with more available economic capital, students from rural communities, and those with fewer college options widen the choice set range; on the other hand, students with less exceptional academic achievement, students with lower financial resources, students living in more urban communities, and students with more local postsecondary opportunities will often constrict this choice set range. Figure 5.1 provides a visualization of the geospatial framework of college choice. In the figure, the

student (the white dot) is located at the center of their choice set mobility range, with institutions located throughout the choice set mobility range (represented as black dots).

Figure 5.1: Geospatial Framework of College Choice



In this conceptual framework, students will primarily consider colleges within their choice set mobility range. Characteristics at the student-level serve as "push" factors, which drive them towards their postsecondary options, while characteristics at the institution level that attract students towards them are "pull" factors. The distance between the student and the institution they choose to enroll in are impacted by these push and pull factors, where varying levels of each will drive students towards or away from home. The push and pull factors are generated from both the literature and the findings within this dissertation.

Of course, institutions exist beyond the choice set mobility range, and students will often consider (and even enroll in) colleges they never thought they would consider based on how distal they were from home. The permeability of the choice set mobility range is determined by a student's distance elasticity. As mentioned in the literature review in chapter 2, distance elasticity refers to sensitivity of college choices and college enrollment related to the distance between them and the institutions in question (Hillman 2016; 2017). In the framework, students who are distance inelastic will either consider institutions outside their range more strongly or extend their range much farther than their peers, whereas distance elastic students will have that range be less permeable—and thus have the dashed circle representing the choice set mobility range in Figure 5.1 become a more solid border. Institutions outside of each student's choice range can still attract that student and pull factors can draw the institution into the choice set by expanding the choice set mobility range.

New Directions for Future Research

The geospatial framework of college choice is in its infant stage as proposed here. While this dissertation drew heavily from the available research and provided analyses that filled extant gaps in the research body, there is still more to explore and solidify in creating a conceptual framework that positions geospatial factors above others. Future research projects can continue to explore this framework in several ways. First, projects that focus on larger choice sets than those explored here can confirm, reject, or expand the findings used to create this framework and further develop interpretations of what impacts the size of the choice set mobility range. Further, research can explore the process through which students generate college choice sets while focusing on distance and examining qualitatively how distance is considered during choice

set creation. Differences between various cultural groups—particularly racially and ethnically diverse groups—and how they decide on their choice set mobility range is ripe ground for future research, as this project did not fully explore cultural capital and cultural differences in relation to student mobility. Study one and three found that there were significant interactions between race and ethnicity, with Hispanic racial minority students differing from their non-Hispanic peers—the reasons behind these differences remain available for exploration.

REFERENCES

Hillman, N. W. (2016). Geography of college opportunity: The case of education deserts. *American Educational Research Journal*, *53*(4), 987–1021. https://doi.org/10.3102/0002831216653204

Hillman, N. W. (2017). Geospatial analysis in higher education research. In M. B. Paulsen (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 32, pp. 529–575). Springer International Publishing. https://doi.org/10.1007/978-3-319-48983-4_11

Hillman, N., & Weichman, T. (2016). Education deserts: The continued significance of "place" in the twenty-first century. *Viewpoints: Voices from the Field*.

Turley, R. N. L. (2009). College proximity: Mapping access to opportunity. *Sociology of Education*, *82*(2), 126–146. https://doi.org/10.1177/003804070908200202

APPENDIX 3.1: FREQUENCY STATISTICS FOR CATEGORICAL ANALYTIC VARIABLES FOR TRADITIONAL-NONTRADITIONAL STUDENTS—AGE WHEN ENTERING COLLEGE

	Age when entering college					
	Traditional			Nontraditional		
Variable	**N**	**Unweighted percent**	**Weighted percent**	**N**	**Unweighted percent**	**Weighted percent**
Race (Census Categories)						
White	7,770	67.18	69.33	2,960	64.69	67.28
Black or African American	1,920	16.63	14.95	1,100	24.01	21.8
Asian	670	5.82	6.45	130	2.82	3.78
American Indian or Alaska Native	360	3.08	2.63	120	2.69	2.34
Native Hawaiian/other Pacific Islander	220	1.92	1.61	60	1.27	1.02
More than one race	620	5.38	5.03	210	4.52	3.78
Hispanic or Latino origin						
Non-Hispanic or Latino	9,050	78.2	80.61	3,770	82.3	82.74
Hispanic or Latino	2,520	21.8	19.39	810	17.7	17.26
Gender						
Male	4,940	42.67	43.88	2,110	46.14	48.25
Female	6,630	57.33	56.12	2,470	53.86	51.75
Disability status: Has long-lasting disability or condition in 2011-12						
No	10,440	90.26	90.34	3,820	83.37	82.91
Yes	1,130	9.74	9.66	760	16.63	17.09
First-generation college student						
No	7,920	68.44	71.35	2,000	43.72	47.7
Yes	3,650	31.56	28.65	2,580	56.28	52.3
Had sibling who attended college first						
No	6,230	53.85	52.23	2,530	55.23	52.46
Yes	5,340	46.15	47.77	2,050	44.77	47.54
Immigrant generation status						
First generation immigrant	770	6.65	6.95	400	8.72	12.01
Second generation immigrant	2,620	22.61	21.37	580	12.69	12.22

See notes at end of table.

Variable	Age when entering college: Traditional N	Traditional Unweighted percent	Traditional Weighted percent	Nontraditional N	Nontraditional Unweighted percent	Nontraditional Weighted percent
Third generation immigrant or higher	8,190	70.74	71.68	3,600	78.59	75.76
Marital status in 2011-12						
Single, divorced, separated, or widowed	11,530	99.66	99.69	3,590	78.37	76.02
Married	40	0.34	0.31	990	21.63	23.98
Dependency status in 2011-12						
Dependent student	10,850	93.8	94.53	1,100	24.03	28.09
Independent student	720	6.2	5.47	3,480	75.97	71.91
Applied for any financial aid in 2011-12						
No	770	6.63	10.92	170	3.71	9.89
Yes	10,800	93.37	89.08	4,410	96.29	90.11
Attendance intensity pattern in 2011-12						
Full-time	8,570	74.02	69.88	2,930	64.08	48.05
Part-time or mixed	3,010	25.98	30.12	1,640	35.92	51.95
Took online, night, or weekend classes at first institution						
No	7,550	65.22	66.69	2,190	47.85	50.29
Yes	4,020	34.78	33.31	2,390	52.15	49.71
Took remedial courses in 2011-12						
No	8,930	77.18	78.14	3,380	73.74	68.96
Yes	2,640	22.82	21.86	1,200	26.26	31.04
Recent (2011) high school graduate						
No	1,270	11	9.23	4,140	90.39	89.72
Yes	10,300	89	90.77	440	9.61	10.28
Student urbanicity						
City	3,940	34.02	33.63	1,940	42.39	42.85
Suburb	5,480	47.33	47.49	1,640	35.83	34.96
Town	790	6.86	6.75	440	9.7	10.67
Rural	1,360	11.78	12.14	550	12.08	11.52
Institution level						
4-year	7,620	65.85	62.14	2,570	56.08	33.95
At least 2-year but less than 4-year	3,950	34.14	37.86	2,010	43.85	65.9

See notes at end of table.

Variable	Age when entering college: Traditional N	Traditional Unweighted percent	Traditional Weighted percent	Nontraditional N	Nontraditional Unweighted percent	Nontraditional Weighted percent
Less-than-2-year	#	0.01	#	#	0.07	0.15
Institution control						
Public	6,590	56.91	75.91	1,820	39.65	70.8
Private, not-for-profit	2,870	24.77	20.72	240	5.33	4.86
Private, for-profit	2,120	18.32	3.37	2,520	55.01	24.33
Institution selectivity						
Open admission or not 4-year institution	4,880	42.13	45.04	3,330	72.69	86.23
Minimally selective	1,950	16.82	7.37	1,140	24.95	9.46
Moderately selective	2,670	23.09	28.62	80	1.7	3.2
Very selective	2,080	17.96	18.98	30	0.66	1.11
Institution urbanicity						
City	6,420	55.52	55.37	2,570	56.15	53.24
Suburb	2,930	25.28	24.78	1,180	25.85	24.42
Town	1,250	10.8	12.31	250	5.35	8.77
Rural	970	8.4	7.54	580	12.65	13.56

Rounds to zero.
NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

APPENDIX 3.2: SUMMARY STATISTICS FOR CONTINUOUS ANALYTIC VARIABLES FOR TRADITIONAL-NONTRADITIONAL STUDENTS—AGE WHEN ENTERING COLLEGE

	Age when entering college											
			Traditional						Nontraditional			
			Unweighted		Weighted				Unweighted		Weighted	
Variable	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**
Dependent Variable												
Distance from home to college attended in 2011-12	1.00	3,762.00	138.38	357.24	137.16	5.06	1.00	4,390.00	210.76	491.29	176.95	13.97
Log Distance	0.00	8.23	3.21	1.80	3.23	0.03	0.00	8.39	3.01	2.06	2.77	0.06
Independent Variables												
Age in 2011-12	17.00	19.00	18.31	0.50	18.31	0.01	15.00	75.00	28.01	8.92	27.95	0.30
Number of Student's Dependents	0.00	7.00	0.05	0.34	0.03	0.00	0.00	10.00	0.98	1.33	0.86	0.04
Household Size	1.00	13.00	4.01	1.47	4.04	0.02	1.00	12.00	2.88	1.63	2.94	0.05
Direct subsidized and Unsubsidized loans (in thousands)	0.00	23.54	3.18	3.95	2.15	0.06	0.00	22.00	4.97	5.18	2.73	0.15
Adjusted gross income in 2011-12 (in thousands)	0.00	1,000.00	67.88	72.31	73.95	0.95	0.00	500.00	23.28	28.78	29.17	1.22
Total federal Title IV aid received (in thousands)	0.00	69.50	7.53	8.41	5.77	0.16	0.00	53.15	7.74	5.79	4.98	0.25
Amount of institutional need-based grants received (in thousands)	0.00	49.46	1.51	4.99	1.38	0.05	0.00	43.15	0.19	1.83	0.20	0.05
Amount of institutional non-need-based grants received (in thousands)	0.00	50.00	1.88	5.06	1.69	0.07	0.00	40.00	0.15	1.32	0.15	0.03
Amount of institutional tuition and fee waivers received (in thousands)	0.00	40.00	0.16	1.94	0.16	0.02	0.00	12.10	0.02	0.37	0.02	0.01

See notes at end of table.

	Age when entering college											
			Traditional						Nontraditional			
			Unweighted		Weighted				Unweighted		Weighted	
Variable	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**
Net price after grants as percentage of income	0.00	100.00	40.12	33.12	33.36	0.48	0.00	100.00	67.13	35.00	53.29	1.27
Veterans' benefits amount received (in thousands)	0.00	21.07	0.00	0.21	0.00	0.00	0.00	54.18	0.99	5.04	0.61	0.08
In-state tuition for full-time undergraduates (in thousands)	0.00	44.22	12.65	11.57	10.10	0.15	0.00	43.09	10.07	6.58	6.50	0.21
In-state fees for full-time undergraduates (in thousands)	0.00	10.90	0.80	1.07	0.90	0.03	0.00	9.69	0.40	0.53	0.47	0.03
Census tract: median household income (in thousands)	7.08	250.00	63.90	30.77	66.11	0.45	4.51	185.90	49.18	20.82	50.44	0.77
Census tract: percent White	0.00	100.00	71.40	25.37	73.13	0.43	0.00	100.00	67.67	27.18	69.05	0.93
Census tract: percent Black	0.00	100.00	13.52	21.51	12.16	0.32	0.00	100.00	19.59	26.18	18.12	0.84
Census tract: percent Asian	0.00	88.36	5.82	9.83	5.99	0.18	0.00	92.05	3.81	7.25	3.99	0.25
Census tract: percent American Indian/Alaska Native	0.00	97.19	0.62	2.85	0.64	0.04	0.00	80.29	0.74	2.93	0.91	0.15
Census tract: percent with a Bachelor's degree or higher	0.00	91.57	30.25	18.64	31.96	0.30	0.00	85.75	22.42	14.07	23.91	0.47
Census tract: percent unemployed	0.00	35.77	6.21	3.27	5.93	0.05	0.00	33.49	7.07	3.67	6.81	0.10
Commuting zone: number of postsecondary institutions	1.00	380.00	99.81	107.55	93.92	2.95	1.00	380.00	76.26	95.05	75.82	4.01
Commuting zone: number of 4-year institutions	0.00	141.00	41.73	43.04	39.21	1.14	0.00	141.00	31.75	37.29	31.57	1.57
Commuting zone: number of 2-year institutions	0.00	108.00	25.57	29.32	24.12	0.81	0.00	108.00	20.33	26.23	19.95	1.06
Commuting zone: number of less-than-2-year institutions	0.00	131.00	32.50	37.56	30.59	1.03	0.00	131.00	24.18	33.28	24.30	1.44

SD = Standard deviation.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

APPENDIX 3.3: FREQUENCY STATISTICS FOR CATEGORICAL ANALYTIC VARIABLES FOR TRADITIONAL-NONTRADITIONAL STUDENTS—COLLEGE ENROLLMENT TIMING

	College enrollment timing					
	Traditional			Nontraditional		
Variable	**N**	**Unweighted percent**	**Weighted percent**	**N**	**Unweighted percent**	**Weighted percent**
Race (Census Categories)						
White	7,190	67	69.25	3,540	65.43	67.89
Black or African American	1,810	16.85	15.03	1,210	22.44	20.22
Asian	640	5.98	6.59	160	2.96	3.91
American Indian or Alaska Native	330	3.04	2.56	150	2.83	2.6
Native Hawaiian/other Pacific Islander	200	1.82	1.51	90	1.57	1.43
More than one race	570	5.33	5.06	260	4.77	3.96
Hispanic or Latino origin						
Non-Hispanic or Latino	8,400	78.26	80.31	4,410	81.55	83.16
Hispanic or Latino	2,330	21.74	19.69	1,000	18.45	16.84
Gender						
Male	4,610	42.97	43.92	2,440	45.01	47.26
Female	6,120	57.03	56.08	2,980	54.99	52.74
Disability status: Has long-lasting disability or condition in 2011-12						
No	9,710	90.43	90.27	4,550	84.1	84.57
Yes	1,030	9.57	9.73	860	15.9	15.43
First-generation college student						
No	7,380	68.72	71.1	2,540	46.97	53.07
Yes	3,360	31.28	28.9	2,870	53.03	46.93
Had sibling who attended college first						
No	5,740	53.48	51.86	3,020	55.75	53.47
Yes	5,000	46.52	48.14	2,390	44.25	46.53
Immigrant generation status						
First generation immigrant	740	6.85	7.14	430	8	10.48
Second generation immigrant	2,430	22.61	21.54	770	14.21	13.56

See notes at end of table.

Variable	College enrollment timing: Traditional N	Traditional Unweighted percent	Traditional Weighted percent	Nontraditional N	Nontraditional Unweighted percent	Nontraditional Weighted percent
Third generation immigrant or higher	7,570	70.53	71.32	4,210	77.78	75.96
Marital status in 2011-12						
Single, divorced, separated, or widowed	10,690	99.55	99.66	4,430	81.87	80.8
Married	50	0.45	0.34	980	18.13	19.2
Dependency status in 2011-12						
Dependent student	10,010	93.2	94.41	1,950	35.97	41.55
Independent student	730	6.8	5.59	3,460	64.03	58.45
Applied for any financial aid in 2011-12						
No	710	6.59	10.68	230	4.23	10.75
Yes	10,030	93.41	89.32	5,180	95.77	89.25
Attendance intensity pattern in 2011-12						
Full-time	7,990	74.41	70.82	3,510	64.84	49.67
Part-time or mixed	2,750	25.59	29.18	1,900	35.16	50.33
Took online, night, or weekend classes at first institution						
No	7,080	65.91	67.16	2,660	49.17	52.22
Yes	3,660	34.09	32.84	2,750	50.83	47.78
Took remedial courses in 2011-12						
No	8,270	76.98	77.92	4,040	74.66	71.4
Yes	2,470	23.02	22.08	1,370	25.34	28.6
Recent (2011) high school graduate						
No	†	†	†	5,410	100	100
Yes	10,740	100	100	†	†	†
Student urbanicity						
City	3,650	34.02	33.62	2,220	41.11	41.03
Suburb	5,090	47.36	47.46	2,030	37.54	37.53
Town	730	6.82	6.66	510	9.35	10.13
Rural	1,270	11.8	12.25	650	12	11.31
Institution level						
4-year	7,040	65.58	62.71	3,150	58.13	37.88
At least 2-year but less than 4-year	3,700	34.41	37.29	2,260	41.81	62

See notes at end of table.

Variable	College enrollment timing					
	Traditional			Nontraditional		
	N	Unweighted percent	Weighted percent	N	Unweighted percent	Weighted percent
Less-than-2-year	#	0.01	#	#	0.06	0.12
Institution control						
Public	6,220	57.9	75.74	2,180	40.35	72.29
Private, not-for-profit	2,750	25.64	21.24	360	6.6	6.51
Private, for-profit	1,770	16.46	3.02	2,870	53.05	21.2
Institution selectivity						
Open admission or not 4-year institution	4,510	41.97	44.43	3,700	68.3	79.82
Minimally selective	1,680	15.67	7.23	1,410	25.97	9.45
Moderately selective	2,540	23.66	28.97	210	3.86	7.22
Very selective	2,010	18.69	19.37	100	1.87	3.52
Institution urbanicity						
City	5,920	55.15	55.4	3,070	56.78	53.59
Suburb	2,710	25.27	24.61	1,390	25.77	24.98
Town	1,190	11.09	12.45	300	5.62	9.08
Rural	910	8.48	7.54	640	11.83	12.35

† Not applicable.
# Rounds to zero.
NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

APPENDIX 3.4: SUMMARY STATISTICS FOR CONTINUOUS ANALYTIC VARIABLES FOR TRADITIONAL-NONTRADITIONAL STUDENTS—COLLEGE ENROLLMENT TIMING

	College enrollment timing											
			Traditional						Nontraditional			
			Unweighted		Weighted				Unweighted		Weighted	
Variable	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**
Dependent Variable												
Distance from home to college attended in 2011-12	1.00	3,762.00	137.53	355.13	137.07	5.53	1.00	4,390.00	201.31	476.76	169.35	11.57
Log Distance	0.00	8.23	3.22	1.80	3.24	0.03	0.00	8.39	3.03	2.03	2.83	0.06
Independent Variables												
Age in 2011-12	15.00	57.00	18.45	1.37	18.41	0.03	15.00	75.00	26.25	8.98	25.79	0.29
Number of Student's Dependents	0.00	8.00	0.06	0.38	0.04	0.00	0.00	10.00	0.82	1.27	0.68	0.03
Household Size	1.00	13.00	4.01	1.47	4.05	0.02	1.00	12.00	3.06	1.67	3.15	0.05
Direct subsidized and Unsubsidized loans (in thousands)	0.00	23.39	3.08	3.86	2.15	0.06	0.00	23.54	4.90	5.11	2.62	0.14
Adjusted gross income in 2011-12 (in thousands)	0.00	1,000.00	68.58	72.95	74.25	0.94	0.00	873.76	28.75	38.35	37.14	1.50
Total federal Title IV aid received (in thousands)	0.00	60.09	7.21	7.76	5.79	0.16	0.00	69.50	8.34	7.71	5.08	0.25
Amount of institutional need-based grants received (in thousands)	0.00	49.46	1.55	5.07	1.42	0.06	0.00	43.15	0.30	2.27	0.33	0.05
Amount of institutional non-need-based grants received (in thousands)	0.00	50.00	1.93	5.12	1.71	0.07	0.00	40.00	0.32	2.06	0.39	0.05
Amount of institutional tuition and fee waivers received (in thousands)	0.00	40.00	0.17	1.99	0.16	0.02	0.00	19.00	0.03	0.54	0.03	0.01

See notes at end of table.

	College enrollment timing											
			Traditional						Nontraditional			
			Unweighted		Weighted				Unweighted		Weighted	
Variable	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**
Net price after grants as percentage of income	0.00	100.00	39.61	32.86	33.24	0.46	0.00	100.00	63.99	35.88	49.71	1.16
Veterans' benefits amount received (in thousands)	0.00	33.19	0.01	0.52	0.00	0.00	0.00	54.18	0.82	4.60	0.48	0.06
In-state tuition for full-time undergraduates (in thousands)	0.00	44.22	12.66	11.73	10.24	0.16	0.00	43.09	10.45	7.11	6.84	0.21
In-state fees for full-time undergraduates (in thousands)	0.00	10.90	0.81	1.09	0.91	0.03	0.00	10.90	0.43	0.59	0.54	0.03
Census tract: median household income (in thousands)	7.08	250.00	64.12	31.01	66.18	0.48	4.51	250.00	51.01	22.37	53.32	0.69
Census tract: percent White	0.00	100.00	71.26	25.54	72.94	0.45	0.00	100.00	68.52	26.66	70.40	0.78
Census tract: percent Black	0.00	100.00	13.72	21.81	12.44	0.35	0.00	100.00	18.25	25.17	16.15	0.66
Census tract: percent Asian	0.00	88.36	5.80	9.85	5.92	0.18	0.00	92.05	4.16	7.71	4.59	0.32
Census tract: percent American Indian/Alaska Native	0.00	97.19	0.62	3.00	0.66	0.06	0.00	95.33	0.72	2.58	0.79	0.08
Census tract: percent with a Bachelor's degree or higher	0.00	91.57	30.35	18.68	31.93	0.30	0.00	85.75	23.43	14.94	25.57	0.46
Census tract: percent unemployed	0.00	35.77	6.20	3.29	5.93	0.05	0.00	33.49	6.96	3.60	6.64	0.09
Commuting zone: number of postsecondary institutions	1.00	380.00	99.17	106.39	93.50	2.83	1.00	380.00	81.15	100.19	80.58	3.97
Commuting zone: number of 4-year institutions	0.00	141.00	41.55	42.75	39.04	1.10	0.00	141.00	33.65	39.11	33.54	1.53
Commuting zone: number of 2-year institutions	0.00	108.00	25.33	28.91	23.94	0.77	0.00	108.00	21.63	27.71	21.27	1.06
Commuting zone: number of less-than-2-year institutions	0.00	131.00	32.29	37.13	30.51	0.99	0.00	131.00	25.88	35.10	25.77	1.43

SD = Standard deviation.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

APPENDIX 3.5: FREQUENCY STATISTICS FOR CATEGORICAL ANALYTIC VARIABLES FOR TRADITIONAL-NONTRADITIONAL STUDENTS—INSTITUTIONAL CONTROL

	Institutional control					
	Traditional			Nontraditional		
Variable	**N**	**Unweighted percent**	**Weighted percent**	**N**	**Unweighted percent**	**Weighted percent**
Race (Census Categories)						
White	7,960	69.14	69.68	2,780	59.85	59.63
Black or African American	1,860	16.13	15.41	1,170	25.14	27.91
Asian	680	5.89	6.2	120	2.67	2.24
American Indian or Alaska Native	280	2.42	2.5	200	4.31	3.37
Native Hawaiian/other Pacific Islander	170	1.43	1.48	120	2.48	1.64
More than one race	570	4.98	4.74	260	5.54	5.2
Hispanic or Latino origin						
Non-Hispanic or Latino	9,430	81.89	81.35	3,390	73.09	77.55
Hispanic or Latino	2,080	18.11	18.65	1,250	26.91	22.45
Gender						
Male	4,980	43.28	45.18	2,070	44.59	40.18
Female	6,530	56.72	54.82	2,570	55.41	59.82
Disability status: Has long-lasting disability or condition in 2011-12						
No	10,330	89.75	89.27	3,930	84.73	83.05
Yes	1,180	10.25	10.73	710	15.27	16.95
First-generation college student						
No	7,800	67.79	68.4	2,120	45.64	42.6
Yes	3,710	32.21	31.6	2,520	54.36	57.4
Had sibling who attended college first						
No	6,050	52.52	51.71	2,710	58.52	59.09
Yes	5,470	47.48	48.29	1,920	41.48	40.91
Immigrant generation status						
First generation immigrant	830	7.24	8.17	340	7.24	6.13
Second generation immigrant	2,250	19.5	19.78	950	20.53	15.7

See notes at end of table.

Variable	Institutional control					
	Traditional			Nontraditional		
	N	Unweighted percent	Weighted percent	N	Unweighted percent	Weighted percent
Third generation immigrant or higher	8,430	73.26	72.06	3,350	72.23	78.17
Marital status in 2011-12						
Single, divorced, separated, or widowed	11,050	96	95.58	4,070	87.73	84.81
Married	460	4	4.42	570	12.27	15.19
Dependency status in 2011-12						
Dependent student	9,600	83.42	84.07	2,350	50.71	39.75
Independent student	1,910	16.58	15.93	2,290	49.29	60.25
Applied for any financial aid in 2011-12						
No	910	7.88	11.53	30	0.65	0.85
Yes	10,600	92.12	88.47	4,610	99.35	99.15
Attendance intensity pattern in 2011-12						
Full-time	7,870	68.38	64.47	3,630	78.2	75.34
Part-time or mixed	3,640	31.62	35.53	1,010	21.8	24.66
Took online, night, or weekend classes at first institution						
No	7,540	65.51	64.76	2,200	47.37	45.42
Yes	3,970	34.49	35.24	2,440	52.63	54.58
Took remedial courses in 2011-12						
No	8,590	74.6	75.62	3,720	80.19	83.42
Yes	2,920	25.4	24.38	920	19.81	16.58
Recent (2011) high school graduate						
No	2,540	22.07	22.26	2,870	61.88	71.24
Yes	8,970	77.93	77.74	1,770	38.12	28.76
Student urbanicity						
City	3,900	33.89	35.04	1,980	42.6	41.66
Suburb	5,250	45.61	45.6	1,870	40.25	36.13
Town	870	7.57	7.24	370	7.91	11.41
Rural	1,490	12.93	12.11	430	9.23	10.79
Institution level						
4-year	6,120	53.14	54.18	4,070	87.77	80.7
At least 2-year but less than 4-year	5,390	46.86	45.82	560	12.14	18.89

See notes at end of table.

Variable	Institutional control					
	Traditional			Nontraditional		
	N	Unweighted percent	Weighted percent	N	Unweighted percent	Weighted percent
Less-than-2-year	#	#	#	#	0.09	0.42
Institution control						
Public	8,400	72.98	81.13	†	†	†
Private, not-for-profit	3,110	27.02	18.87	†	†	†
Private, for-profit	†	†	†	4,640	100	100
Institution selectivity						
Open admission or not 4-year institution	6,120	53.21	52.83	2,080	44.8	63.44
Minimally selective	610	5.32	5.45	2,480	53.39	35.78
Moderately selective	2,670	23.16	25.19	80	1.81	0.78
Very selective	2,110	18.31	16.52	#	#	#
Institution urbanicity						
City	5,940	51.62	54.16	3,050	65.8	64.05
Suburb	2,970	25.8	25.09	1,140	24.54	20.15
Town	1,420	12.29	12.09	80	1.72	5.38
Rural	1,180	10.28	8.66	370	7.93	10.42

† Not applicable.
# Rounds to zero.
NOTE: Counts rounded to the nearest 10. Percentages reflect unrounded counts. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).

APPENDIX 3.6: SUMMARY STATISTICS FOR CONTINUOUS ANALYTIC VARIABLES FOR TRADITIONAL-NONTRADITIONAL STUDENTS—INSTITUTIONAL CONTROL

	Institutional control											
			Traditional						Nontraditional			
			Unweighted		Weighted				Unweighted		Weighted	
Variable	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**
Dependent Variable												
Distance from home to college attended in 2011-12	1.00	3,762.00	122.10	335.59	121.12	5.20	1.00	4,390.00	250.21	518.74	435.22	20.37
Log Distance	0.00	8.23	3.05	1.78	3.05	0.03	0.00	8.39	3.41	2.08	4.10	0.09
Independent Variables												
Age in 2011-12	15.00	75.00	19.96	5.31	19.91	0.16	16.00	69.00	23.79	8.08	25.36	0.27
Number of Student's Dependents	0.00	10.00	0.18	0.68	0.15	0.01	0.00	9.00	0.64	1.17	0.82	0.05
Household Size	1.00	13.00	3.87	1.56	3.87	0.03	1.00	12.00	3.24	1.62	3.14	0.05
Direct subsidized and Unsubsidized loans (in thousands)	0.00	23.39	2.24	2.96	1.94	0.06	0.00	23.54	7.29	5.26	6.19	0.26
Adjusted gross income in 2011-12 (in thousands)	0.00	1,000.00	65.60	72.48	67.87	1.15	0.00	500.00	29.52	36.10	25.49	0.88
Total federal Title IV aid received (in thousands)	0.00	60.09	5.90	6.91	5.23	0.16	0.00	69.50	11.78	8.15	10.01	0.33
Amount of institutional need-based grants received (in thousands)	0.00	49.46	1.56	5.11	1.22	0.05	0.00	21.60	0.06	0.62	0.07	0.02
Amount of institutional non-need-based grants received (in thousands)	0.00	50.00	1.92	5.11	1.48	0.06	0.00	17.82	0.08	0.72	0.07	0.02
Amount of institutional tuition and fee waivers received (in thousands)	0.00	40.00	0.16	1.92	0.14	0.02	0.00	17.82	0.03	0.57	0.03	0.01

See notes at end of table.

	Institutional control											
			Traditional						Nontraditional			
			Unweighted		Weighted				Unweighted		Weighted	
Variable	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Mean**	**SD**
Net price after grants as percentage of income	0.00	100.00	37.26	32.24	34.43	0.45	0.00	100.00	73.87	30.45	74.46	1.09
Veterans' benefits amount received (in thousands)	0.00	33.19	0.06	0.98	0.08	0.02	0.00	54.18	0.82	4.80	0.67	0.10
In-state tuition for full-time undergraduates (in thousands)	0.00	44.22	10.50	11.93	8.92	0.15	4.80	28.00	15.44	3.24	14.46	0.16
In-state fees for full-time undergraduates (in thousands)	0.00	10.90	0.82	1.09	0.85	0.03	0.00	2.71	0.35	0.37	0.43	0.04
Census tract: median household income (in thousands)	7.08	250.00	63.02	30.89	63.95	0.50	4.51	178.08	51.56	21.93	49.45	0.49
Census tract: percent White	0.00	100.00	72.89	24.62	72.90	0.45	0.00	100.00	64.04	28.02	64.91	0.85
Census tract: percent Black	0.00	100.00	13.32	21.18	12.79	0.36	0.00	100.00	19.99	26.67	20.72	0.82
Census tract: percent Asian	0.00	92.05	5.42	9.56	5.70	0.18	0.00	82.12	4.81	8.31	4.12	0.18
Census tract: percent American Indian/Alaska Native	0.00	95.33	0.62	2.86	0.70	0.05	0.00	97.19	0.73	2.88	0.68	0.05
Census tract: percent with a Bachelor's degree or higher	0.00	91.57	30.19	18.59	30.98	0.33	0.00	87.30	22.67	14.42	21.89	0.32
Census tract: percent unemployed	0.00	35.77	6.11	3.26	6.01	0.05	0.00	33.49	7.30	3.62	7.29	0.10
Commuting zone: number of postsecondary institutions	1.00	380.00	87.63	98.18	89.97	2.97	1.00	380.00	106.80	118.26	92.10	4.51
Commuting zone: number of 4-year institutions	0.00	141.00	36.99	39.99	37.58	1.15	0.00	141.00	43.65	45.43	37.95	1.72
Commuting zone: number of 2-year institutions	0.00	108.00	22.11	26.30	23.11	0.81	0.00	108.00	28.99	33.05	24.87	1.25
Commuting zone: number of less-than-2-year institutions	0.00	131.00	28.53	34.29	29.28	1.04	0.00	131.00	34.16	41.49	29.28	1.61

SD = Standard deviation.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17).