

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Educational Administration: Theses, Dissertations,
and Student Research

Educational Administration, Department of

Fall 11-21-2015

A CORRELATIONAL CASE STUDY ON DISTANCE FROM HOME AND ATTRITION OF FIRST-TIME, FULL-TIME STUDENTS

James L. Baldwin

University of Nebraska-Lincoln, jl20@pitt.edu

Follow this and additional works at: <http://digitalcommons.unl.edu/cehsedaddiss>



Part of the [Higher Education Administration Commons](#)

Baldwin, James L., "A CORRELATIONAL CASE STUDY ON DISTANCE FROM HOME AND ATTRITION OF FIRST-TIME, FULL-TIME STUDENTS" (2015). *Educational Administration: Theses, Dissertations, and Student Research*. 250.
<http://digitalcommons.unl.edu/cehsedaddiss/250>

This Article is brought to you for free and open access by the Educational Administration, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Educational Administration: Theses, Dissertations, and Student Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

A CORRELATIONAL CASE STUDY ON DISTANCE FROM HOME AND
ATTRITION OF FIRST-TIME, FULL-TIME STUDENTS

by

James L. Baldwin

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: Educational Studies
(Educational Leadership and Higher Education)

Under the Supervision of Professor Brent Cejda

Lincoln, Nebraska

November, 2015

A CORRELATIONAL CASE STUDY ON DISTANCE FROM HOME AND
ATTRITION OF FIRST-TIME, FULL-TIME STUDENTS

James L. Baldwin, Ph.D.

University of Nebraska, 2015

Advisor: Brent Cejda

As institutions face increasing demands to maintain or increase enrollments, colleges and universities have begun to recruit students from greater distances. The purpose of this *ex-post facto* case study was to determine the existence of a relationship between the institutional distance from home and the attrition of traditional-aged, first-time, full-time students prior to the second year at the University of Pittsburgh at Bradford, one of the four-year campuses of the University. Following the correlational analysis, further analysis was conducted to determine if a non-linear relationship existed between the institutional distance from home and attrition of first-year students prior to their second year when controlling for sex, race or ethnicity, expected family contribution (EFC), parental education levels (specifically, first-generation college student status), residency status, housing status, SAT or ACT score, or market segmentation as defined by the institution. The literature review showed that prior exploration of this relationship has been sporadic and the methodology used was incomplete.

The case study examined attrition from first to the second year for 2,837 domestic, traditional-aged, first-time, full-time students (freshmen) matriculated and enrolled during the fall semesters of 2005 through and including 2013 at the University of Pittsburgh at Bradford.

Three statistical approaches were used: point-biserial correlation, partial correlation, and binary logistic regression analysis. Results indicated that there is no significant relationship between the institutional distance from home and attrition.

This case study adds to the literature: A new methodology for measuring distance in miles and travel time was utilized. Furthermore, the results will help to inform the future use of the variable of institutional distance from home in future studies of attrition, retention, and the development of predictive models. The study has practical implications for admissions officers, orientation planners, first-year experience planners, student support services, and learning community practitioners. It is recommended that this study be replicated at other institutions to contribute to the enrollment management, retention, and attrition literature.

Acknowledgements

I want to thank the committee members, Dr. Wesley, Dr. Xia, and Dr. Steckelberg for their time, their suggestions, and their support with this process.

I want to thank my advisor, Dr. Cejda for his guidance, his advice, his support, and his help. You have been a great mentor to me, and I will always appreciate the attention and time that you gave me.

I want to thank my friends and colleagues who always had a kind and inspirational word to say.

I also want to thank my extended family members who were supportive and always willing to help if they could.

I'd like to thank my parents, Margaret and Thomas, who have always been supportive, caring, and loving. They were always rooting me on.

I want to thank my daughters, Lillian and Lauren for their love and understanding when Dada was so busy with his studies. I hope that I served as a good role model.

Finally, I want to thank my loving wife, Catherine, who believed in me, cheered me on, supported me, cared for me, and helped me to make this happen. I could not and would not have done this without her.

Table of Contents

Chapter 1—Introduction	1
Background.....	5
Statement of the Problem.....	8
Purpose Statement.....	8
Conceptual Framework.....	9
Attrition Versus Retention	17
Research Questions.....	18
Definition of Terms.....	19
Assumptions.....	23
Delimitations.....	23
Limitations	23
Significance of Study.....	25
Chapter 2—Review of Literature.....	27
Retention and Persistence	27
Analysis of Distance on Retention and Persistence.....	35
Distance in Other Student Persistence Studies	51
Commute Time	55
Chapter 3—Methodology	56
Research Hypotheses	57
Research Design.....	60
Site of Research	62
Population and Sample	62

Institutional Review Board	64
Research Questions and Hypotheses Development.....	65
Primary research question.....	65
Hypothesis Setup	66
Secondary research questions	66
Hypothesis Setup	68
Data Elements	69
Source of Data and Collection Procedures	76
Google APIs.....	76
Missing Data	77
Descriptive Statistics.....	79
Data Analyses	83
Point-Biserial Correlation Process.....	85
Partial Correlation Process.....	86
Binary Logistic Regression Process.....	88
Statistical Analysis Limitations	92
Point-Biserial Correlation	92
Partial Correlation.....	93
Binary Logistic Regression.....	93
Summary	94
Chapter 4—Results of Data Analysis	97
Introduction.....	97
Point-Biserial Correlation.....	99

Point-Biserial Correlation Summary.....	107
Partial Correlation.....	108
Partial Correlation Summary	111
Binary Logistic Regression.....	112
Overall Evaluation of Models.....	114
Goodness-Of-Fit Statistics.....	115
The Chi-Square Goodness of Fit.....	117
Kendall’s Tau-b, Goodman-Kruskal’s Gamma, Somer’s D.....	117
Receiver Operating Characteristic (ROC) Curve	117
Odds Ratio	118
Statistical Tests of Individual Predictors	120
Binary Logistic Regression Summary	121
Results of the Research Questions.....	122
Chapter 5—Findings, Conclusions and Implications	126
Introduction.....	126
Discussion of Results.....	127
SAT/ACT Score.....	129
Expected Family Contribution (EFC).....	131
In-State Residency	132
Domestic Underrepresented Minority Students.....	134
On-campus Housing Status.....	136
Market Segmentation	137
Sex and First-Generation Status.....	138

Limitations	138
Implications for Practice	140
Dissemination	140
Marketing and Advertising	141
Admissions.....	144
Advising.....	148
Student Affairs and Student Activities	148
Recommendations for Further Research.....	149
Conclusion	154
References	156

List of Tables

Table 1	Description of Independent Predictor Variables.....	74
Table 2	Population Descriptive Statistics	79
Table 3	Race and Ethnicity: Descriptive Statistics	80
Table 4	Residency and Market: Descriptive Statistics.....	80
Table 5	Descriptive Statistics for Interval Independent Variables	81
Table 6	Proposed Analyses and Findings to be Reported.....	83
Table 7	Point-Biserial Correlation Matrix (Cumulative).....	99
Table 8	Point-Biserial Correlation Analyses	103
Table 9	Point-Biserial Correlation Analyses of DistanceT in 30-Minute Intervals (Cumulative)	105
Table 10	Partial Correlation Analyses of Distance x Attrition	109
Table 11	Tolerance and Variance Inflation Factor	113
Table 12	Binary Logistic Regression Results Incorporating DistanceT.....	118
Table 13	Binary Logistic Regression Results Incorporating DistanceD	118
Table 14	Binary Logistic Regression Results Incorporating DistanceF	119

Chapter 1

Introduction

In 2012, the retention of first-year students into the second year at all four-year institutions of higher education dropped to 72%, its lowest rate since 1988, and 4.8% below the high mark of 74.9% in 1991 (ACT, 2008; ACT, 2013a). Nearly four million first-year students who started in the fall of 2011 did not return for their second year (ACT, 2013a; Knapp, Kelly-Reid, & Ginder, 2012). Retention continues to be a challenge for higher education even after countless studies, the development of numerous theories and models, attempts at predictive modeling, remediation, and services designed to increase student success (Lenning, 1982; Tinto, 1993; Habley, Bloom, & Robbins, 2012; Seidman, 2012).

Institutions are committed to addressing the issue of retention. Increasing the retention rate of first-time, full-time students is a goal that was reported by 68.4% of four-year private institutions and 57.1% of four-year public institutions of higher education in 2012 (Noel-Levitz, 2012). However, recent and ongoing changes in demographics, public financing, and the attraction of online education make the goal of retention even more critical as many institutions struggle to maintain enrollments.

Enrollments in secondary schools, grades nine through twelve, are projected to increase only 3% between 2010 and 2021, and the number of high school graduates is projected to increase by only 2.5% between 2009 and 2012 (Hussar & Bailey, 2013). Increases are projected in parts of the Midwest and most of the West, Southwest, and Southeast. However, most of the states in the Northeast will see decreases with some areas seeing declines of between 5% and 15% (Prescott & Bransberger, 2012). The

National Center for Education Statistics (NCES) reported a decrease of 1.6% in post-secondary education enrollment from 2010 to fall 2011. It was the first decline in enrollments since 1996 (Knapp, Kelly-Reid, & Ginder 2012). Decreases occurred at public and for-profit institutions; only private institutions experienced an increase. The NCES projected a 15% increase in post-secondary enrollment between fall 2010 and fall 2012, which was less than half of the actual increase (46%) that occurred between 1996 and 2010 (Hussar & Bailey, 2013). The projected growth in post-secondary enrollment assumes significant increases in the post-secondary attendance rates for both traditional- and nontraditional-aged students.

The State Higher Education Executive Officers Association (SHEEO) reported that, when adjusted for inflation, education appropriations per FTE fell to a 25-year low in 2012. Conversely, tuition, as a percentage of educational revenue generated, increased 23.7% between 1987 and 2012 and reached an all-time high of 47%. Most recently, between 2011 and 2012, state and local support per student decreased by 7.9% (SHEEO, 2013). The decrease in state and local appropriations as a percentage of the overall operational budget means that institutions must rely more heavily on revenue generated from tuition. In this budgetary climate, failing to retain a student beyond the first year results in three years of net tuition revenue loss. In addition to lost tuition revenue, there may also be a loss in the revenue that would have been generated from fees associated with room, board, and other services.

Performance-based funding, the allocation of some state funding based on performance measures, has often been tied to enrollments. However, an increasing number of states are tying allocations to course completion, retention, and graduation

rates. As of February 2013, 16 states had some form of performance-based funding in place or were about to begin a transition to performance-based funding, and 19 other states were engaged in formal discussions regarding the practice (National Conference of State Legislatures, 2013).

The prevalent opinion is that retaining students is far more efficient and cost-effective than recruiting new students to replace those lost to attrition (Lovitts, 2001; Raisman, 2002; Schuh & Gansemer-Topf, 2012). The median cost to recruit a single undergraduate student in 2011 was \$2,185 at four-year private institutions and \$457 at four-year public institutions (Noel-Levitz, 2011). The costs associated with recruitment include travel, supplies, publications, marketing, staff salary and benefits, and consulting and vendor services. An additional associated cost is the institutional financial aid used to recruit and retain new students (both merit and need-based). Institutions that utilize a front-load model for merit or need-based institutional aid (i.e., provide the greatest amounts in the first year and replacing institutional money with other types of financial aid as a student progresses) stand to lose the most through attrition as institutional dollars are spent before other types of financial aid (e.g., loans) are utilized (Schuh & Gansemer-Topf, 2012). Other indirect costs associated with attrition include overhead costs for under-utilized facilities, less auxiliary revenue (e.g., housing, dining services, bookstore), public image, and the costs (i.e., actual costs of recruitment or severances, and morale) associated with the hiring or elimination of faculty and staff positions as enrollments change (Swail, 2004).

Massive open online courses (MOOCs) and other distance education endeavors pose new opportunities for those institutions with the infrastructure, resources, and

motivation to attempt online programming, and yet, they have the potential to create new competitive threats for many other institutions. Allen and Seaman (2013) surveyed 2,800 chief academic officers and reported that while only 2.6% of institutions were offering a MOOC in 2012, an additional 9.4% were planning to engage in a MOOC in the future. In 2012, 32% of students in higher education were enrolled in at least one online course, and 69.1% of chief academic officers reported that online education was critical to their institution's long-term enrollment strategy. This was the highest reported rate since data collection began in 2002. However, studies show that completion rates for online courses are 9%-25% lower than courses taught and delivered in a traditional classroom setting (Lorenzetti, 2002; Simpson, 2004; Wojciechowski & Palmer, 2005; Nora & Plazas Snyder, 2008), and the attrition rates for online programs often exceed 50% (Bauman, 2002; Lorenzetti, 2002). Completion rates for MOOCs are even lower. Early studies suggest that completion rates for MOOCs range from 5% to 13% (Koller, Ng, Do, & Chen, 2013; Jordan, 2013). Online education and, to a greater extent, MOOCs will continue to appeal to students who perceive the benefits of asynchronous learning at a lower cost, which is not location bound. Institutions will continue to develop and deliver a product that meets this need and appeal and may be willing to sacrifice completion rates, which may impact retention rates, for the ability to scale the enrollment capacities of these kinds of courses to a much larger audience.

These are just a few of the contemporary factors that make maintaining, let alone increasing, enrollments and retention more challenging. The competition among institutions for recruitment and retention will increase. In response, institutions will have to allocate more money to staffing, travel, increased marketing, more communications,

increased offers of merit and need-based institutional aid, additional student-success services, and student programming and activities just to maintain the size and quality of the incoming first-year class and overall enrollments. Are there overlooked factors that could better inform recruitment and help to identify students at greater risk for attrition?

Background

The University of Pittsburgh at Bradford (UPB) is a public, baccalaureate regional campus of the University of Pittsburgh. The retention of first-time, full-time students into their second year has been a concern. During the past ten years, the retention rate has been consistently between 67% and 73%, which is 5-7% less than the rates of the other baccalaureate regional campuses of the University of Pittsburgh (i.e., Greensburg and Johnstown), and 20% less than the rate at the Pittsburgh campus of the University of Pittsburgh (Baldwin, 2013).

UPB implemented a series of initiatives designed to affect student success and increase retention. Following the implementation of a Federal TRIO Programs Student Support Services grant, a redesign of the first-year seminar course, and development of a pre-registration process for all incoming first-year students, one-year retention of first-time, full-time students increased from 68.8% in 2006 to 71.3% in 2007. It increased again in 2008 (73%) and 2009 (73.2%). However, in fall of 2010, it dropped to 68% and remained at that level in the fall of 2011 and 2012, which was an eight-year low. During this period, there was one noticeable change in the profile of incoming students. Students comprising the first-time, full-time class were being recruited from greater distances from the campus, located in Bradford, Pennsylvania.

The primary market at the University of Pittsburgh at Bradford (UPB) comprises five Pennsylvania counties. Those counties include McKean County, where UPB is located, and the counties that surround McKean County. The secondary market includes counties in Pennsylvania, upper-New York, and one county in Ohio. The secondary market includes counties bordering the primary market within a radius of 98 to 178 miles. The tertiary market comprises all other potential recruitment areas beyond the primary and secondary markets. The average distance from home to UPB for students in the tertiary market is approximately 230 miles. The mean distance from institution to home for first-time, full-time students who matriculated at UPB in the fall of 2010 was 108 miles, and 39.2% of that incoming class lived at least 100 miles away from the campus. The two other regional campuses of the University of Pittsburgh are the University of Pittsburgh at Greensburg and the University of Pittsburgh at Johnstown. The mean distance from institution to home for the same populations at those campuses were 42 and 55 miles, respectively. At UPG, only 12.6% of that incoming class lived 100 or more miles from campus. At UPJ, only 18.2% lived 100 or more miles from campus. The median distance at UPB was 69.2 miles, at UPG it was 20.2 miles, and at UPJ, it was 47.5 miles. Nationally, the mean distance is 94 miles. ACT (2013b) reported that the median distance from home for all students who completed the ACT college readiness assessment was 51 miles.

A Potential Factor

ACT (2013b) reported a correlation between assessment score and the institutional distance from home (i.e., the higher the score, the farther a student's

institution was from home). ACT also noted that the median distance from home for first-generation college students was 24 miles.

The Higher Education Research Institute (HERI) (2007) reported that 18.7% of the incoming first-year class that completed the Cooperative Institutional Research Program (CIRP) in 2006 indicated that living near home was a “very important reason” when selecting a college. This was the highest percentage since the question was added to the survey in 1983. Additionally, HERI noted that greater proportions of first-generation college students were selecting higher education institutions within 50 miles of home. While the cost of attendance is often cited as one factor, research suggests that there are other factors that are leading more students to choose institutions that are closer to home. The greater the distance between a student’s home and their institution of higher education, the less likely the student can travel home or have friends and family travel to see them.

Howe and Strauss (2000) noted that the millennial generation (those born after 1981) have stronger and closer relationships with their parents than any previously polled generation. Parents and students both reported more collaborative decision making in the students’ lives, and both parents and students reported that they enjoyed the close, emotionally. Fry (2013) noted that data collected by the Pew Research Center showed that in 2012, 36% of people between the ages of 18 and 31 were living in their parents’ home. This rate is the highest recorded in over 40 years.

Finally, Sivak and Schoettle (2012) reported that the current generation of college-age teens is significantly less likely to be driving than the previous generation. Between 1983 and 2000, there was a 17% drop (from 87% to 70%) in the number of 19-

year-olds who had their driver's license. Historically, mobility and the ability to drive have been linked.

Statement of the Problem

The previous exploration and research of potential attributes related to attrition that could then be incorporated into the predictive modeling of student attrition and retention have rarely included the distance between the student's permanent home address and the post-secondary institution. In all studies reviewed to date, distance has been calculated based on a straight line from a student's permanent home address to the post-secondary institution.

Purpose Statement

The purpose of this case study was to determine if a relationship existed between student attrition prior to the second year and the institutional distance from students' permanent home addresses for domestic, first-time, full-time, traditional-aged students at the University of Pittsburgh at Bradford. A secondary purpose was to determine if the relationship continued to exist having controlled for the following variables: sex, race and ethnicity, expected family contribution (EFC), parental education level, residency status, housing status, SAT and ACT scores, and market segmentation, as defined by the institution. Demographic, socioeconomic, and academic variables have been routinely studied and tested in previous studies spanning the history of retention research (McNeely, 1940; Spady, 1970; Tinto, 1975; Bean, 1980; Astin, 1984; Pascarella, 1985; Cabrera, Nora, and Castaneda, 1992).

Conceptual Framework

The theoretical foundation of this case study is derived from research in which contributing factors to retention and attrition have been hypothesized, tested and analyzed using longitudinal data. In general, the purpose has been to better describe the patterns of student retention and attrition. However, a clear secondary purpose exists—the development of predictive models that could lead an institution to implement specific intervention strategies that would mitigate those detrimental factors correlated with attrition or to enhance those factors found to influence retention positively. Astin and Oseguera (2005) asserted that empirical studies of college student retention and the development of theories or models result in two pragmatic functions: prediction and control. Prediction is the estimation of a particular student’s chance of success (retention and degree completion). Control is an institution’s capacity to positively affect a student’s chance of completion.

The first significant, multi-institutional study of undergraduate “student mortality” (i.e., attrition) was conducted by McNeely (1940). His study examined demographic, social engagement, institutional characteristics, completion, and departure data for 15,535 freshmen enrolled at 25 universities during the fall of 1931. The study, “College Student Mortality,” was published by the U.S. Department of Interior and the Office of Education (Berger, Ramirez, & Lyons, 2012; Morrison & Silverman, 2012). McNeely (1940) found that 62.1% of the entering class studied did not persist for four consecutive years. The first-year to second-year retention rate was 79.1%. McNeely reported that the primary cause of mortality was dismissal due to academic failure. Other factors that McNeely studied were age, major, financial ability, place of lodging,

employment status, extra-curricular activity participation, the location of the institution, and distance from home. McNeely noted that mortality was lower among students who attended institutions in the same county as their permanent home address, and that mortality was higher for students who were from a different state than where the institution was located.

In his review of previous studies and literature regarding student retention, Summerskill (as cited in Berger, Ramirez, & Lyons, 2012) stated that most of the studies conducted between 1920 and 1960 focused on the relationship between personality characteristics and attrition or persistence. These early studies were largely descriptive, atheoretical, and did not generate hypotheses from models or theories to predict relationships among the variables studied (Terenzini, Pascarella, Theophilides, & Lorang, 1985). His significant contribution to the development of the theoretical framework of retention studies was to suggest that future research be based upon psychological and sociological theories.

However, beginning in 1970, the seminal, conceptual models that provide the theoretical framework for empirical study of retention were developed by Spady (1970), Tinto (1975), Bean (1980), Astin (1984), Pascarella (1985), and culminated with Cabrera, Nora, and Castaneda (1992) who developed a more comprehensive model that combined the previous models developed by Tinto and Bean (Morrison & Silverman, 2012).

Spady (1970) developed a sociological model of student dropout in higher education, which is widely recognized as the first significant step in developing a systematic research and theory-based foundation in the study of retention (Berger, Ramirez, & Lyons, 2012). Spady's model, informed by Durkheim's (1951) social theory

of suicide, underscored the interaction between student characteristics and campus environment and suggested that relationships between certain variables might be indicative of an individual (i.e., student) who was severing ties with a social system (i.e., institution). Spady proposed a model that included the following social process factors, which were likely correlated with attrition: “family and previous educational background, academic potential, normative congruence, friendship support, intellectual development, grade performance, social integration, satisfaction, and institutional commitment” (Spady, 1971, p. 38). Spady conducted a multiple regression analysis of the entering freshmen class of 1965 at the University of Chicago (n = 683) and found that the two strongest contributing factors that affected first-year attrition for men were grade performance (5.91%) and institutional commitment (2.52%). Institutional commitment was a measurement of the degree to which a student felt it was important to graduate from the University. Among women, institutional commitment (11.97%) and subcultural orientations (3.62%) were the strongest factors. Subcultural orientations, a part of normative congruence, were measurements of political orientation and attitude, extracurricular activities (specifically, athletic or political), and academics (field of major interest).

Tinto (1975) used Spady’s (1970) adaptation of Durkheim’s (1951) concepts as the source for the development of his interactionist theory of student departure and developed a more expansive connection to Durkheim’s work by applying the association of different types of suicides to different types and reasons for attrition. Tinto noted that certain distinctions were lacking: Attrition could be voluntary (e.g., a temporary leave of absence or transferring to a different institution) or involuntary (e.g., dismissal for

academic or behavioral reasons). Voluntary withdrawal from an institution was likely the result of a student's lack of academic (value congruence) and social (social support) integration into the institution. The greater the level of integration, the more likely that a student would persist. Tinto stated that his interactionist theory of student departure accounted for voluntary departure and not involuntary departure.

Tinto (1987) noted that his theory did not account for factors such as the student's external communities (e.g., family, neighborhood, and secondary school) as well as the organizational attributes of the institution in which a student had enrolled (e.g., selectivity, faculty-to-student ratios, institutional size, and institutional type and control). Tinto revised and expanded the scope of his initial work, and developed his theory of individual departure, incorporating ideas from the work of Arnold Van Gennep (1960).

Van Gennep (1960), in *The Rites of Passage*, described stages of transition that a person experiences from "youthful participation to full adult membership in society" (p. 92) in terms of "separation, transition, and incorporation" (p. 11). Tinto (1993), believed that these transitional stages (and the overall transitional process) could be applied to the incorporation of students within the college environment ("especially in the first year") (p. 94). Tinto believed that the stages of separation from a student's previous community could be described in the same terms that Van Gennep used (i.e., separation, transition, and incorporation) and that the failure to complete these stages would result in an early departure from an institution.

Additional research led Tinto (1993) to develop a longitudinal, explanatory model of departure. Tinto expanded upon the environmental conditions (i.e., intellectual and social integration) influencing departure to include "adjustment, difficulty, incongruence,

isolation, finances, learning, and external obligations or commitments” (p. 112).

However, Tinto stated that his theory “is not a systems model of departure” (p. 112).

Tinto warned that his theory should neither be used in a post-facto review of an individual’s departure nor does it explain non-voluntary withdraw from an institution (e.g., suspension or dismissal). It is meant to describe how interactions among different students within the various communities that make up an institution can affect the departure process. Tinto recognized that not all colleges and universities (like students) are the same, and, therefore, there cannot be a single, successful retention strategy.

More recently, Tinto (2000) has criticized studies and existing theories of student departure for failing to look at the link between the classroom experience and student persistence. Failure to engage in the smallest of the academic communities (i.e., the classroom) may result in a lack of institutional academic and social engagement. Tinto stated that this risk is greatest in the first year of college, and strategies such as improved pedagogy, learning communities, and cooperative teaching can increase classroom engagement.

Astin (1975, 1984, 1991, 1993) developed a theory of student involvement that was based upon “the amount of physical and psychological energy a student devotes to the academic experience” (1984, p. 397). In addition to the degree to which a student interacted with the subject matter, services, faculty, and institution, Astin also identified both personal (e.g., past grades, aspiration, study habits, parental level of education) and environmental (e.g., residence, employment, institutional characteristics) factors that were predictive of retention. Astin used longitudinal data that was collected by the Higher Education Research Institute (HERI) to test his model and found that academic

involvement, involvement with faculty members, and involvement with social peer groups were the three most important types of involvement affecting retention.

Bean (1980) developed a causal model of student departure, which placed greater emphasis on variables that were environmental in nature. Bean's model was based on the work of Price (1977) who had developed a model describing the factors that led to workplace turnover. Bean, critical of the use of Durkheim's social theory of suicide in the work of Spady and Tinto, respectively, developed a model that incorporated many more variables than had been used in previous models or theories. Bean identified attrition as the dependent variable; satisfaction and institutional commitment as the intervening variables; background variables including academic achievement and socioeconomic status; and organizational determinants including integration, grade-point average, goal commitment, and living on campus. Bean postulated that satisfaction, affected by the organizational determinants, affected the likelihood of attrition. Bean (1985) later revised his model to include three empirical findings after having tested 14 variables. Bean stated, "a student's peers are more important agents of socialization than are informal faculty contacts, that students may play a more active role in their socialization than previously thought, and that college grades seem more the product of selection than socialization" (p. 35). Bean noted that his model was consistent with his earlier studies and studies conducted by Tinto. The differences were mostly variable placement within larger categories, variable order, and the combining of certain variables that had previously been distinctly measured.

Bean's research extended beyond traditional-aged, first-time, full-time students. Bean and Metzner (1985) argued that social variables were predominant in most of the

prevailing theories, and the existing models did not adequately describe the attrition and retention behaviors of the nontraditional student. Bean and Metzner found that environmental factors such as finances, hours of employment, external encouragement, family responsibilities, and opportunities to transfer to a different institution, impacted nontraditional-aged students to a greater extent.

Pascarella (1985), building on the work of Astin, developed a general causal model consisting of five sets of variables that measured or described direct and indirect student change: student background and pre-college traits (e.g., aptitude, personality, previous school experiences); structural and organizational characteristics; institutional environments; interactions with agents of socialization (e.g., faculty, staff, and peers); and quality of student effort. Pascarella and Terenzini (2005) later utilized this model in the exploration of student retention and withdrawal from an institution.

Cabrera, Nora, and Castaneda (1992) developed a more comprehensive model that combined the previous models developed by Tinto and Bean (Morrison & Silverman, 2012). Cabrera et al. incorporated additional variables regarding student finances after having analyzed the models developed by both Bean and Tinto. Besides finding that finances affected the degree to which students could engage in social activities at the institution, and that a relationship existed between finances and persistence, the results of the survey of 2,453 full-time freshman students at a large public institution led Cabrera et al. to conclude that the two models developed by Tinto and by Bean had common ground and were mutually supportive in explanatory value.

Braxton, Hirschy, and McClendon (2004) extended the work of Tinto by incorporating findings from research that looked at organizational, psychological,

sociological, and economic factors. Their premise was that these three categories of factors affecting attrition differed based on college type (residential versus commuter). Specifically, each differed with regards to the degree to which student characteristics affected outcomes. The residential model placed more emphasis on academic achievement and ability, race, sex, parental education, and family socioeconomic status than the commuter model. The commuter model emphasized psychosocial factors such as motivation, control issues, self-efficacy, empathy, affiliation needs, and anticipatory socialization. As an extension of this research, in developing a Theory of Commuter Student Departure, Braxton, et al. sought to show differences in the importance of both the internal campus environment and the life circumstances outside of campus in influencing student persistence.

This case study incorporates a conceptual framework that builds upon previous longitudinal causal research on the factors influencing attrition and retention. The relationship between retention and attrition and the distance from home to first-time, full-time students' post-secondary institution was explored. This case study focuses on the variable of distance, which has not routinely been analyzed in previous studies. As previously stated, there is no single successful retention program, and although retention and attrition models should be uniquely designed based on longitudinal institutional-specific data, (Hossler, 1991; Tinto, 1993), previous research suggests that there is a core set of demographic, socioeconomic, and academic variables to use when developing predictive models. This case study examines distance as a potential variable that should be examined by other institutions of higher education when analyzing retention and attrition factors.

Attrition Versus Retention

The conceptual framework consists of models and theories with titles suggesting that some research is devoted to determining the reasons for student persistence and retention, while other research is devoted to determining causes of attrition (McNeely, 1940; Spady, 1970; Tinto, 1975; Bean, 1980; Bean & Metzner, 1985; Astin, 1984; Pascarella, 1985; Cabrera, Nora, and Castaneda, 1992). The predominant theme of research findings involves determining which factors are most closely correlated with retention and utilizing this data to develop predictive models. Ultimately, all studies seek an outcome in which factors are identified which might impact the rate of attrition (reducing it) and the rate of retention (increasing it). It is also true that, mathematically, affecting one would have the equal and opposite effect on the other: the retention rate is, mathematically, the opposite of the attrition rate.

However, while some studies have stated that retention rates and attrition rates are opposites, others have suggested that attrition is the reciprocal or inverse of retention (Wilcox, Estes, & Buter, 2010). While it is true that in terms of describing a student (observing each event), he or she is either retained or is not (attrition), there is nuanced complexity in the examination of the potential factors and variables that influence a student's decision to remain at or to leave an institution.

The purpose of this case study is to determine if a relationship exists between two factors; attrition prior to the second year and the institutional distance from students' permanent home addresses for first-time, full-time, traditional-aged students. In this case study, the methodology used in data analysis includes point-biserial correlation and binary logistic regression. Data preparation requires that attrition after the first or second

semester be coded with a value of *1*. This assignment was chosen because analysis using binary logistic regression “predicts the *1* value of the dependent variable using the *0* level as the reference value” (Garson, 2010, p. 2). Prediction of attrition is the goal. Therefore, an observation of attrition will result in a value of *1*.

Research Questions

The central research question that this case study aimed to answer was, does a relationship exist between the institutional distance from home and the attrition of domestic, traditional-aged, first-time, full-time students prior to the second year at the University of Pittsburgh at Bradford? Following the correlational analysis, further analysis was conducted to determine if the relationship was non-linear, and if it existed between the institutional distance from home and attrition of first-year students prior to their second year when controlling for sex, race or ethnicity, expected family contribution (EFC), parental education levels (specifically, first-generation college student status), residency status, housing status, SAT or ACT score, or market segmentation as defined by the institution.

To examine the relationship while controlling for these variables, the following sub-questions were developed:

1. Will controlling for sex affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
2. Will controlling for race or ethnicity affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?

3. Will controlling for expected family contribution (EFC) affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
4. Will controlling for parental education levels (specifically, first-generation college student status) affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
5. Will controlling for residency status affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
6. Will controlling for housing status affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
7. Will controlling for SAT or ACT score affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
8. Will controlling for market segmentation, as defined by the institution, affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?

Definition of Terms

Attribute: Synonymous with variable or data element, attributes, such as sex, distance, or SAT score, were tested as possible predictors of attrition.

Attrition: For the purposes of this study, attrition is the result of the observation that a traditional-aged, first-time, full-time student (see below) failed to return for their respective second year. Attrition is the opposite of first-year retention (also colloquially known as the retention of freshmen to the sophomore year).

Distance: Distance was calculated in three different ways and stored in three separate variables. The first calculation was the distance from home to the institution based on latitude and longitude. Known as the great-circle or orthodromic distance, a colloquialism for this measurement is the “as the crow flies” distance. The second calculation was the distance from home to the institution based on the recommended driving route as defined by the Google Maps Distance Matrix Application Programming Interface (API) and the Google Maps API. However, Smith, Spinelli, and Zhou (2002) noted in their research that “commuting time between students' homes and various universities may be more important than mere intervening physical distance” (p. 39). Therefore, the third calculation was the time that it takes to commute from home to the institution based on the recommended driving route as determined by the Google Maps API and calculated by the Google Maps Distance Matrix API.

Expected Family Contribution (EFC): EFC is an estimate of the parents', guardians', and student's ability to contribute to post-secondary education expenses. EFC is calculated using the Federal Methodology, which takes into account income, some assets, expenses, family size, and other factors.

Market Segmentation: The institution has defined a primary market, a secondary market, and a tertiary market. The primary market comprises the county in which the institution is located (McKean (PA)) and those counties that are contiguous to McKean

County (Cameron (PA), Elk (PA), Forest (PA), Potter (PA), Warren (PA), Allegany (NY) and Cattaraugus (NY)). The secondary market comprises those counties that are contiguous to the primary market (Allegheny (PA), Armstrong (PA), Beaver (PA), Bradford (PA), Butler (PA), Centre (PA), Clarion (PA), Clearfield (PA), Clinton (PA), Crawford (PA), Erie (PA), Indiana (PA), Jefferson (PA), Lawrence (PA), Lycoming (PA), Mercer (PA), Tioga (PA), Union (PA), Venango (PA), Chautauqua (NY), Chemung (NY), Erie (NY), Livingston (NY), Steuben (NY), Wyoming (NY)). The tertiary market comprises all other areas (counties, states, and countries). The category of Market was assigned based on the student's Permanent Home Address (see below).

Housing Status: This variable indicated whether the student is living on campus in an institutionally owned and controlled residence hall or if the student is commuting (either from home or a rented facility not owned or controlled by the institution).

Parental Education Levels: Parental education levels determine whether a student is considered first-generational, which is defined as neither parent (or guardian) having graduated from a four-year college or university (National Center for Education Statistics, 2010). Only students who completed a Free Application for Federal Student Aid (FAFSA) were included in this study.

Permanent Home Address: The permanent home address is the student's address of record at the time of matriculation. Only students with permanent domestic home addresses were included in this study. Students who provided an international address as their permanent home address were removed, as were U.S. residents of Guam, Puerto Rico, the U.S. Virgin Islands, the Northern Marianna Islands, and students who were the

dependents of U.S. enlisted personnel using an APO address outside of the continental United States of America, and Alaska and Hawaii.

Race/Ethnicity: This variable was based on self-identified and reported information and categorized as the following: Nonresident alien; Race and ethnicity unknown; Black or African American, non-Hispanic; American Indian or Alaska Native, non-Hispanic; Asian, non-Hispanic; Native Hawaiian or other Pacific Islander, non-Hispanic; Hispanic; Two or more races, non-Hispanic; and White, non-Hispanic. These classifications are defined as part of the National Center for Education Statistics (NCES) Integrated Postsecondary Education Data System (IPEDS) annual survey.

Residency: This variable indicated whether the student is a resident of Pennsylvania (PA) or a resident of another state (OS) and is used for the determination of the rate used in the calculation of tuition. Non-residents (OS) are charged a higher tuition rate.

SAT or ACT score: This variable was calculated by adding the SAT Critical Reading and Mathematics scores. ACT Composite Scores were converted based on Concordance Tables designed by ACT and the College Board (ACT, 2007; The College Board, 2009).

Traditional-aged, First-time, Full-time Students (FTFT): The purpose of this case study was to determine if a relationship existed between the institutional distance and attrition for domestic, traditional-aged, first-time, full-time students (FTFT) at the University of Pittsburgh at Bradford. FTFT students were defined as undergraduate students, under the age of 25 at the time of enrollment, who matriculated and enrolled as baccalaureate degree-seeking students at the University in 12 or more credits, and who

had previously completed 12 or fewer post-secondary credits. The delineation of age is based on the 2013 NCES publication, *The Condition of Education 2013*, in which three age categories are described: under 25, between 25 and 34, and 35 and older.

Assumptions

1. The selection of undergraduate student records from the University of Pittsburgh at Bradford resulted in participants who were as similar as possible in all characteristics except the independent variable (McMillan, 2008).
2. The data extracted for use in this case study were accurate.
3. Self-reported data such as race/ethnicity, first-generation status, and financial data used to determine EFC were accurate.
4. The concordance between the results of the ACT and SAT tests are valid and reliable, and, therefore, it was appropriate to use this table to provide missing SAT values (Pommerich et al., 2000).

Delimitations

1. The subjects of this case study were traditional-aged, first-time, full-time students who matriculated and enrolled at the University of Pittsburgh at Bradford from 2005 to 2013. Only students who completed the FAFSA and had a permanent domestic address were included. The findings are limited to this population.

Limitations

1. Confirmation bias, the tendency to search for or interpret data and information in a way that confirms the researcher's preconceptions, leading to statistical errors, is a potential limitation (Sandelowski, 1986; McMillan, 2008). The author of this case study has worked at the University of Pittsburgh at Bradford for 15 years as assistant

dean, and registrar, and is now the Vice President for Enrollment Management. He oversees several areas including institutional research. Perceptions, assumptions, relationships, and experiences may have affected the researcher's ability to examine the institution critically. However, efforts were made to reduce any bias.

2. Large sample sizes may identify significant relationships for statistical reasons and not because the relationships really exist (McMillan, 2008).

3. The fundamental limitation of all correlational research applies; causality cannot be inferred from results that are statistically significant.

4. Recoding data to assure that data fields and definitions are consistent may have introduced errors that threaten reliability and validity. However, every effort was made to clean and code data consistently.

5. The subjects of this case study were limited to all traditional-aged, first-time, full-time students who reported a domestic address and matriculated and enrolled at the University of Pittsburgh at Bradford from 2005 to 2013. Students without a valid address within the United States of America were removed from this study. Subjects that did not complete the FAFSA were also removed from this study. Findings are limited to this population and cannot be applied to other campuses of the University or other post-secondary institutions.

6. There are multiple reasons that may affect a student's decision to leave an institution or to persist. Predictor variables or attributes outside of the focus of this case study were not explored. The variables being studied are predominately individual characteristics or group assignments. Institutional practices, student behaviors and

affiliations, and student utilization of services are not being studied, but all may influence attrition.

7. For the purposes of this case study, no differentiation was made between students who left the institution with the intent of returning at a later date and students who left and had no intention of returning. Both are considered attrition for the purposes of this case study.

Significance of Study

In the application of strategic enrollment management, student persistence is now as important, if not more so, than recruitment strategies (Peltier, Laden, & Matranga, 1999). Previous theory, literature, and research have focused primarily on the processes of marketing and of recruiting students. An opportunity to identify and recruit those students who are most likely to persist has recently been the subject of research and theory (Kotler, 1976; Litten, Sullivan, & Brodigan, 1983; Hossler, 2000). Institutions have enhanced existing programs and implemented new programs and support services to increase student success and reduce attrition. However, nationally, retention rates have remained relatively flat. Between 1983 and 2010, freshman to sophomore retention rates at public four-year institutions ranged from 66.4% in 1996 and 2005 to 70% in 2004 (ACT, 2010).

Changes in demographics and attitudes toward post-secondary education are resulting in increased competition among institutions for a shrinking college-going population. Reduced public funding and institutional budget cutbacks have compelled institutions to explore techniques and resources for reducing attrition. Much of the recent focus has been on the relationship between student backgrounds and institutional success.

Students who have a history of academic success, are not economically challenged, and whose parents have completed post-secondary degrees tend to persist at higher rates (Bean, 1990; Astin, 1991; Astin, 1993; Astin, 1997; Pascarella & Terenzini, 2005; Pascarella & Terenzini, 1991). However, data that is geographic in nature as it pertains to student success and persistence has rarely been studied and discussed in the research literature (Litten, Sullivan, & Brodigan, 1983; Ousley, 2010). If a relationship is identified, the value of this case study may reach beyond that of the University of Pittsburgh at Bradford, having implications for other institutions, especially those institutions serving rural regions. Although this is a single case study, The University of Pittsburgh at Bradford represents the most common type of institution found in the Carnegie Classification System (rural serving).

This case study adds to the existing research and literature on student persistence and predictive modeling. It enhances limited research and findings on the topic of the potential relationship between distance and attrition and introduces two new methods by which distance is measured (actual distance and travel time).

A relationship between distance and attrition has implications beyond a simple awareness that could include new or enhanced programming specific to students who are enrolled from farthest away. New programming might include topics such as homesickness, transportation options, communication, building relationships, taking advantage of new and different surroundings, and enhancing focus on objectives and goals. Other implications could include influence on the marketing and recruitment processes, patterns, and strategies that institutions and systems employ.

Chapter 2

Review of Literature

The major purpose of this case study was to determine if a relationship existed between attrition and the institutional distance from students' permanent home addresses to the University of Pittsburgh of Bradford. This chapter presents a review of literature relevant to the topic of persistence, previous studies in which the relationship of institutional distance to persistence has been explored, and possible factors contributing to a potential relationship.

1. Retention and Persistence
2. Analysis of Distance on Retention and Persistence
3. Distance in Other Student Persistence Studies

Retention and Persistence

The review of the literature regarding student retention and persistence results in a vast array of studies with an equally vast array of focuses and methodologies. The literature review that was conducted included both longitudinal and short-term studies, studies of a single post-secondary institution and multiple campuses and institutions, and subsets of populations (e.g., based on major, age, sex, race and ethnicity, socioeconomic backgrounds).

It is critical to note that the previous approaches used and variables measured often appeared as part of studies of retention, attrition, and persistence factors. In some instances, researchers have used the words "retention" and "persistence" interchangeably. However, there is a difference. Retention refers to the rate at which students re-enroll at the same institution in sequential semesters or years. Persistence refers to repetitive

behavior that leads to the completion of requirements and ultimately graduation. Additionally, The National Center for Education Statistics defines retention as an institutional measure, while persistence is defined as a student measure. Conversely, “attrition is the diminution in numbers of students resulting from lower student retention” (Hagedorn, 2012, p. 85).

The purpose of previous research has been to better describe the patterns of student retention and attrition. An important secondary purpose exists—the development of predictive models that could lead an institution to implement specific intervention strategies that would mitigate those detrimental factors correlated with attrition or to enhance those factors found to enhance retention. The research that has been conducted has often begun by starting with those factors that have been previously analyzed and found to have a statistically significant relationship with retention or attrition. Researchers have then explored other factors, some of which may be unique to the institution being studied. Theorists have hypothesized that there is no single successful retention program, and retention and attrition models should be uniquely designed based on longitudinal institutional-specific data (Hossler, 1991; Tinto, 1993). Therefore, factors in previous studies that have not been found to have had a statistically significant relationship with retention or attrition should not be overlooked. This practice is evident in the review of the literature (i.e., commonality of certain factors and variables).

This literature review included studies that focused on retention, attrition, and persistence. The following section of the literature review examines existing studies, but unless otherwise noted, the distance between home and the institution was not a variable that was measured or analyzed.

Mattox (1983) reviewed data from students at a primarily residence-oriented, four-year institution (Ohio State University) and analyzed factors at the time of matriculation that could help predict attrition. Using stepwise discriminant analysis, Mattox found that academic background characteristics such as ACT composite, high school grades, and highest degree planned were the strongest predictors. Mattox also identified several sex-specific factors: study habits, mother's level of education, mechanical ability, and plans to change the major field of study” for men, and “marriage plans, an attitude assessment related to laws prohibiting homosexual relations, and self-perceived ‘understanding of others’” for women.

Morris (1988) used a step-wise regression analysis to explore potential predictors of retention at an urban university (Temple University). Morris looked at the entire entering freshman classes from 1982 through and including 1985 and found that undergraduate grade-point average was the most important factor. Additional but less significant attributes included student's age, scores on standardized tests, family income, and the amount of financial aid received.

Gillock (1998) looked at how the process by which students made the transition from high school to college affected attrition. The hypothesis tested was that variables associated with the transition would be more predictive of attrition than individual, environmental, and socio-demographic variables. Gillock surveyed 552 students during their first semester. Additional data were gathered throughout the first year to assess persistence and attrition. Analyses were completed using correlational techniques, analysis of variance, and regression techniques. Gillock found that although the

information that was gathered and analyzed was useful in informing programming, the data were not predictive.

Goodman (1999) looked at students enrolled at Walters State Community College from 1992 to 1997. Four groups of variables were tested: demographic, pre-matriculation (e.g., high school GPA and test scores), post-matriculation variables (e.g., grades), and post-matriculation variables related to enrollment (e.g., absences, credit load). Each variable was evaluated using univariate approaches (i.e., chi-square and t-test). Goodman discovered that the following variables affected persistence: age and sex, high school test scores and GPA, timing of admissions application, lack of absences, required developmental or remedial coursework, course load, change of major, and having received financial assistance.

Hines (1999) developed and tested an instrument for reliability in predicting persistence and administered it to 436 Davenport College students (65 males and 371 females). Only one section of the instrument (i.e., social consciousness) was found to be reliable. However, 12 variables were identified that, when combined, increased the prediction of attrition by 57%. Those variables included unique criteria such as Protestant affiliation, cigarette smoking, and parents paying tuition.

Orazem (2000) conducted a qualitative grounded theory study to determine why undecided students either persisted until graduation or left Montana State University-Bozeman. Thirteen students were interviewed for this study. Orazem's findings were that, "(a) the societal forces to attend college were pervasive, and (b) the first year was one of massive social education" (p. 104), and that one of the most crucial factors for student persistence was that students developed a sense of purpose. Students who were best able

to develop a personal sense of purpose beyond external influencers were most likely to succeed and persist. Reflecting on the transition from adolescence to adulthood, Orazem stated that successful social education was another significant factor in the persistence of students who participated in this study. In addition to this social gap, another example of social and cultural distance was suggested by Orazem,

The students in this study came to the university with little understanding of the educational enterprise. They were under-prepared and overwhelmed by the enormity of the experience. They were not savvy to the bureaucracy and were in need of great assistance to navigate this new world, often intimidated by the rules and regulations (p. 105).

Orazem's participants consisted of undecided students (i.e., students who had not yet declared a major); however, his findings are likely applicable to all traditional-aged undergraduates.

Orazem made an insightful observation: He stated, "In loco parentis may be an institutional policy that is no longer practiced, but according to these students' comments, they seemed to want some sort of pseudo-parental role to help them be successful at the university" (p. 109). This observation concerning students' desire for the institution to play a pseudo-parental role is supported by many of the findings of the Pew Research Center regarding the generation known as the "millennials."

Kaiser (2005) sought to measure the effectiveness of the high school grade point average and SAT score as predictors of retention at Seton Hall University. Findings included that quartile assignment and analysis of both high school GPA and SAT score resulted in greater predictability of retention. Kaiser excluded high school rank even

though he cited several studies that concluded, “HS GPA, coupled with rank in high school class, have been found to have a higher relationship to student attrition than any other combination of predictive factors” (p. 35). Kaiser gave no explanation for his omission of high school rank from his study. However, it is possible that rank will not be used as often in future retention research (in analysis and as a predictor) because secondary schools are less frequently recording and reporting it. For example, in 2013, rank was reported for only 67.2% of the entering traditional-aged, first-time, full-time class at the University of Pittsburgh at Bradford (Baldwin, n.d.).

Milon (2006) looked at the relationship between completeness of the retention plans (or strategic plans that included retention as a specific objective) at New York State institutions and retention and graduation rates. Milon developed a retention plan rubric based on a model used by Noel-Levitz, Inc. and models that were discovered in the existing literature. Of the 177 institutions that were approached, 38 provided documentation regarding retention plans. Milon found that there was no significant relationship between retention plans and outcomes. Milon did suggest a possible flaw in the study, although he did not suggest that the flaw was significant. However, the flaw may have drastically affected his outcomes. The retention plans that were submitted were developed between 2002 and 2008. Retention and graduation rates were from 2004. Even plans developed and implemented in 2002 would not necessarily have been in place long enough for their potential effects to be fully realized. One of the findings from Milon’s literature search was that many leading researchers warn against the adoption of enrollment plans from other institutions. The uniqueness of each campus profile, mission,

and population requires individualized plans. This finding is consistent with the findings of both Tinto and Hossler (Hossler, 1991; Tinto, 1993).

Johnson's (2007) objective was to develop a mathematical equation using multivariate logistic regression analysis for predicting freshmen retention based on characteristics such as high school GPA, SAT score, first semester GPA, sex, and ethnicity. Class rank was omitted, although Johnson did cite its potential relationship to retention in his literature review. His findings showed that sex and first semester GPA were statistically significant at a "private university in the Midwest in the fall term 2004" (p. xi).

Atkinson (2008) focused on identifying factors that affect retention on the regional campuses of Ohio University. This research is especially applicable to this study given location and institutional characteristics. Atkinson noted differences in the student profile including age, socioeconomic status, and family and employment responsibilities. Atkinson compared self-reported information collected through the administration of the *Noel-Levitz: Student Satisfaction Survey 4-year College and University Version* and student demographic data with both the rates of enrollment and retention rates. Significant factors included financial aid awarded, campus appearance, and the opportunity to play sports. Demographic factors included credit load, progression (class level), and grade point average (GPA).

Campbell (2008) selected 76 attributes to include in four predictive models of retention at Northern Arizona University after having analyzed 250 potential data fields. Institutional distance was not one of the 250 original data fields that were analyzed. However, in addition to at-risk attributes such as lower high school GPA, lower SAT

score, provisional admission, lower math placement scores, lower parental adjusted gross income (AGI), Pell eligibility, sex (male), and race/ethnicity (white), Campbell found a relationship between risk of attrition and students with origins from non-urban counties.

Radney (2009) sought to identify those variables that would best predict persistence of first-time freshmen students at California State University. Using Multivariate Discriminant Analysis (MDA), Radney could only correctly classify students 55.3% of the time when using the full model. However, using a reclassification model, in which he relegated transfers to community colleges and dropouts to separate categories, respectively, the accuracy of prediction increased to 67%. Radney was later able to predict students who persisted with 90.8% accuracy. Although Radney looked at 17 different variables, only credits completed, high school GPA, and first-year college GPA were significant.

Davis (2010) had sought to develop a model of college persistence by synthesizing behavioral theory research and research from college persistence theory. Davis' population consisted of first-time transfer students who most resembled traditional college students. Using logistic regression, Davis found a significant negative correlation between distance and the decision to attend, or conversely, as Davis noted, the closer a student lived to the institution, the greater the degree to which the student was confident in the decision to attend. Furthermore, there was a significant correlation between confidence in the decision to attend and persistence.

The research on retention and persistence has focused on possible contributing factors affecting retention and attrition. Pre-matriculation variables such as demographic (e.g., sex, race/ethnicity, location), socioeconomic (e.g., household income, educational

achievement, and poverty status), and academic variables (e.g., SAT or ACT score, high school grade point average, and high school rank) are routinely analyzed. Occasionally, post-matriculation variables (e.g., first-semester grade point average, credit load, and the number of absences) have been included. These and other factors have been hypothesized, tested and analyzed using longitudinal data. Both theories and models, as were described in the theoretical foundations section of chapter 1, have been developed based on the outcomes. The primary purpose of this research has been two-fold: first, to identify and describe patterns of student retention and attrition; and second, the development of predictive models based on contributing factors that could lead an institution to implement specific intervention strategies that would counter negative factors that correlated with attrition and enhance positive factors that correlated with retention. Simply stated, the aim of this research is prediction and control (Astin & Oseguera, 2005).

Analysis of Distance on Retention and Persistence

Prior to recent advances in mapping applications, the connection between retention and distance from home was more difficult to measure. Therefore, distance from home as a continuous variable was not widely referred to in most of the studies before the start of this decade. Distance, when cited, was often reported and measured categorically. Summerskill (1962) noted that very early studies found a statistically significant correlation between attrition and the rural location of a student's hometown (Cuff, 1929; Strang, 1937; West, 1928). Few, if any, studies measured distance in terms of actual miles. Most studies focused on the location of the hometown, hometown population size, and whether or not the institution was out-of-state as compared to the

student's hometown. Early results suggested that students from larger communities tended to have higher rates of success (Fredericksen & Schrader, 1951), while students who were considered out-of-state were likely to have academic difficulties (Feder, 1940). However, the results of later studies have varied. Many have shown no statistically significant relationship between distance and either retention or attrition. Some have shown an inverse relationship.

Iffert (1957) initially reported an inverse relationship between state residency and attrition. His research showed that out-of-state students had lower attrition rates. However, the majority of out-of-state students that were included in his sample were from private institutions. At private institutions, attrition rates were lower, and graduation rates were higher than the rates at public institutions. Iffert also realized that the differential in cost and financial resources between out-of-state students at private and public institutions could potentially affect attrition rates. Iffert stated that "[t]he weight of the evidence points to the conclusion that location of a student's home in relation to college had no significant bearing on his chances of graduation" (p. 74).

Aiken (1964) developed and administered an inventory to 1,006 incoming freshmen at the Woman's College of the University of North Carolina in 1962. Aiken then conducted additional analysis on two samples of 100 students, each, who had completed the first semester. Utilizing multiple regression analysis, Aiken found a statistically significant relationship between attrition and each range of distance from home to the institution that was greater than 100 miles (i.e., 100-199, 200-299, 300-399, 400-499, etc.).

While Astin (1975) reported that attrition rates were higher for students living at home, Pantages and Creedon (1978) found that a correlation existed between the distance from home and the probability of attrition. This suggests that a nonlinear relation may exist between distance and attrition (i.e., students very close to home and students very far from home are less likely to be retained).

Stork and Berger (1978) collected data from 120 freshmen who were retained (classified as “persisters”) and 120 freshmen who attrited (i.e., left the institution) (classified as “leavers”) from the College of Liberal Arts at Utica College in 1972. Stork and Berger’s analysis, using a stepwise discriminant analysis, found a statistically significant ($p < .01$) difference in distance from home between persisters and leavers when distance was defined as a zone. The results showed that the “distance zone means translate to approximately 210 miles for the persisting group and 360 miles for the leavers” (p. 286).

Bean’s (1980) causal model of student departure placed greater emphasis on variables that were environmental in nature. Utilizing multiple regression analysis and path analysis, Bean found that state residence and distance from home, two of several organizational determinants, were among several influential background variables that were found to be related to attrition among only the women studied. Neither variable showed a statistically significant relationship with attrition among the men in the study.

Ramist (1981) analyzed the College Board's Admissions Testing Program Summary Reporting Service (ATP-SRS) data for factors and variables affecting retention and attrition. He concluded that the hometown location was correlated with persistence if the hometown was in a state that was contiguous to the state in which the institution was

located. Conversely, a relationship between location and attrition existed when the hometown was located in a state not contiguous to the state in which the institution was located. Ramist noted that the rates of attrition and persistence for students whose hometown was in the same state as that of the institution were between those of the contiguous out-of-state hometown, and the non-contiguous, out-of-state hometown.

French (1982) measured the effects of five environmental factors among a sample of first-year students who started their studies at Durham and Loughborough universities (England) in 1975. Actual physical distance from home to the institution was measured and categorized as one of the variables under “The Personal Environment.” French included this variable (distance) because his literature review showed several examples of studies that led to the conclusion that separation from home was a potential source of transitional stress, especially among students who reported that they were unprepared for being separated from friends and family. However, French’s results suggested that separation (in terms of distance) was not a significant factor.

Lenning (1982) reviewed attrition studies from 1975 through 1981 and found that the results of those studies indicated a correlation between students from rural areas and attrition. However, institutional characteristics such as size and nature affected the validity of this result. Lenning also noted that there was a correlation between attrition and students who were from non-contiguous states.

Hatch’s (1983) objective was to develop a procedure for analyzing retention at any institution. An analysis of factors affecting retention was conducted at different institutions, each representing a different type: small college or large university. Although Hatch was able to achieve a prediction with 70-80% accuracy using multiple

variables identified through discriminant analysis, the variables used in each predictive model differed significantly among institutional types and were influenced by other variables such as sex. In his findings, distance from home was a factor that only affected the retention of men at the large university that was included in the study. Six other variables were stronger predictors (part-time employment had the greatest effect).

Cash (1990) looked at the variables affecting freshmen retention at two Seventh-day Adventist higher education institutions. The testing of Tinto's model of retention was the primary objective of this study. In the literature review, Cash noted that there were conflicting findings regarding the effect of distance on retention. However, the examples provided were from research published in 1970, 1973, and 1981, respectively. Recent studies of the "millennial" suggest that there are significant differences between students entering college today and those who enrolled 30 or 40 years ago. Cash also cited research that looked at the locale (e.g., city vs. town, urban vs. rural) as a factor and noted that findings often showed a connection between small town size, the degree to which the town was rural, campus size, and retention. Cash also sought to determine if social distance might play a role. Social distance comprises several different variables: "the amount of secondary education in church-operated high schools, the church membership of parents, whether parents were employed by the church or not, and geographical distance of the home from the university" (p. 83). Essentially, this is the analysis of the effect of a combination of specific variables on retention, not unlike a regression analysis. In this study, distance, a sub-variable of social distance, was not separately analyzed for its effect on retention. The findings of Cash's study were that factors affecting student enrollment and retention at these two institutions were mostly similar to students at other

public and private institutions. Factors that differed included those related to a commitment to a Christian education, and social distance.

Bank, Slavings, and Biddle (1990) examined the effects of social influences on undergraduates' decisions to leave or remain at the institution of first matriculation. Bank et al. used a four-stage regression analysis of the data—a process in which a regression model is built using an initial set of variables. The residuals of the first analysis then become the dependent variable, and the next set of variables is then analyzed. The population consisted of 1,240 first-year students at a large Midwestern state university. Bank et al. found that while normative influences had a stronger effect than modeling influences on students' intentions, both, when combined with background variables, increased the predictive model in a statistically significant way. One of the background variables that was analyzed was the distance between home and school. Bank et al. found that distance was not a statistically significant background variable in predicting attrition after the first or second semester and resulted in regression coefficients of less than .05.

Although Paulsen's (1990) research focused on factors affecting college choice, his findings indicated that there were relationships between distance and prior academic success, aspiration, parent's educational attainment, and socioeconomic status—all factors that have been found to relate to student persistence. Paulsen reviewed previous studies on college choice and found that students who had stronger academic backgrounds were more likely to choose highly selective and out-of-state institutions (Jackson, 1978; Gilmour, Spiro and Dolich, 1978; Dahl, 1982; Zemsky and Oedel, 1983; Hearn, 1984). Paulsen (1990) noted that a correlation existed between the institutional distance from home and parents' level of education. He suggested that distance served as

an additional proxy for socioeconomic status. Paulsen also stated that students saw colleges or universities as less attractive when expenses (tuition, room, and board) and distance from home increased. However, attractiveness increased when the availability of aid increased. Conversely, Paulsen stated that the attractiveness of colleges, in general, as well as a specific institution, increased as the distance from home decreased. This research implies that distance not only influences the choice of institution, but it also influences the perceived attractiveness or fit of a chosen institution prior to matriculation and initial enrollment.

Mooney, Sherman, and LoPresto (1991) surveyed 88 female undergraduates to determine if academic locus of control, self-esteem, and distance from home were variables that predicted adjustment to college. They found that an internal locus of control, high self-esteem, and the perception that the distance from home was "just right" were all related to the four dimensions of college adjustment when assessed using the Student Adaptation to College Questionnaire (SACQ). However, Mooney et al. did not observe a correlation between students' actual distance from home and the four dimensions of adjustment.

Kelly (1996) conducted a study to examine persistence at the United States Coast Guard Academy using Astin's (1993) Input-Environment-Output (I-E-O) model. He examined 619 persisters and non-persisters who were members of the entering classes of 1991 or 1993. Although Kelly was aware of the inconsistencies in the findings of previous studies involving distance between a student's home and the institution as a predictor of attrition or retention, he hypothesized that distance, in the case of cadets, might serve as a proxy for the opportunity to maintain relationships with high school

friends. Since shore leave and liberty opportunities were limited (both in duration and quantity), the closer that a cadet lived to the academy, the more often he could return home and maintain those previous relationships. Maintaining previous external relationships prevented the cadet from socially integrating within the academy. Additionally, the likelihood of integration into the social culture of the academy decreased as the number and the quality of friendships outside of the academy increased. Kelly then linked that social integration with persistence based on previous studies and literature reviews. Kelly used government mileage charts to record distance from a cadet's hometown to the academy. Using univariate and multivariate statistical techniques, Kelly confirmed statistically significant relationships between variables associated with the academic dimension as well as the social dimension. However, the variables associated with the academic dimension were stronger and more consistent than those associated with the social dimension. Kelly did not provide any specific analysis or conclusions concerning the variable of distance (DIST) from the cadet's hometown to the academy.

Papa (1996) studied academic preparedness as a predictor of retention. The primary focus was on the measure "predicted grade point average" (pGPA). The pGPA comprises the following variables: SAT-V score, SAT-M score, high school rank, and the number of high school courses completed. Papa found that openness to new ideas, study habits, leadership, high school GPA, and high school type collectively predicted retention with 80.1% accuracy for those students with a pGPA below the median. However, for students with a pGPA above the median, college GPA, initial impression, and distance from home predicted retention with 65.8% accuracy. Papa noted in his review of the

existing literature that few studies had investigated the relationship between distance and retention. As was stated in the introduction, distance, with the advent of better GIS software, is now more easily measured.

Cobb (2001) examined students' perceptions of six brand identity attributes and the relationship of the attributes to factors affecting student retention. Citing business research, the six brand attributes included the following: "(a) vision of the institution's brand identity, (b) brand-customer relationship, (c) total employee commitment, (d) quality of programs, (e) commitment of financial resources, and (f) pricing" (p. 8). Cobb utilized descriptive statistics, measured the differences in the means using ANOVA and Pearson's r , and used both correlation and multiple regression analyses. Cobb discovered a correlation between the brand attributes and traditional variables associated with retention research (e.g., sex, financial aid status, housing, cumulative GPA). Of special interest is the finding that "Personal reason (distance from home) and difficulty with academic integration were cited as the top two reasons for students' non-persistence" (p. xiv). Applying Tinto's work and findings regarding student integration, Cobb noted that institutions "could focus their orientation courses to assist such students cope with loneliness away from home" (p. 108) as a means of affecting distance as a negative factor in persistence.

Bebergal (2003) found no statistically significant relationship between students' county of residence and persistence at a mid-size, public, four-year institution in southeast Florida. Bebergal randomly selected 1,500 students from 6,000 who entered the institution as first-time, full-time students. In addition to common variables that are routinely analyzed as possible predictors of persistence to the second year (academic,

socioeconomic, demographic), she also looked at location. She defined location as one of three possible outcomes: Palm Beach and Broward, Other Florida County, and Outside of Florida. Incorporating demographic and academic variables, Bebergal used discriminant analysis to develop a predictive model of persistence for incoming freshmen at Florida Atlantic University. Although the model failed to predict persistence beyond the probability of chance, it successfully predicted attrition at a rate that was statistically better than chance.

Stillman (2007) analyzed selected factors to determine the level of association with freshmen retention at Southern Oregon University. Data were collected regarding demographic characteristics, secondary school experiences, finances, and socioeconomic status. High school grade average, SAT/ACT score, and parental educational level were found to be statistically significant factors related to retention. Utilizing Chi-square tests for independence, Stillman noted that the lack of a statistically significant relation between retention and native language, college distance from home, living arrangements, religious preference, ethnicity, and parental income were not consistent with findings from other studies. Focusing on just one variable, distance from home, and comparing data from Stillman's research to that of others, one obvious difference is noticeable: coding and grouping of the independent variable of distance. Stillman's categories of distance were the following ranges: 5 or fewer, 6-10, 11-50, 51-100, 101-500, and over 500 miles. Had Stillman followed the protocols used in many of the previous studies reviewed, his six categories would have been collapsed into only three or four categories. For example, if Stillman had utilized Johnson's (2010) categories (albeit a later study), Stillman's first three categories would have been condensed into a single category.

Stillman, as part of his analysis, then reduced his number of categories to just two and reported that of his sample, “Students living less than 50 miles from their home had a higher persistence rate (65.9% versus 62.2%)” (p. 59). Stillman did not provide a breakdown by category, so further conjecture regarding this variable cannot be made.

Arreguin (2008) applied gravity models to track and predict retention at Stephen F. Austin State University. Using freshmen enrollment data from 2001 to 2005, Arreguin calculated the distance from home to the University to determine the effect of distance on enrollment and retention. Arreguin noted, as part of his literature review, that he was unable to find examples of retention research that utilized the gravity model. However, Arreguin found that distance, which was rarely referred to in other studies, positively correlated, albeit weakly, with retention. These findings implied that the nearer that a student lived to Stephen F. Austin State University, the more likely that he or she would be retained. However, the correlations were weak and not statistically significant. Regression analysis, combining the variables related to distance, previous graduates, and competition, resulted in values that were not significant. Arreguin then categorized counties based on current enrollments (low, moderate, and high) and found no strong or statistically significant correlations between retention and distance when enrollment was added as a factor.

Davis (2010) sought to develop a model of college persistence by synthesizing behavioral theory research and research that formed the basis for college persistence theory. Four specific areas for analysis were identified:

- 1) pre-collegiate variables that influence perceptions of higher education experiences;
- 2) sociological, psychological, organizational, and economic

perceptions of higher education experiences that influence intentions regarding participation in higher education; 3) interactions between pre-collegiate variables and perceptions of higher education experiences that influence intentions for participation in higher education; and 4) development of causal models that resulted from the observed relationships among pre-collegiate variables, perceptions of higher education experiences, and intentions for participation in higher education (p. xvi).

Davis identified several factors that were consistent among the various models researched and tested including mother's level of education, certainty of major, satisfaction with high school life, distance from institution to home, and years between graduation and matriculation. The variable associated with distance was a self-reported approximation of the physical distance between home and college. Davis found a significant negative correlation between distance and the decision to attend, or conversely, as Davis noted, the closer a student lived to the institution, the greater the degree to which the student was confident in the decision to attend. Furthermore, there was a significant correlation between confidence in the decision to attend and persistence. Additionally, the self-reported perception of the "difficulty in transferring" influenced the degree to which distance negatively affected retention. The relationship between distance and attrition was weaker among students who perceived transferring to a different institution as difficult, due to location, major, or credits earned. This variable and the finding were unique among the studies that were reviewed as part of the literature review.

Johnson (2010) sampled freshmen at Ball State from 2004-2008 to "determine what combination of demographic, academic, and athletic variables best predicted first-

year GPA and retention into the second year of college” (p. 119). Distance from home was cited as one of the variables for consideration; however, as part of his literature review, Johnson cited only one study that directly linked distance from home to rates of attrition. However, Johnson did cite several studies in which distance from home was a factor in the choice of institution. Johnson also cited 2008 Higher Education Research Institute (HERI) data from the Cooperative Institutional Research Program (CIRP) showing that, nationally, 49.4% of students reported that it was important or somewhat important to attend a college or university that was close to home with 52.8% of freshmen attending a college or university within 50 miles from home. Links between distance and choice, as well as the strength of choice and persistence, were made again in this study, although not directly proven. Johnson cited research that suggested two possible models that described the distance factor. Citing Fisher (1989), Johnson (2010) noted that culture shock and feelings of isolation were two potential results of homesickness that could vary depending on the distance from home. However, Johnson also cited research in which the self-reported perception of distance was a better predictor than the actual distance itself.

Values for the variable of distance from home were coded as “short distance” (fewer than 100 miles away), “medium distance” (101-250 miles away), and “long distance” (more than 250 miles away). Johnson showed that following completion of the first term, post-secondary GPA was a significant factor in predicting retention and distance was no longer significant. Johnson failed to hypothesize reasons for that result. One might surmise that those students who were academically successful, and for whom distance was an issue, were able to overcome homesickness due to positive feelings about

their academic success. Therefore, this factor lost statistical significance because those who were academically successful, regardless of perceived or actual distance, were more likely to be retained. A follow-up study might focus only on those students who were not academically successful to determine the significance of factors for that specific sub-population. It is also possible that there was a correlation with distance and GPA (i.e., as the distance from home increased, GPA decreased), and if a correlation between GPA and persistence existed, then a relationship between persistence and GPA might exist (i.e., a transitive relationship). Additional post-term qualitative research might also yield additional results.

Johnson found that distance and retention were significantly correlated. He reported that students were 1.5 times less likely to be retained for each distance level studied (i.e., short, medium, long). When combined with race, distance became an even stronger predictor of retention. Johnson does not explain why that might be, but given the culture and student profile of Ball State University, located in Muncie, Indiana, cultural shock and feelings of isolation may have been factors resulting in homesickness and leading to attrition.

Ousley (2010) explored and tested the use of spatial analysis in enrollment management to predict persistence and graduation. Ousley proposed that exploring U.S. Census and geographic information science (GIS) data would result in new variables that could become part of retention outcomes predictions. This study looked at enrollments and retention of freshmen at the University of Arizona between 2004 and 2007 and included traditional variables in retention studies (e.g., sex, race/ethnicity, residency, high school GPA, SAT/ACT score). The study also included a few variables that are less

frequently cited (if at all) among the retention studies reviewed. Those included the following: distance from home, the locale (as a percentage of a block group classified as urban area, as a percentage of a block group classified as urban cluster, and percentage of neighborhood that is an ethnic minority), timing of application submission, number of standardized tests submitted, and median household income level for the block group. Ousley's findings were that GPA was the most predictive variable of student persistence. Cultural and social capital proxies could be helpful in predicting persistence (especially neighborhood educational level of males). He also found that the number of standardized tests that had been taken (i.e., submitted) and the timing of the submission of the admission application was significant. Variables and the degree to which they were accurate in predicting persistence did differ when race/ethnicity was added as a dependent variable. However, distance from home to the University was not a statistically significant predictor of persistence from the first to the second year for any sub-population or when controlled for other variables with the exception of Native American students and nonresidents.

Walke (2010) explored the relationship between college choice and retention at historically black colleges and universities (HBCUs). This study included distance from home as a variable. The study acknowledged the potential effects of social distance (Cash, 1990) and cultural distance (Malocsay, 2004). In these two examples, "distance" refers the perception of the degree to which the setting is different from one's origin. Therefore, distance can include a perceived integrational "distance" based on differences in the locale, population, education levels, culture, as well as actual distance.

Walke's variables included amount and type of financial aid, college choice measures (e.g., first choice, number of colleges applied to, number of colleges admitted), perception of institutional image (e.g., friendly, diverse, supportive, isolated, prestigious), ratings of the institution (e.g., reputation, quality of housing, cost, campus surroundings, campus attractiveness), and classifying data (residency, sex, housing status, high school GPA, and SAT score). Neither distance from home nor any other kinds of GIS data were actually analyzed, although data was recorded. Walke found the strongest relationships to retention to be among financial aid types and amount awarded, pre-matriculation expectations, and ratings of the institution. Walke stated, "This study was unable to confirm high school GPA and SAT scores as predictors of freshman retention outcomes" (p. 74).

Before the start of this decade, very few studies attempted to test for a relationship between the institutional distance from home and attrition. Distance was rarely measured as a continuous variable and instead was described in categorical terms. The results of those limited studies are mixed. The results of some studies suggested that although a correlation existed, it was not statistically significant or at best, the relationship was weak. The results of other studies showed that a statistically significant relationship did exist. However, even the results of those studies are mixed: some results showed a direct relationship while other results showed an inverse relationship.

Although recent technology allows researchers to more easily measure distances between points within very large data sets, analysis using this type of data is still not common. The author of this case study could not find any examples in higher education research where the commute time of residential students from home to institution was

measured and included as a potential factor affecting retention, attrition, or persistence. New technology affords researchers the opportunity to explore this potential relationship and enhance the existing literature. For the purposes of this case study, exploration outside of traditional retention and persistence studies is necessary to enhance the knowledge of the researcher and to better inform the research.

Distance in Other Student Persistence Studies

Research involving homesickness is relatively recent. During the last two decades, research has been conducted in describing what homesickness is, its causes and effects, and ways in which it might be mitigated.

Thurber (1995, 1999) stated that anticipated or actual separation from home can cause homesickness, which is distress or impairment typified by “preoccupying thoughts of home and attachment objects” (abstract, 1999). Van Tilburg et al. (1996) proposed two different types of homesickness. The first type pertains to the challenges associated with a new environment. The second type pertains to missing the previous environment and the specific, significant people that are associated with that environment. Although mild homesickness is common and can lead to the development of new associations and coping skills, moderate to severe homesickness can create anxiety, depression, social withdrawal (withdrawing from the environment), and the inability to focus.

Homesickness can also exacerbate existing disorders (Thurber & Walton, 2012). Thurber and Walton observed that the majority of research on homesickness addressed students in post-secondary institutions as the population is at significant risk due to leaving their home for post-secondary attendance. Van Tilburg and Vingerhoets (2007) stated that future research on the phenomenon of homesickness would benefit from “... clarifying

and defining the concept of homesickness and by developing a comprehensive theoretical framework” (p. 173). Thurber and Walton (2012) noted that research to date (i.e., 2012) had been primarily phenomenological in nature (i.e., studies focused on the development of consciousness and self-awareness).

Fisher, Frazer, and Murray (1986) studied homesickness in boarding students and found no relationship between distance and homesickness. However, in a previous study of university students by the same researchers, Fisher, Murray, and Frazer (1985) found that geographical distance was a factor, the impact of which could be lessened by other factors such as psychological distance, opportunities for communicating with those at home, ability to communicate with home, and the degree to which the university environment was similar to the student’s home environment.

Fisher (1989) developed a model of homesickness that was multicausal. Fisher believed that there was a critical balance between personal control and social and environmental demands. Fisher’s model stated that separation itself was not the cause of homesickness, but that separation affects a person’s commitment to a new environment. A high degree of commitment to a new environment will mitigate homesickness, while a low degree of personal commitment will likely result in feelings of homesickness.

Strobe, Van Vliet, and Hewstone (2002), using a questionnaire that they developed, surveyed all incoming new students at the Faculty of Social Sciences of Utrecht University in The Netherlands in 1995 and all incoming new students at Cardiff University in The United Kingdom in 1996. The first goal of this study was to add to existing research by thoroughly examining college student homesickness. A second goal was to provide a cross-cultural examination and comparison of homesickness. Distance

from students' hometowns was calculated, only in the United Kingdom sample, based on the self-reported data. This data was not available in the sample from The Netherlands. Physical distance was found to correlate only with the degree to which a student reported missing their family, and not an overall degree of homesickness. Strobe et al. noted that the relationship between distance and homesickness needed further examination citing conflicting studies by Fisher, Frazer, and Murray (1996) in which no relationship was found between distance and homesickness in children and studies by Fisher, Murray, and Frazer (1985) in which a relationship was reported.

Scopelliti and Tiberio (2010) conducted a study of homesickness among 200 Italian university students living in Rome. Half of the students from the sample population were from Rome; the other half were from different regions of Italy. The study was to determine if a predictive model of homesickness could be developed for non-native students using several variables (e.g., demographic, experiential, personality, etc.). Variables, using multiple regression analysis, were tested in a four-step process. Distance from Rome was a statistically significant predictor in steps 1 (residential experience) and 2 (personality factors), but was excluded from the final model when steps 3 (feelings of attachment) and 4 (attachment to Rome) were added. However, in their findings, Scopelliti and Tiberio also noted that the relationship between distance and homesickness needed further examination, citing conflicting studies by Fisher, Frazer, and Murray (1996) in which no relationship was found between distance and homesickness in children and studies by Fisher, Murray, and Frazer (1985) and Strobe, Van Vliet, and Hewstone (2002) in which a relationship was observed between distance and homesickness among college students.

Finally, the definition of distance as it applied to retention and persistence studies can include concepts such as social distance (Cash, 1990) and cultural distance (Malocsay, 2004). Ousley's (2010) research begins to scratch the surface in terms of the use of geodemography in retention and persistence prediction, but it clearly suggests directions for future research. Definitions of "distance" have also been broadened by research that has cited the following: (a) the importance of academic and social integration as espoused by Tinto; (b) Kuh's findings regarding the link between persistence and closing the gap between students' cultures and the culture of the institution; and (c) Astin's recommendations concerning students' academic and social involvement. To some extent, nearly all of the literature reviewed for this case study suggests that successful students are able to bridge the "distance" gap (Astin, 1984). The definition of "distance" could be expanded by recognizing "the gap between a student's culture of origin and the dominant culture of the institution" (Kuh and Love, 2000 in Malocsay, 2004, p. 1). This gap is the sum of many forms of "distances," including the differences between a student's experience and the institutional (and regional) culture regarding the following areas: actual distance in mileage, differences in locale (size, ruralness, access, amenities), racial/ethnic composition, socioeconomic status, expectations in terms of motivation and achievement, educational experience, political, religious, sex composition, and access and experience with technology. In some instances, actual distance could serve as a proxy for measuring potential gaps.

Research into homesickness (its causes and effects) has focused on the psychological consequences of relocation, particularly on student adjustment to a new location. Current models of homesickness focus on loss, interrupted lifestyle, reduced

control, changes in role and self-consciousness, and personal and situational factors. In many of the studies, the relationship between geographical distance (both actual and perceived) and homesickness has been explored. However, conflicting results have emerged, and more research is required.

Commute Time

As previously stated, the author of this case study could not find any examples in higher education research where distance, measured as the commute time of residential students from home to the institution, was calculated and included as a potential factor affecting retention, attrition, or persistence. Furthermore, actual commute time (or travel time) as opposed to geographical distance, was examined as a potential factor in only one recent quantitative study that examined retention, attrition, and persistence rates among students at a commuter institution. However, one study that was reviewed did incorporate commute time in a study of student success and retention.

Johnston (2013) explored academic success and retention in the Middle Tennessee State University Horse Science Program finding that commute time was negatively correlated with campus participation and academic GPA for commuting students (i.e., students not living in campus housing). More research into the potential effect of commute time is required.

Chapter 3

Methodology

As outlined in chapter 1, the purpose of this quantitative case study was to conduct an *ex-post facto* correlational analysis of the independent variable of institutional distance from each respective student's home as it relates to the dependent variable of attrition of traditional-aged, first-time, full-time students prior to their second year at the University of Pittsburgh at Bradford, a four-year campus of the University of Pittsburgh, a state-related, public Research I institution located in Pennsylvania. Following the correlational analysis, further analysis was conducted to determine if a non-linear relationship existed between the institutional distance from home and attrition of first-year students prior to their second year when controlling for sex, race or ethnicity, expected family contribution (EFC), parental education levels (specifically, first-generation college student status), residency status, housing status, SAT or ACT score (economic, demographic, and academic variables that have been routinely studied and tested in previous studies spanning the history of retention research) and market segmentation as defined by the institution (a proxy measure of distance) (McNeely, 1940; Spady, 1970; Tinto, 1975; Bean, 1980; Astin, 1984; Pascarella, 1985; Cabrera, Nora, and Castaneda, 1992).

The correlational method (also referred to as the correlational research strategy) is often utilized when “two different variables are observed to determine whether there is a relationship between them” (Gravetter & Wallnau, 2009, p. 12). The correlation is a common statistic that is often used to measure and describe the relationship. The

correlation is used in several different applications, including prediction, validity, reliability, and theory verification.

This case study utilizes correlation to verify theory: the prediction that a relationship exists between the institutional distance from home and attrition of first-year students prior to their second year of enrollment. However, it is important to note that correlation does not imply (require) causation. A correlation does not assure a cause-and-effect relationship. Although a correlation can suggest a relationship between two variables, it does not explain why the relationship exists. Examples of errant inference of causation include reverse causation, a string of causation, bidirectional causation, common-causal variable, and the ecological fallacy. Nevertheless, correlation is a statistical tool that hints at the possibility that a relationship exists (Tufté, 2003).

Research Hypotheses

The primary reason for this case study was to determine if a relationship existed between attrition and the institutional distance from students' permanent home addresses for first-time, full-time, traditional-aged students at the University of Pittsburgh at Bradford. The following research question and null hypothesis were developed to address this question:

R1: Is there a statistically significant relationship between the institutional distance from home and first-year attrition for traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

H₀: There is not a statistically significant relationship between the institutional distance from home and first-year attrition for traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford.

To address this question, point-biserial correlation analysis was chosen.

Following the point-biserial correlation analysis, further analysis was conducted to examine the relationship to determine if a relationship existed between the institutional distance from home and attrition of first-year students prior to their second year, when controlling for demographic, socioeconomic, and academic, pre-matriculation variables that have been routinely analyzed as possible factors and predictors of attrition and retention, as was outlined in the literature review. Those variables are sex, race and ethnicity, expected family contribution (EFC), parental education level, residency status, housing status, SAT and ACT score, and market segmentation, as defined by the institution. To address this question, the following research sub-questions and null hypotheses were developed.

R1a: Will controlling for sex affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

H0a: Institutional distance from home, when controlling for sex, is not a statistically significant predictor of attrition at the University of Pittsburgh at Bradford.

R1b: Will controlling for race or ethnicity affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

H0b: Institutional distance from home, when controlling for race or ethnicity, is not a statistically significant predictor of attrition at the University of Pittsburgh at Bradford.

R1c: Will controlling for expected family contribution (EFC) affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

H0c: Institutional distance from home, when controlling for expected family contribution (EFC), is not a statistically significant predictor of attrition at the University of Pittsburgh at Bradford.

R1d: Will controlling for parental education levels (specifically, first-generation college student status) affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

H0d: Institutional distance from home, when controlling for parental education levels (specifically, first-generation college student status), is not a statistically significant predictor of attrition at the University of Pittsburgh at Bradford.

R1e: Will controlling for residency status affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

H0e: Institutional distance from home, when controlling for residency, is not a statistically significant predictor of attrition at the University of Pittsburgh at Bradford.

R1f: Will controlling for housing status affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

H0f: Institutional distance from home, when controlling for housing status, is not a statistically significant predictor of attrition at the University of Pittsburgh at Bradford.

R1g: Will controlling for the SAT composite score affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

H0g: Institutional distance from home, when controlling for the SAT composite score, is not a statistically significant predictor of attrition at the University of Pittsburgh at Bradford.

R1h: Will controlling for market segmentation as defined by the institution affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

H0h: Institutional distance from home, when controlling for market segmentation as defined by the institution, is not a statistically significant predictor of attrition at the University of Pittsburgh at Bradford.

To address each of the preceding sub-questions and null hypotheses, binary logistic regression analysis was selected.

Research Design

An *ex-post facto* quantitative research case study design was used to determine if a relationship existed between the institutional distance from home and attrition of traditional-aged, first-time, full-time students prior to their second year at the University of Pittsburgh at Bradford. Following the correlational analysis, further analysis was conducted to examine the relationship and to determine if a non-linear relationship existed between the institutional distance from home and attrition of first-year students

prior to their second year when controlling for sex, race or ethnicity, expected family contribution (EFC), parental education levels (specifically, first-generation college student status), residency status, housing status, SAT or ACT score, and market segmentation as defined by the institution.

An *ex-post facto* case study was used because experimental research was not possible, and the researcher could not control the dependent or independent variables. Group comparisons and additional correlations, based on previous research and existing theory in student retention (Astin, 1991; Astin, 1993; Astin, 1997; Pascarella & Terenzini, 2005; Pascarella & Terenzini, 1991) were made based on the following additional variables: composite SAT scores, high school grades (grade point average), sex, race, parental education levels, housing status, and expected family contribution (EFC). Market segmentation is a unique variable in this study that is assigned by the institution and is based on the county of origin.

Creswell (2009) defined the case study as the process in which “researcher explores in depth a program, an event, an activity, a process, or one or more individuals” (p. 13). Creswell (2007) noted that the case study can be either a single case or a case bounded by time and place. Leedy and Ormrod (2001) stated that case studies attempt to learn “more about a little known or poorly understood situation” (p. 149). Creswell (2009) advised that the structure of a case study should be the problem, context, issues, and lessons learned, and Creswell also noted (2007) that the case study design can be exploratory, explanatory, or descriptive in nature. Gomm, Hammersley, and Foster (2000) noted that the case study was appropriate when the researcher had an interest in

naturally occurring features or variables in context, and that the nature of the case study could be quantitative, qualitative, or both (mixed methods).

The intended outcome of the case study approach is the understanding of the case or cases through an interpretation of the data (Creswell, 2009). The purpose of this specific case study was to determine if a relationship existed between student attrition and the institutional distance from students' permanent home addresses for first-time, full-time, traditional-aged students attending the University of Pittsburgh at Bradford.

Site of research

The population involved in this case study was drawn from the University of Pittsburgh at Bradford, a regional, undergraduate campus of the University of Pittsburgh. The University of Pittsburgh is a state-related, public Research I institution that is located in Pennsylvania.

The University of Pittsburgh at Bradford (UPB) is a regional campus of the University of Pittsburgh located in Bradford, Pennsylvania. It enrolls approximately 1,500 students and offers 46 majors. The Carnegie classification for UPB is baccalaureate colleges - diverse fields. The size and setting designation is small four-year, highly residential. The enrollment profile is listed as exclusively undergraduate four-year, and the undergraduate profile is full-time four-year, selective, higher transfer-in. Approximately 83% of the total degree-seeking enrollment reside in Pennsylvania (University of Pittsburgh at Bradford, n.d.).

Population and sample

The population involved in this case study consisted of all traditional-aged, first-time, full-time students (freshmen) matriculated and enrolled during the fall semesters of

2005 through and including 2013. In order to increase the validity of the study, the entire population of the institution was considered. Each row represented a unique traditional-aged, first-time, full-time student who matriculated and was enrolled in a fall semester between 2005 and 2013, beyond the date on which a student may drop all classes without tuition liability (approximately two weeks after the beginning of each semester).

First-time student (undergraduate) classification was based on the definition used by the Integrated Postsecondary Education Data System (IPEDS). IPEDS is the core postsecondary education data collection program for the National Center for Education Statistics. It collects standardized data from all institutions of higher education that receive federal student financial assistance under Title IV of the Higher Education Act of 1965. The IPEDS definition of a first-time student is as follows:

A student who has no prior postsecondary experience (except as noted below) attending any institution for the first time at the undergraduate level. This includes students enrolled in academic or occupational programs. It also includes students enrolled in the fall term who attended college for the first time in the prior summer term, and students who entered with advanced standing (college credits earned before graduation from high school). (National Center for Education Statistics, 2010)

Full-time student (undergraduate) classification was also based on the definition used by IPEDS, “A student enrolled for 12 or more semester credits, or 12 or more quarter credits, or 24 or more contact hours a week each term.”

Domestic students comprised the study population; students lacking a permanent U.S. residence were eliminated from the population so as not to skew distance analysis.

Finally, only students who completed a Free Application for Federal Student Aid (FAFSA) were included in this study.

Institutional Review Board

Expedited review was requested and approved by the Institutional Review Board at the University of Nebraska, Lincoln. Additionally, the researcher obtained approval from the Institutional Review Board at the University of Pittsburgh. Expedited review was requested as well. The study was reviewed and approved by the expedited review procedure authorized under 45 CFR 46.110 and 21 CFR 56.110. The study itself was approved under 45 CFR 46.110.(5). IRB granted a waiver of informed consent, and the risk level designation was Minimal Risk.

The study involved existing data, and de-identification of identifiable data was proposed in several ways: within the resulting dataset (e.g., data masking, pseudonymization, reducing the precision of information), de-identified through the use of software (e.g., The PARAT tool, μ -ARGUS, Cornell Anonymization Toolkit, University of Texas Anonymisation Toolbox), or de-identified by an independent person who is named in the protocol and who has completed the honest broker certification.

De-identification was accomplished through the deletion of the direct identifiers (data masking). The practice of deleting direct identifiers ensures that unintended disclosures will cause no harm and will reduce the likelihood of researcher bias. The direct identifiers that were deleted were the EMPLID and the student's address. The EMPLID is a unique, sequential number assigned to each student by the University when a bio/demo record is created. Because a single data set was utilized, pseudonymization was not necessary.

De-identification through the deletion of the direct identifiers was accomplished in two steps. The first step was the deletion of the student EMPLID after the data were obtained, and the researcher confirmed that there were no duplicated records. The second step, the deletion of the student address, was completed following verification that the address was domestic and after *DistanceD*, *DistanceF*, and *DistanceT* were calculated. The data was stored as an encrypted file on a desktop computer that was password protected and located in the investigator's office at the University of Pittsburgh at Bradford. The data was seen only by the investigator during the study.

Research Questions and Hypotheses Development

Primary research question: The primary research question that this case study aimed to answer was, does a relationship exist between the institutional distance from home and the attrition of traditional-aged, first-time, full-time students prior to their second year at the University of Pittsburgh at Bradford? Point-biserial correlation, partial correlation, and binary logistic regression analyses were used to analyze the data to answer the central research question.

The point-biserial correlation was the first statistic that was used to determine if a relationship exists between the institutional distance from home and attrition of first-year students prior to their second year of enrollment. The correlation is routinely used when observing two variables to determine whether there is a relationship between them (Gravetter & Wallnau, 2009). The statistic, correlation, describes the relationship. Three characteristics describe the correlation: *direction*, *form*, and *strength*. Direction indicates if the relationship is positive, negative, or no association when no relationship appears to exist. Form describes the pattern of the scatter plot: linear, nonlinear (or curvilinear).

Strength describes the degree to which the variables are related. A perfect correlation is identified with a resulting strength value of ± 1.00 , meaning that with each change in the independent variable, an equal change (either in the same or opposite direction) is observed in the dependent variable.

The first correlation used was the point-biserial correlation. Because the dependent variable, attrition, was dichotomous and the independent variable, distance, was quantitative, the point-biserial correlation coefficient (a special case of the Pearson correlation) was used. The point-biserial correlation coefficient, referred to as r_{pb} , was especially appropriate as the dependent variable, attrition, was dichotomous or binary; the only possible values for this variable were 1 (attrition) and 0 (retention). In a dichotomous or binary variable, any values other than 1 and 0 are undefined and do not have a corresponding data point.

Hypothesis Setup: The null hypothesis (H_0) was defined as: There is not a statistically significant relationship between the institutional distance from home and first-year attrition for traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford. The alternative hypothesis (H_1) was defined as: There is a statistically significant relationship between the institutional distance from home and first-year attrition for traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford: specifically, the likelihood of attrition increases as institutional distance from home increases.

Secondary research questions: Following the correlational analysis, further analysis was conducted to examine the relationship and to determine if a non-linear relationship existed between the institutional distance from home and attrition of first-

year students prior to their second year when controlling for sex, race or ethnicity, expected family contribution (EFC), parental education levels (specifically, first-generation college student status), residency status, housing status, SAT or ACT score, and market segmentation as defined by the institution. Binary logistic regression was selected for this analysis.

Binary logistic regression was used to explore the relationship between the binary dependent variable of attrition and an independent variable, while controlling for the other independent variable(s). Each model consisted of the dependent variable (attrition) and the independent variable (institutional distance from home). Each additional independent variable was added (sex, race and ethnicity, expected family contribution, parental education level, residency status, housing status, SAT and ACT score, and market segmentation). Logistic regression can test interaction effects or curvilinear relationships. An interaction effect exists when the effect of one independent variable on the dependent variable is affected by the level of a second independent variable (Ganzach, 1997). The relationship is usually, but not exclusively, categorized by the effect of the product of the independent variables on the dependent variable. Curvilinear relationships exist when there is an accelerated (positive or negative) effect between the independent variables and the dependent variable.

The binary logistic regression process previously described was used. All independent variables were entered simultaneously. However, additional tests for the statistical significance of the coefficients of the cross-product term are required. The first test examines whether there is a multiplicative interaction between the variables and whether that affects the probability or likelihood of the outcome (interaction). The second

test involves the squaring of the variable to produce a new variable that tests for the curvilinear relationship. Both of the variables (squared and not squared) were entered to test the linear and curvilinear relationships of the predictors to the dependent variable. Following completion of the point-biserial correlation, and subsequent partial correlation analyses, the relationship was tested while controlling for the following variables: sex, race and ethnicity, expected family contribution (EFC), parental education level, residency status, housing status, SAT and ACT score, and market segmentation using additional logistic regression analyses.

Hypothesis Setup: Logistic regression predicts the odds of the outcome based on the values of the independent variables (predictors) such that the odds are defined as the probability (or likelihood) that a particular outcome is a case divided by the probability that it is a non-case. Therefore, the development of a model that predicts the likelihood of the outcome better than the chance or random prediction level becomes the basis for the null hypothesis and hypothesis. The predictor is the institutional distance from home. The outcome is attrition of first-time, full-time students at the University of Pittsburgh at Bradford prior to their second year of enrollment.

The null hypotheses (H_{0a} through H_{0h}) were defined as: There is no relationship between the institutional distance from home and attrition of first-time, full-time students at the University of Pittsburgh at Bradford when controlling for sex, race and ethnicity, expected family contribution (EFC), parental education level, residency status, housing status, SAT and ACT score, and market segmentation, respectively. The alternative hypotheses (H_{1a} through H_{1h}) were defined as: There is a significant relationship between the institutional distance from home and attrition of first-time, full-time students at the

University of Pittsburgh at Bradford when controlling for sex, race and ethnicity, expected family contribution (EFC), parental education level, residency status, housing status, SAT and ACT score, and market segmentation: specifically, the likelihood of attrition increases as institutional distance from home increases when controlling for pre-matriculation information.

Data elements

This *ex-post facto* research, correlational analysis uses pre-existing institutional data or data calculated based on pre-existing institutional data.

Attrition: For the purposes of this case study, attrition is the result of the observation that a traditional-aged, first-time, full-time student fails to return for their respective second year. Retention to the second year was coded with a value of *0*. Attrition after the first or second semester was coded with a value of *1*. This assignment was chosen because the analysis was conducted using binary logistic regression, which “predicts the *1* value of the dependent variable using the *0* level as the reference value” (Garson, 2010, p. 2). Prediction of attrition is the goal. Therefore, an observation of attrition will result in a value of *1*.

Address: Students lacking a permanent U.S. residence were eliminated from the population so as not to skew distance analysis. Eliminations included students with a permanent address in Guam, Puerto Rico, the U.S. Virgin Islands, and U.S. military bases located outside of the continental United States and Hawaii. Addresses listing a post office box pose a problem in the accurate measurement of the institutional distance from home since the post office is not a student’s home. However, the researcher believes that in most cases, the difference in the distance from home to post office is negligible.

DistanceD: This independent variable was derived by calculating the distance from each student's home address to the institution based on latitude and longitude within Microsoft Excel using the Google Maps Distance Matrix Application Programming Interface (API) in Visual Basic (VBA). This measurement is known as the great-circle or orthodromic distance (the shortest distance between two points on the surface of a sphere); a colloquialism for this method is "as the crow flies."

DistanceF: This independent variable was derived by calculating the distance from each student's home address to the institution based on the recommended driving route as defined by the Google Maps API and calculated by the Google Maps Distance Matrix API. This second calculation of distance was accomplished within Microsoft Excel using the Google Maps API and Google Maps Distance Matrix API in VBA. Based on the literature review, this method had not been previously used in measuring distances in educational research. This methodology adds to the research literature.

DistanceT: This independent variable was derived by calculating commute time from each student's home address to the institution based on the recommended driving route as defined by the Google Maps API and calculated by the Google Maps Distance Matrix API. This second calculation of distance was accomplished within Microsoft Excel using the Google Maps API and Google Maps Distance Matrix API in VBA. Based on the literature review, this method also adds to research literature, as it has not been previously used.

Expected Family Contribution (EFC): The independent variable, EFC, is an estimate of the ability of parents, guardians, and student to contribute to post-secondary education expenses. EFC is calculated using the Federal Methodology, which takes into account

income, some assets, expenses, family size, and other factors. The data used to calculate EFC is self-reported and collected as part of the Free Application for Federal Student Aid (FAFSA). The EFC is reported on each respective student's Institutional Student Information Record (ISIR), which was digitally uploaded to the institution's student information system (SIS).

Sex: This independent variable was self-identified by each student, having been requested on the University's application for admissions. Students who had self-identified as female were coded with a value of 0 and students who had self-identified as male were coded with a value of 1. Students for which this value was missing were eliminated from the study.

Housing Status: This independent variable indicates whether the student, during the first semester of full-time enrollment, was living on campus in an institutionally owned and controlled residence hall (coded with a value of 1) or if the student was commuting (from either home or a rented facility not owned or controlled by the institution (coded with a value of 0)).

Market Segmentation: This independent variable indicates the market in which a student's permanent home address is located. The institution has defined a primary market, a secondary market, and a tertiary market. The primary market comprises the county in which the institution is located (McKean (PA)) and those counties that are contiguous to the county in which UPB is located (Cameron (PA), Elk (PA), Forest (PA), Potter (PA), Warren (PA), Allegany (NY) and Cattaraugus (NY)). The secondary market comprises those counties that are contiguous to the primary market (Allegheny (PA), Armstrong (PA), Beaver (PA), Bradford (PA), Butler (PA), Centre (PA), Clarion (PA),

Clearfield (PA), Clinton (PA), Crawford (PA), Erie (PA), Indiana (PA), Jefferson (PA), Lawrence (PA), Lycoming (PA), Mercer (PA), Tioga (PA), Union (PA), Venango (PA), Chautauqua (NY), Chemung (NY), Erie (NY), Livingston (NY), Steuben (NY), Wyoming (NY)). The tertiary market comprises all other areas. The category of Market was coded as 1 for residents and 0 for non-residents for each segmentation category: primary, secondary, and tertiary.

Parental Education Levels: This independent variable designates if a student was considered first-generational, which is defined as neither parent nor guardian having graduated from a four-year college (National Center for Education Statistics, 2010). The data was self-reported and collected as part of the Free Application for Federal Student Aid (FAFSA). The educational attainment level was reported for both mother and father on each respective student's Institutional Student Information Record (ISIR), which was uploaded to the institution's student information system (SIS).

Race/Ethnicity: The data for this independent variable, requested on the University's application for admissions, was self-reported by the student. The University updated its application in 2007 to comply with changes made in IPEDS reporting. Changes in classification and coding options were made to applications for admission to the University starting with the recruiting cycle for the fall of 2008. The categories were changed to the following: race and ethnicity unknown (coded with a value of 1), Hispanic or Latino or Spanish origin of any race (coded with a value of 2), American Indian or Alaskan Native (coded with a value of 3), Asian (coded with a value of 4), Black or African American (coded with a value of 5), Native Hawaiian or Other Pacific Islander (coded with a value of 6), White (coded with a value of 7), and designations of more than

one racial/ethnic category (coded with a value of 8 if neither race was Hispanic).

Furthermore, a binary category was created. Values that corresponded to the student self-reported category of White, Non-Hispanic were coded with a value of 1. Those students who self-reported a category other than White, Non-Hispanic were coded with a value of 0. Self-reported data came from admissions applications. Students identifying as non-resident alien (of any race or ethnicity) were not included in this case study (National Center for Education Statistics, 2010).

Residency: This independent variable indicated the student's home state. The University assesses tuition charges based on the residency data. However, in some cases, students who self-identified their permanent home address as *Pennsylvania* may still have been assessed out-of-state tuition charges if residency had not been established in accordance with university policy. There is a difference of an approximately \$10,000 between in-state tuition and out-of-state tuition. This variable was coded as 0 for in-state tuition assessment and 1 for out-of-state tuition assessment.

SAT and ACT Scores: The SAT composite score is an independent variable based on the sum of the scores earned on the mathematics and critical reading sections of the test (i.e., Math and Critical Reading). The highest Math and the highest Verbal scores, for each student, were combined to create the composite score used for analysis. The highest ACT Composite test scores were converted to an SAT equivalent composite score using the ACT-SAT Concordance Tables (ACT, 2007; The College Board, 2009). Although both sets of tests measure similar outcomes, the scores cannot be exchanged because ranges and measures differ. However, scaling is used to correspond scores based on percentile position or distribution rank (Dorans, 2000). The ACT and The College Board used

scaling to develop concordance tables that have a correlation value of .92. The concordance table that was used in this study “provides concordance between the ACT composite scores and SAT critical reading and mathematics scores” (Dorans, 2008; The College Board, 2009).

Table 1

Description of Independent Predictor Variables

Variable	Description
DistanceT	Scale value (1-2,280)-Minutes of driving time from home to the institution based on the self-reported address at the time of matriculation and calculated using the Google Maps API and Google Maps Distance Matrix API scripts and recommended route
DistanceD	Scale value (.21-2,318.92)-Great-circle or orthodromic distance in miles from home to the institution based on the self-reported address at the time of matriculation and calculated using Google Maps Distance Matrix API script.
DistanceF	Scale value (.31-2656.98)-Distance in miles from home to the institution based on the self-reported address at the time of matriculation and calculated using the Google Maps API and Google Maps Distance Matrix API scripts and recommended route.
EFC	Scale Value (0-99,410)-Expected Family Contribution as reported in the ISIR based on self-reported data from the FAFSA.
SAT/ACT	Scale Value (510-1,510)-SAT Score (Math + Verbal) or conversion of ACT score based on the Concordance. Highest combined totals reported are used.
In-State Residency	Binary value-Students whose permanent address at the time of matriculation was in the state of Pennsylvania (1) versus those whose were not (0)
Primary Market	Binary value-Students whose permanent address at the time of matriculation was in the primary market as defined by the institution (1) versus those whose were not (0)
Secondary Market	Binary value-Students whose permanent address at the time of matriculation was in the secondary market as defined by the institution (1) versus those whose were not (0)
Tertiary Market	Binary value-Students whose permanent address at the time of matriculation was in the tertiary market as defined by the institution (1) versus those whose were not (0)
Sex-Male	Binary value-The student is male (1) or is female or did not report (0). This datum is based on self-reported data from the admissions application.
First-Generation Status	Binary value-The student does not have a parent with a college degree (1) or does (0). This datum is derived from ISIR data based on self-reported data from the FAFSA.
On-campus Housing	Binary value-The student lives on-campus (1) or does not (0).
White-Non-Hispanic	Binary value-The student reports White, Non-Hispanic (1) or does not (0). This datum is based on self-reported data from the admissions application.

Source of data and collection procedures

Data for this case study were obtained from the University of Pittsburgh data warehouse, an exact copy of the University's student information system (SIS) that is stored in an Oracle database. The data warehouse is updated nightly. The researcher obtained approval from the University of Nebraska, Lincoln, Institutional Review Board and the University of Pittsburgh Institutional Review Board to conduct this study and use the database.

The researcher wrote Structured Query Language (SQL) scripts to extract the data from the data warehouse; the scripts were run in a report writer called Sybase InfoMaker. The resulting data were saved to the researcher's desktop as comma-delimited files and then converted for use in Microsoft Excel and SPSS.

Google APIs

The Google Maps Application Programming Interface (API) and Google Maps Distance Matrix API are services that developers use to calculate the travel distance and time between an origin address and a destination address. Results calculated by the Google Maps Distance Matrix API are based on the recommended route between points as determined by the Google Maps API (The Google Distance Matrix API, n.d.).

The Google Maps API and the Google Maps Distance Matrix API were accessed with Microsoft Excel using VBA. The researcher modified and combined VBA that was developed by Ashish Koul (2012) and Maurice Calvert (2011).

There were limitations to the free version of this service: 100 elements per query, 100 elements per 10 seconds, and 2,500 elements per 24 hour period (The Google Distance Matrix API, n.d.). The researcher used the free API services and worked within

the limits in two ways: first, to overcome the limitation of 100 elements per 10 seconds, a VBA function, *Application.Wait(Time)*, was added. This function paused the process after each element such that the 100-element limit would never be reached. Second, the limit of 2500 elements per 24 hour period was overcome by dividing the data into subsets smaller than 2500 elements, processing each subset over a series of three days, and then recombining the results into a single data set.

Missing Data

Cohen and Cohen (1983) popularized the use of dummy variables to code for missing data. However, that methodology is no longer in popular practice because it can produce biased parameter estimates and reduce statistical power (Jones, 1996). Rubin (1976) defined three different categories of missing data: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).

Data missing completely at random represent those missing Y values for which there is no relationship between the probability of the missing Y value and either an X value or the Y value itself. Data missing at random represent missing Y values for which there is no relationship between the probability of the missing Y value and an X value. However, data missing not at random represent missing Y values because of the unobserved value of Y.

Rubin stated that data missing at random (MAR and MCAR) can be ignored. While random missing data do reduce sample size and degrees of freedom, they do not skew the results. However, data that is missing not at random (MNAR) could potentially introduce a strong bias.

For the purposes of this case study, an assumed likelihood existed that independent variables would be missing. Independent variables that were less likely to be missing included sex, race or ethnicity, SAT and ACT scores, and high school grades (grade point average).

The most likely missing independent variables were those associated with parental education levels and expected family contribution (EFC). These data are not collected from students who do not file a FAFSA (Free Application for Federal Student Aid). Research has shown a strong positive correlation between adjusted gross income and education levels (Baum & Ma, 2007). It is logical (albeit not foolproof) to assume that students who do not file a FAFSA are using alternative personal and familial resources to pay for their education. Therefore, the researcher believes that missing data pertaining to parental education levels, and expected family contribution (EFC) could be classified as missing not at random (MNAR) and could result in a scenario in which students at the upper extreme are more likely to be missing (i.e., students with the greatest proportion of parents with a college education, and students from families with the highest adjusted gross incomes and highest expected family contributions).

Traditional approaches to handling missing data include listwise deletion, pairwise deletion, mean imputation, regression-based imputation, multiple imputation, hotdecking, use of selection models, and application of EM algorithm. The recommended course of action is to use listwise deletion if fewer than 10% of the values are missing. In this approach, each row (representing an observation (i.e., a student)), in which data is missing, is deleted and not used in the analysis (Lynch, 2003). Listwise deletion was implemented in this study, as less than 3% of data were missing.

Descriptive Statistics

The total number of all domestic, traditional-aged, first-time, full-time students (freshmen) matriculated and enrolled during the fall semesters of 2005 through and including 2013 at the University of Pittsburgh at Bradford was 3,001. The number of observations used was 2,837. The 164 observations that were not used were removed because the data was incomplete. Final coding, calculation of DistanceT, DistanceD, and DistanceF, and review of the data were conducted in Excel. The resulting dataset was imported into SPSS, which was used for all statistical analyses.

As described in Chapter 3, the dependent variable in this study was attrition of traditional-aged, first-time, full-time students prior to their second year. Attrition is a dichotomous or binary dependent variable for which a value of 1 was assigned if attrition occurred; 0 was assigned if attrition did not occur prior to the second year of enrollment (i.e., retention): attrition was treated as a case and non-attrition (or retention) as a non-case. Of the 2,837 observations, the frequency distribution showed that 836 (29.5%) students left the institution (resulting in attrition) prior to the second year of enrollment (case), and 2,001 (70.5%) remained (non-case).

Tables 2 through 5 list the population's descriptive statistics for the independent variables included in the analyses. Descriptive statistics and histograms reveal that the categorical independent variables and the non-categorical variable SAT/ACT score contained normally distributed data. However, skewness, the degree to which the distribution is symmetrical, and kurtosis, the degree to which the shape of the data distribution matches a Gaussian distribution, were greater than the values associated with a normal distribution (0 for skewness and 3 for kurtosis) for the variables of EFC,

DistanceD, DistanceF, and DistanceT. For each of those values, skewness was greater than 0, meaning that there was a right-skewed distribution (i.e., most of the values were concentrated to the left of the mean, with extreme values to the right), and kurtosis was greater than 3, which indicated a leptokurtic distribution: one with values clustered around the mean creating a higher peak and thicker tails. However, Wheeler (2004), citing his research and also the research of Walter Shewhart, stated that skewness and kurtosis statistics simply lend added insights into the shape of the data and are not especially useful otherwise. In this case study, the resulting skewness and kurtosis values showed that a large percentage of the first-time, full-time students enrolled at the University of Pittsburgh at Bradford were from similar distances (both in terms of mileage and travel times): a large proportion of students came from distances near the mean.

Table 2

Population Descriptive Statistics

Cohort (Fall)	<i>n</i>	Attrition	Female %	Male %	1st Generation %	Resident %	Commuter %
2005	208	27.4%	57.2%	42.8%	47.1%	75.0%	25.0%
2006	327	29.4%	50.8%	49.2%	41.9%	80.7%	19.3%
2007	285	27.4%	55.1%	44.9%	41.4%	87.0%	13.0%
2008	340	26.2%	50.6%	49.4%	45.9%	88.5%	11.5%
2009	356	31.2%	52.2%	47.8%	38.5%	83.1%	16.9%
2010	307	28.7%	49.5%	50.5%	43.3%	82.7%	17.3%
2011	338	29.9%	56.2%	43.8%	39.9%	86.1%	13.9%
2012	322	31.7%	54.3%	45.7%	38.5%	83.9%	16.1%
2013	354	32.2%	52.5%	47.5%	37.9%	87.6%	12.4%
Total	2837	29.5%	53.0%	47.0%	41.3%	84.2%	15.8%

Table 3

Race and Ethnicity: Descriptive Statistics

Fall Cohort	<i>n</i>	Asian, non-Hispanic %	Black or African American, non-Hispanic %	American Indian or Alaska Native, non-Hispanic %	Native Hawaiian or other Pacific Islander, non-Hispanic %	Hispanic %	Two or more races, non-Hispanic or Other %	White, non-Hispanic %
2005	208	0.5%	4.8%	0.5%	0.0%	1.9%	5.8%	86.5%
2006	327	1.5%	6.1%	0.3%	0.0%	1.8%	12.6%	77.7%
2007	285	1.4%	7.0%	0.4%	0.0%	0.4%	7.3%	83.5%
2008	340	2.4%	7.9%	0.6%	0.3%	3.5%	2.1%	83.2%
2009	356	1.7%	7.0%	0.8%	0.0%	2.2%	3.5%	84.8%
2010	307	2.9%	13.7%	0.3%	0.0%	3.3%	1.3%	78.5%
2011	338	2.7%	12.1%	0.0%	0.0%	3.6%	5.6%	76.0%
2012	322	3.4%	14.3%	0.3%	0.0%	4.3%	7.2%	70.5%
2013	354	3.1%	18.1%	0.6%	0.3%	5.9%	5.9%	66.1%
Total	2837	2.3%	10.4%	0.4%	0.1%	3.1%	5.6%	78.1%

Table 4

Residency and Market: Descriptive Statistics

Cohort (Fall)	<i>n</i>	In-State Residency %	Out-of-State Residency %	Primary %	Secondary %	Tertiary %
2005	208	84.6%	15.4%	51.0%	24.0%	25.0%
2006	327	78.6%	21.4%	38.2%	32.1%	29.7%
2007	285	81.8%	18.2%	34.7%	32.6%	32.6%
2008	340	81.5%	18.5%	32.1%	30.6%	37.4%
2009	356	83.4%	16.6%	36.5%	29.5%	34.0%
2010	307	76.9%	23.1%	31.6%	27.7%	40.7%
2011	338	82.2%	17.8%	36.7%	25.4%	37.9%
2012	322	82.0%	18.0%	35.7%	25.5%	38.8%
2013	354	83.9%	16.1%	29.9%	24.3%	45.8%
Cumulative	2837	81.6%	18.4%	35.6%	28.1%	36.3%

Table 5

Descriptive Statistics for Interval Independent Variables

Cohort (Fall)	<i>n</i>	<i>M</i>	Median	SD	Min.	Max.	Skewness	Kurtosis
SAT								
2005	208	1005.28	1010.00	122.09	640	1290	.046	-.281
2006	327	983.36	980.00	126.14	610	1310	-.044	.123
2007	285	974.32	980.00	142.23	523	1410	-.111	-.061
2008	340	990.00	990.00	140.08	570	1460	-.049	.247
2009	356	983.46	985.00	122.27	620	1450	.253	.890
2010	307	989.97	990.00	149.11	610	1510	.207	.078
2011	338	977.66	980.00	139.94	680	1470	.412	.334
2012	322	970.43	980.00	149.81	510	1470	-.018	.126
2013	354	966.84	960.00	153.85	530	1500	.125	.320
Cumulative	2837	981.38	980.00	139.57	510	1510	.076	.268
EFC (\$)								
2005	208	10022.00	6248.50	14270.47	0	96497	3.498	15.393
2006	327	9080.31	5542.00	14960.05	0	94687	3.820	18.053
2007	285	10648.12	6153.00	14313.68	0	96426	3.133	13.690
2008	340	12741.09	7107.00	16734.66	0	96622	2.266	5.800
2009	356	12123.71	5634.50	17484.69	0	96457	2.471	6.828
2010	307	10854.09	5582.00	17356.99	0	99410	3.195	11.914
2011	338	10008.28	5556.00	14575.62	0	96753	3.105	12.977
2012	322	11807.98	6349.50	16052.42	0	99332	2.473	8.142
2013	354	9192.36	4364.00	14975.49	0	96753	3.415	14.723
Cumulative	2837	10828.86	5664.50	15732.97	0	99410	2.935	10.677
DistanceD (Miles)								
2005	208	98.01	37.26	189.83	.036	1890.40	6.622	54.047
2006	327	106.75	59.96	207.17	.215	2197.90	7.174	61.339
2007	285	100.98	70.77	126.75	.348	1258.93	5.459	43.967
2008	340	119.99	101.14	183.81	.227	2178.78	7.628	75.639
2009	356	109.14	75.27	171.79	.366	1929.89	7.548	73.319
2010	307	126.39	103.41	175.75	.541	1909.07	5.892	48.003
2011	338	125.06	92.34	215.38	.354	2166.14	7.186	63.114
2012	322	125.89	81.81	206.00	.210	2196.94	6.963	62.579
2013	354	149.09	113.56	265.18	.480	2318.92	6.676	49.455
Cumulative	2837	119.18	81.71	198.82	.210	2318.92	7.224	65.404
DistanceF (Miles)								
2005	208	131.17	49.24	229.48	.81	2196.54	5.876	44.971
2006	327	150.42	83.26	280.52	.62	2530.84	6.280	46.324
2007	285	138.72	96.93	160.54	.93	1490.05	4.381	31.212
2008	340	164.00	138.26	225.94	.43	2635.85	6.759	64.064
2009	356	148.06	102.22	207.65	.87	2231.34	6.540	59.148
2010	307	173.07	137.32	217.00	.99	2236.31	4.981	37.544
2011	338	171.49	130.49	258.94	.31	2518.42	6.283	52.087
2012	322	170.99	112.78	249.55	.37	2534.57	6.032	50.433
2013	354	202.84	152.24	310.97	.93	2656.98	6.044	43.024
Cumulative	2837	163.00	109.98	243.94	.31	2656.98	6.311	52.976
DistanceT (Minutes)								
2005	208	139.71	65.00	206.14	2	1920	5.269	54.047

2006	327	157.34	98.00	246.30	2	2160	5.853	41.731
2007	285	149.49	111.00	148.04	3	1343	3.754	24.959
2008	340	174.41	160.00	201.87	2	2280	5.992	54.267
2009	356	159.37	115.00	188.64	3	1980	5.824	50.459
2010	307	181.43	157.00	194.19	3	1920	4.263	29.933
2011	338	178.44	145.00	227.25	1	2160	5.606	44.448
2012	322	180.02	132.50	220.89	2	2160	5.294	41.675
2013	354	208.43	178.00	270.70	2	2280	5.611	38.888
Cumulative	2837	171.63	127.00	216.34	1	2280	5.643	45.136

Data Analyses

In this study, the dependent variable, attrition, is dichotomous (or binary). A student will either not enroll beyond the first year (attrition) or will (retention). Because the purpose of this case study was to determine if a relationship exists between attrition (dependent variable or DV) (prior to the second year of enrollment) and the institutional distance from home for a student (independent variable or IV), the dependent variable was assigned the value of *1* if the student was not retained beyond the first year or a value of *0* if the student was retained. The nature of the dependent variable lends itself to analysis utilizing both point-biserial correlation and binary logistic regression. The results of the point-biserial correlation prompted the researcher to further investigate the possible role of the independent variable, SAT/ACT score, through a partial correlation analysis. The results of that analysis additionally supported the use of the binary logistic regression analysis to further explore the possibility of a relationship between attrition and institutional distance from home while controlling for pre-matriculation variables.

Table 6 shows the analyses that were proposed and ways in which the findings will be reported.

Table 6

Proposed Analyses and Findings to be Reported

Statistic	Measurement	Relevance
Point-Biserial Correlation	Point-Biserial Correlation Coefficient (r_{pb}) to determine the type of relationship	A positive number will indicate a positive relationship, and a negative number will indicate an inverse relationship.
	One-tailed t-test to determine significance	Correlation is significant if the value of the t-test is greater than the critical value with degrees of freedom (n-2) based on the researcher's level of significance ($p < 0.05$).
	Coefficient of Determination (r_{pb}^2) to determine strength of association between IV and DV	$r_{pb}^2 > 0.81$ = very strong, 0.49-0.80 = strong, 0.25-0.48 = moderate, and 0.00-0.08 = weak.
Partial Correlation	Partial Correlation Coefficient (r) to determine the type of relationship	A positive number will indicate a positive relationship, and a negative number will indicate an inverse relationship.
	One-tailed t-test to determine significance	Correlation is significant if the value of the t-test is greater than the critical value with degrees of freedom (n-2) based on the researcher's level of significance ($p < 0.05$).
Binary Logistic Regression	Overall Model Evaluation: Likelihood Ratio Test, Score Test, and Wald Test	Values that are significant at the 0.05 level will indicate the model offers improved prediction over chance.
	Statistical Tests of Individual Predictors: Wald Chi-Square Statistic	A value of $p < 0.05$ indicates that the predictor is significant, and the null hypothesis can be rejected.
	Goodness-Of-Fit Statistic: Hosmer-Lemeshow (H-L) Test	A value of $p > 0.05$ indicates that we fail to reject the null hypothesis that the model is not significantly different from observed values.
	Validations of Predicted Probabilities: Kendall's Tau-b, Goodman-Kruskal's Gamma, Somer's D, and the c statistic	The degree (%) to which the model results in fewer false predictions than chance can be reported if Kendall's Tau-b, Goodman-Kruskal's Gamma, and Somer's D values > 0 . An area under the receiver operating characteristic (ROC) curve (AUC or c statistic) result $> .50$, indicates that the model performance is better than chance.

Point-Biserial Correlation Process: Following standard practice (NCSS Statistical Software, n.d.) the alpha level was set at .05. The alpha level is the value for the probability of a type-I error (i.e., the researcher rejects the null hypothesis when it is, in fact, true—an effect that is not present is detected). A 0.05 alpha level means that the test will falsely reject the null hypothesis once in every twenty tests.

Power is the probability of rejecting a false null hypothesis. Power is equal to 1 minus Beta. Beta is the probability of a type-II error (i.e., when a false null hypothesis is not rejected—an effect that is present is not detected). A power of 0.80 (i.e., Beta = 0.20) was used.

The Pearson formula was used to calculate the point-biserial correlation as there is a close relationship between the resulting Pearson r^2 and the r_{pb}^2 of the point-biserial correlation (Pett, 1997). The r_{pb}^2 is used to measure the effect size. It is the percentage of variance in attrition that can be predicted from the variance in the institutional distance from home. Pett stated that the same criteria for evaluating the Pearson r^2 could be utilized for the evaluation of the point-biserial correlation r_{pb}^2 . The following intervals were utilized: $\geq .81$ = very strong, .49-.80 = strong, .25-.48 = moderate, and .00-.08 = weak.

However, even if the correlation coefficient suggests that a relationship exists, the correlation may not be statistically valid. To test the significance of r_{pb} , a one-tailed or two-tailed t-test for independent means is applied as part of the process. Significance is determined only when the null hypothesis can be rejected. Although r_{pb} equals 0 for the null hypothesis, a one-tailed or two-tailed t-test for independent means is used to determine if the null hypothesis can be rejected. The one-tail t-test was selected for the

analysis in this study because correlations are almost always, though not necessarily, directional. A research or alternative hypothesis usually states that an independent variable is positively or negatively correlated with the dependent variable as opposed to simply stating that the independent variable and the dependent variable are related. In this case study, the researcher hypothesizes that attrition and distance are positively correlated (i.e., as distance increases, the likelihood of attrition increases). The null hypothesis can be rejected if the value of the t-test is greater than the critical value associated with the relevant degrees of freedom ($n-2$) based on the researcher's level of significance (usually $p < .05$) (Gravetter & Wallnau, 2009).

To summarize, using the point-biserial correlation, the researcher initially explored the relationship between the institutional distance from home for traditional-aged, first-time, full-time students and first-year attrition at the University of Pittsburgh at Bradford. A relationship was determined if the r_{pb}^2 was greater than 0 and was considered significant if the t-test value was greater than the critical value associated with the relevant degrees of freedom ($n-2$).

Partial Correlation Process:

The results of the initial point-biserial correlation analysis of the data suggested a potential relationship between the SAT/ACT score and all other independent variables that were included in the study. The resulting point-biserial correlation coefficients between SAT/ACT scores and each of the other independent variables included in this study were significant at the ($p < .01$) except for the correlation between SAT/ACT score and secondary market designation ($p < .05$). The results suggested the possibility that the SAT/ACT score may have been acting as a confounding variable.

A confounding variable is an extraneous variable that correlates with both the dependent variable and the independent variable being analyzed and causes the observed relationship to be obscured or accentuated (MacKinnon, Krull, & Lockwood, 2000). To address the possibility that the predictor or independent variable SAT/ACT score was acting as a confounding variable, the researcher proceeded with an additional analysis: the partial correlation analysis.

The predictor variables were examined by computing partial correlations between the predictor variables and attrition, controlling for the effects of sex, race and ethnicity, expected family contribution (EFC), parental education level, residency status, housing status, SAT and ACT score, and market segmentation. The partial correlation was chosen so as to be able to eliminate the potential influence or distortion of a third variable on the measurement of any potential relationship between attrition and distance (Gravetter & Wallnau, 2009). Partial correlation analyses were used to further explore the relationships between the dependent and independent variables in this study. Partial correlation analysis examines the relationship between two variables of interest after statistically removing or partialling out the effects of a third (or more) variable (Collican, 2014). It is a way to attempt to eliminate confounding variables.

In this study, this analysis enabled the researcher to determine if a third (or more) variable affected or even created the relationship between the dependent variable, Attrition, and the independent variables of interest: DistanceD, DistanceF, and DistanceT. If the relationship between a dependent variable and an independent variable is decreased by eliminating the effects of a second independent variable, then it is possible to conclude

that the relationship between the variables of interest is spurious (Bay & Hakstianz, 1972).

Upon completion of the partial correlation analysis, the researcher determined that the SAT/ACT scores was acting as a confounding variable. Besides the partial correlation, there are several other methods by which confounding variables can be controlled; one of which is logistic regression analysis.

Binary Logistic Regression Process: As was stated in the literature review, work by Astin (1975) and Pantages and Creedon (1978) suggested the possibility of a nonlinear relationship between distance and attrition (i.e., students very close to home and students very far from home are less likely to be retained). An additional statistic was required to test the hypothesis to determine if a non-linear (or curvilinear) relationship existed.

The third statistic that was utilized was the binary logistic regression. Binary logistic regression measures the relationship between a categorical dependent variable and one or more independent variables. The independent variables are often, but not exclusively, continuous. The relationship is measured by using probability scores as the predicted values of the dependent variable. Logistic regression predicts the odds of the outcome based on the values of the independent variables. The odds (or likelihood) are defined as the probability that a particular outcome is a case divided by the probability that it is a non-case. Binary (or Binomial) logistic regression is utilized when the observed outcome for a dependent variable can have only two possible types. Logistic regression is often used when the relationship between the dependent variable and a predictor may be non-linear. Logistic regression uses a maximum likelihood method to find an equation that is “best-fitting” (Burns & Burns, 2009). Furthermore, it avoids four

potential problems that result from using a simple linear regression model when modeling dichotomous outcome variables. First, nonconforming predicted probabilities can occur (i.e., the linear probability model can predict probabilities outside the range of 0 and 1, which is outside of the logical range). Second, heteroscedasticity can occur.

Heteroscedasticity is the unequal variability of the dependent variable across the range of values of the independent variable (i.e., errors are low near $Y = 0$ or 1 but high in the middle). Third, non-normal errors occur: errors can never be normally distributed due to the nature of the dichotomous outcome and therefore, hypothesis testing is problematic. Fourth, linear regression models the probability of a linear relationship. The functional form of the probability of the dichotomous outcome is an S-shaped curve (Rueda, 2014; Lunt, 2013).

This study satisfied the assumptions that Aldrich and Forrest (1984) described: the dependent variable is dichotomous and binary; the outcomes are mutually exclusive, and irrelevant predictors will be excluded; and the population that is studied will include a minimum of 50 data points per predictor. The benefits of using logistic regression over ordinary least squares regression or linear discriminate function analysis include the following: a linear relationship between the independent and dependent variables is not assumed, and a nonlinear effect can be measured; neither the dependent nor the independent variables must be normally distributed; and homogeneity of variance is not assumed (Morgan & Teachman, 1988). Although logistic regression does compute correlation measures to estimate the strength of the relationship, measures of correlations do not inform us of the accuracy or errors associated with the model.

To determine the effectiveness of the model, the following must be evaluated and should be reported as part of the findings: “(a) overall model evaluation, (b) statistical tests of individual predictors, (c) goodness-of-fit statistics, (d) validations of predicted probabilities” (Peng, Lee, & Ingersoll, 2002, p 5.).

The overall model evaluation is the process by which the logistic regression model is compared to the intercept-only model. If an increase in effectiveness is shown after examining the likelihood ratio, score, and Wald tests, then the logistic regression model “provides a better fit to the data” (Peng, Lee, & Ingersoll, 2002, p 5.; Menard, 1995). Peng et al. (2002) stated that p-values smaller than 0.05 (i.e., $p < .05$) indicate that a relationship between the independent and dependent variables likely exists (i.e., the independent variable influences the dependent variable). The statistical test of individual predictors (i.e., variables) involves examination of the Wald chi-square statistic. A test result ($p < .05$) indicates that the predictor is significant and that the null hypothesis can be rejected (Burns & Burns, 2009).

The goodness-of-fit test statistics, which involve testing the fit of the model to the outcomes, may also be used to reject the null hypothesis. The Hosmer-Lemeshow (H-L) test is used. The H-L statistic is a Pearson chi-square statistic, and a result ($p > .05$) indicates that there is a good fit between model and data. The result is that we fail to reject the null hypothesis that the estimated outcome probabilities of the model agree with the empirical outcome probabilities.

A validation of predicted probabilities involves the revalidation of the predicted probabilities against the actual outcomes. Peng et al. (2002) recommended the inclusion of a measure of association or classification table as part of the findings. Measures of

association describe the degree to which two variables are related. The classification table is recommended if classification is a stated goal of the analysis. Most statistical analysis software packages provide a classification table and measures of association as part of the results of a logistic regression analysis: Kendall's tau-b, Goodman-Kruskal's Gamma, Somer's D statistic, and c statistic. Kendall's tau-b is a rank order correlation coefficient without adjustment for ties. Goodman-Kruskal's Gamma (symmetric) and Somer's D (asymmetric) are used with ordinal level variables and can account for concordant and discordant pairs and ties. The c statistic (concordance index) is the "proportion of observation pairs with different observed outcomes for which the model correctly predicts a higher probability for observations with the event outcome than the probability for nonevent observations" (Peng & So, 2002, p. 46). The c statistic is also known as A^1 (i.e., a-prime) or as the area under the receiver operating characteristic (ROC) curve (AUC).

A primary goal of logistic regression is to generate a model that can reliably assign observations into one of the two outcomes. Visualizing predictions versus data can be accomplished using a receiver operating characteristic (ROC) curve. The ROC curve is a plot of sensitivity versus 1-specificity (Hosmer & Lemeshow, 2000). Sensitivity is the proportion of observations correctly classified as an event (i.e., the ability of the model to correctly predict attrition, also known as the true positive rate) whereas specificity is the proportion of observations correctly classified as a nonevent (i.e., the ability of the model to correctly predict non-attrition (or retention)) and 1-specificity is the false positive rate. The ROC curve shows the tradeoff between sensitivity and specificity. As the cutoff or threshold changes, the increase in sensitivity will be

accompanied by a decrease in specificity. The researcher selected to utilize the area under ROC curve (AUC) (i.e., c statistic) as an additional overall measure of the model performance and to report the validation of the predicted probabilities. An AUC = 1 is a perfect model. An AUC = 0.50 is a model that is no better than chance.

Statistical Analyses Limitations

Point-Biserial Correlation: There are limitations to the point-biserial correlation. The first is that the point-biserial coefficient is sample-dependent and assumes a degree of homogeneity. A second limitation is that a broad range of values for the continuous variable will result in an increase in the size of the correlation. Conversely, a small range will reduce the size of the correlation. A third limitation is that predicted values are limited to the range of the original values. A fourth limitation is that the point-biserial correlation cannot reach a value of ± 1 unless you are working with a dataset that includes only 2 data points or observations. The maximum correlation for large datasets in which the independent variable is normally distributed is .85.

The last limitation is linearity assumption (Miles & Shevlin, 2001). The correlation coefficient (r_{pb}) requires a linear relationship between the two variables being observed. The correlation coefficient provides a reliable measure of the strength of the relationship (r_{pb}^2) if the observed relationship appears to be linear. If the observed relationship appears to be nonlinear, then the correlation coefficient is not useful. This final limitation is important in this study and was the impetus for analyzing the data using a scatter plot. As was stated in the literature review, work by Astin (1975) and Pantages and Creedon (1978) suggest that a nonlinear relation may exist between distance and attrition (i.e., students very close to home and students very far from home are less likely

to be retained). A different statistic was used to determine if a non-linear (or curvilinear) relationship existed.

Partial Correlation: The partial correlation analysis is somewhat limited by three basic assumptions of the method. First, like the point-biserial correlation, a linearity assumption exists. The gross or zero-order correlations must have linear regressions. The correlation coefficient (r) requires a linear relationship between the two variables being observed. The correlation coefficient provides a reliable measure of the strength of the linear relationship (r^2) if the observed relationship appears to be linear. A second limitation of this type of analysis is that it assumes that the effects of the independent variables are separate, distinctive, and additively, and not jointly, related. The third limitation is that partial correlation analysis requires a large number of observations because the reliability of partial coefficients decreases as order increases.

Binary Logistic Regression: There are limitations on the use of binary logistic regression. First, the observed outcome must be discrete, and the desired outcome should be coded a 1 . In this study, instances of attrition are observed and coded as 1 or 0 . A second limitation is that large sample sizes are generally required. This limitation is especially important as independent variables are added. Small sample sizes result in low power when the Hosmer–Lemeshow test is applied. Hosmer and Lemeshow (2000) recommend sample sizes greater than 400. A third limitation is that the model should be fitted correctly using a stepwise method so that all significant variables are included, and only significant variables are included. A fourth limitation is that logistic regression requires that each observation be independent and that there is little or no multicollinearity. However, multicollinearity can be factored into the model and

adjustments may be made by centering the variables or conducting a factor analysis with orthogonally-rotated factors (Lani, n.d.). Another limitation is that logistic regression may only be used in between-subject designs.

The last limitation is linearity assumption (Miles & Shevlin, 2001). Although logistic regression does not assume linearity between the dependent and independent variables, there is an implicit assumption of linearity in terms of the logit function among the independent variables.

Summary

The purpose of this *ex-post facto* case study was to determine the existence of a relationship between the institutional distance from home and the attrition of traditional-aged, first-time, full-time students prior to the second year at the University of Pittsburgh at Bradford, one of the four-year campuses of the University of Pittsburgh. Findings were generated as a result of three statistical analyses.

First, the point-biserial correlation was used to determine if a linear relationship existed between the independent variable (the institutional distance from home) and the dependent variable (attrition). That test was conducted three times: for orthodromic distance, for actual distance measured using the recommended driving route as defined by the Google Maps API and calculated by the Google Maps Distance Matrix API, and as measured in travel time (minutes) using the same recommended route. The findings of those tests are reported in Chapter 4.

Second, because results suggested a potential relationship of the SAT/ACT score to all other independent variables that were included in the study, partial correlation analyses were conducted between the independent variables and attrition, controlling for

the effects of sex, race and ethnicity, expected family contribution (EFC), parental education level, residency status, housing status, SAT and ACT score, and market segmentation. The partial correlation was conducted so that confounding or suppressing variables could be eliminated.

Next, because a nonlinear or curvilinear relationship not detected by the point-biserial correlation could have existed, binary logistic regression analysis was used to further explore the potential for a relationship between the institutional distance from home (in mileage and in travel time) and attrition. Logistic regression results in the development of a model that predicts whether the likelihood of the outcome is better than chance. The findings of that test are reported in Chapter 4.

Finally, any relationship discovered between the institutional distance from home and the attrition of traditional-aged, first-time, full-time students prior to the second year at the University of Pittsburgh at Bradford was further examined, using binary logistic regression analysis, while controlling for the following variables: sex, race and ethnicity, expected family contribution (EFC), parental education level, residency status, housing status, SAT and ACT score, and market segmentation, as defined by the institution.

Regardless of the outcomes, this study adds to the literature and the body of knowledge about attrition and retention factors, the methodology for determining distances, and quantitative research. Recent studies exploring the possible relationship between the distance from home and attrition are lacking, even though the profile of post-secondary education students has changed as access has expanded. Additionally, traits and values associated with the millennial generation differ (Howe & Strauss, 2000) from

those of the generations with which the majority of previous research had been conducted.

The utilization of a new methodology for measuring distance and exploring the relationship between distance and attrition using point-biserial correlation and binary logistic regression also adds to the literature. The results guide the future use of the variable of institutional distance from home in future studies of attrition, retention, and the development of predictive models. Furthermore, the study may serve as the basis for similar research to be conducted at other post-secondary institutions. Certainly, additional research will be necessary; institutional characteristics, including but not limited to size, location, urban-rural classification, affiliation, offerings, amenities, proximity to transportation, student profile, and the degree to which the institution is populated by legacy, all likely impact results.

Finally, the results of this study may provide information for the institution being studied that could lead to changes in recruitment, first-year experience programming, services provided, or strategic planning.

Chapter 4

Results of Data Analysis

Introduction

An *ex-post facto* quantitative research case study design was used to determine if a relationship existed between the institutional distance from home and attrition of traditional-aged, first-time, full-time students prior to their second year at the University of Pittsburgh at Bradford. Following the correlational analysis, further analysis was conducted to determine if the relationship was non-linear and if it existed between the institutional distance from home and attrition of first-year students prior to their second year when controlling for sex, race or ethnicity, expected family contribution (EFC), parental education levels (specifically, first-generation college student status), residency status, housing status, SAT or ACT score, or market segmentation as defined by the institution. Complete data for 2,837 individuals from a population of 3,001 domestic, traditional-aged, first-time, full-time matriculated and enrolled students (freshmen) were used in the analyses.

This chapter provides the findings used to answer the research questions in this study. The central research question that this case study aimed to answer is the following: Does a relationship exist between the institutional distance from home and the attrition of traditional-aged, first-time, full-time students prior to the second year at the University of Pittsburgh at Bradford, and if so, would additional analyses show that the relationship would still exist when attempting to control for other variables that, according to a review of the literature, have been shown to influence attrition and retention? The following sub-

questions were developed to examine the relationship while controlling for these variables:

1. Will controlling for sex affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
2. Will controlling for race or ethnicity affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
3. Will controlling for expected family contribution (EFC) affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
4. Will controlling for parental education levels (specifically, first-generation college student status) affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
5. Will controlling for residency status affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
6. Will controlling for housing status affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?

7. Will controlling for SAT or ACT score affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?
8. Will controlling for market segmentation, as defined by the institution, affect the relationship, if any, of distance from home as it relates to traditional-aged, first-time, full-time attrition before the second year at the University of Pittsburgh at Bradford?

Point-Biserial Correlation

The central research question that this case study aimed to answer is: Does a relationship exist between the institutional distance from home and the attrition of traditional-aged, first-time, full-time students prior to the second year at the University of Pittsburgh at Bradford? Point-biserial correlation analyses were conducted for the cumulative data set and for each year of data. As previously stated, the point-biserial correlation is a special case of the Pearson correlation. The Pearson correlation in SPSS was used and resulted in the outcomes of the analysis of the cumulative dataset in Table 7. The outcomes of the analysis by year are presented in Table 8.

Table 7

Point-Biserial Correlation Matrix (Cumulative)

	Correlation													
	ATTRITION	DistanceT	DistanceD	DistanceF	EFC	SAT	IS-RES	PRIM	SEC	TERT	MALE	First Gen	Campus	White
ATTRITION	1	.001	.001	.000	-.058**	-.121**	-.044**	-.029	-.010	.038*	.003	.043*	-.058**	.006
DistanceT	.001	1	.972**	.996**	.114**	.142**	-.281**	-.428**	-.100**	.520**	.050**	-.049**	.277**	-.296**
DistanceD	.001	.972**	1	.979**	.098**	.133**	-.293**	-.346**	-.103**	.440**	.043*	-.043*	.219**	-.258**
DistanceF	.000	.996**	.979**	1	.104**	.135**	-.284**	-.385**	-.111**	.487**	.048**	-.043*	.244**	-.280**
EFC	-.058**	.114**	.098**	.104**	1	.177**	-.099**	-.155**	.018	.138**	.075**	-.224**	.110**	.070**
SAT	-.121**	.142**	.133**	.135**	.177**	1	-.128**	-.150**	.043*	.109**	.170**	-.137**	.118**	.051**
IS-RES	-.044**	-.281**	-.293**	-.284**	-.099**	-.128**	1	.353**	-.060**	-.296**	-.045**	.099**	-.158**	.219**
PRIM	-.028	-.428**	-.346**	-.385**	-.155**	-.150**	.353**	1	-.465**	-.562**	-.077**	.101**	-.517**	.248**
SEC	-.010	-.100**	-.103**	-.111**	.018	.043*	-.060**	-.465**	1	-.471**	.031*	-.043*	.227**	.114**
TERT	.038*	.520**	.440**	.487**	.138**	.109**	-.296**	-.562**	-.471**	1	.048**	-.060**	.302**	-.354**
MALE	.003	.050**	.043*	.048**	.075**	.170**	-.045**	-.077**	.031*	.048**	1	-.072**	.027	.003
First Gen	.043*	-.049**	-.043*	-.043*	-.224**	-.137**	.099**	.101**	-.043*	-.060**	-.072**	1	-.077**	.015
Campus	-.058**	.277**	.219**	.244**	.110**	.118**	-.158**	-.517**	.227**	.302**	.027	-.077**	1	-.126**
White	.006	-.296**	-.258**	-.280**	.070**	.051**	.219**	.248**	.114**	-.354**	.003	.015	-.126**	1

** Correlation is significant at the 0.01 level (1-tailed).

* Correlation is significant at the 0.05 level (1-tailed).

	Sig. (1-tailed)													
	ATTRITION	DistanceT	DistanceD	DistanceF	EFC	SAT	IS-RES	PRIM	SEC	TERT	MALE	First Gen	Campus	White
ATTRITION		.479	.470	.495	.001	.000	.009	.062	.305	.022	.438	.010	.001	.383
DistanceT	.479		0.000	0.000	.000	.000	.000	.000	.000	.000	.004	.004	.000	.000
DistanceD	.470	0.000		0.000	.000	.000	.000	.000	.000	.000	.012	.011	.000	.000
DistanceF	.495	0.000	0.000		.000	.000	.000	.000	.000	.000	.005	.012	.000	.000
EFC	.001	.000	.000	.000		.000	.000	.000	.174	.000	.000	.000	.000	.000
SAT	.000	.000	.000	.000	.000		.000	.000	.010	.000	.000	.000	.000	.003
IS-RES	.009	.000	.000	.000	.000	.000		.000	.001	.000	.009	.000	.000	.000
PRIM	.062	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
SEC	.305	.000	.000	.000	.174	.010	.001	.000		.000	.049	.011	.000	.000
TERT	.022	.000	.000	.000	.000	.000	.000	.000	.000		.005	.001	.000	.000
MALE	.438	.004	.012	.005	.000	.000	.009	.000	.049	.005		.000	.072	.428
First Gen	.010	.004	.011	.012	.000	.000	.000	.000	.011	.001	.000		.000	.216
Campus	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.072	.000		.000
White	.383	.000	.000	.000	.000	.003	.000	.000	.000	.000	.428	.216	.000	

The null hypothesis presented in Chapter 1 is as follows: There is not a statistically significant relationship between the institutional distance from home and first-year attrition for traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford. Upon completion of the initial analysis of the cumulative dataset, no statistically significant relationship was found between Attrition and DistanceT ($r_{pb} = .001, p = .479$), Attrition and DistanceD ($r_{pb} = .001, p = .470$), or Attrition and DistanceF ($r_{pb} = .000, p = .495$). The result is that we fail to reject the null hypothesis. However, the matrix showed a statistically significant, but weak relationship between attrition and the

tertiary market, as defined by the institution. This result suggested that a possible relationship between distance and attrition when grouping cases might exist.

The institution defined a primary market, secondary market, and a tertiary market. As stated before, the primary market comprises McKean County, the county in which the institution is located, and those counties that are contiguous to McKean County. The secondary market comprises those counties that are contiguous to the primary market. The tertiary market comprises all other areas (counties, states, and countries).

Results ($p = .022$) showed a positive correlation ($r_{pb} = .038$, $r_{pb}^2 = .0014$) between tertiary market designation (based on a student's location) and attrition. Results ($p = .009$) also showed a negative correlation ($r_{pb} = -.044$, $r_{pb}^2 = .0019$) between attrition and in-state status. Although this could be indicative of a relationship between attrition and distance when grouped, the relationship might also be the outcome of the difference in the cost of tuition between in-state and out-of-state residency. Exploration of this conjecture is further supported by a negative correlation ($p = .001$) ($r_{pb} = -.058$, $r_{pb}^2 = .0033$) between attrition and EFC. The results also showed a negative correlation ($p = .001$) ($r_{pb} = -.058$, $r_{pb}^2 = .0033$) between attrition and on-campus housing status. Finally, results showed a positive correlation ($p = .01$) ($r_{pb} = .043$, $r_{pb}^2 = .0018$) between attrition and first-generation college student status.

Statistically significant correlations resulted between all three measures of distance (DistanceT, DistanceD, and DistanceF) and the other predictor variables. Of course, correlations were expected between distance and markets (primary, secondary, and tertiary), residency status (in-state and out-of-state), and housing status (on-campus or off-campus) as those independent variables are closely associated with distance (i.e.,

the farther one's home address is from campus, the more likely they are to live on-campus and be considered from the secondary or tertiary market). However, there were statistically significant positive correlations for all three measures of distance and EFC ($p < .01$), SAT/ACT ($p < .01$), sex (DistanceT and DistanceF at $p < .01$ and DistanceD at $p < .05$), first-generation college student status (DistanceT at $p < .01$ and DistanceF and DistanceD at $p < .05$), and a positive correlation between all three measures of distance and minority student status ($p < .01$).

A negative correlation ($p < .01$) ($r_{pb} = -.121$, $r_{pb}^2 = .0146$) was observed between attrition and SAT/ACT score. However, the correlation matrix yielded results showing a statistically significant relationship between the independent variable of SAT/ACT score and all other independent variables that were included in the study. All correlations were significant at the ($p < .01$) except for the correlation between SAT score and secondary market designation ($p < .05$). However, the relationship did not suggest the possibility of multicollinearity as a limitation in this analysis or the binary logistic regression analysis.

Multicollinearity, a statistical phenomenon in which a perfect or exact relationship between the predictor variables exists, makes it difficult for the researcher to determine reliable estimates of individual predictor variable coefficients. However, none of the SAT/ACT correlation coefficients exceeded .7 and therefore multicollinearity is not believed to have been a factor (Hosmer & Lemeshow, 2000). Yet, the SAT/ACT score may be acting as a confounding variable. A confounding variable is an extraneous variable that correlates with both the dependent variable and the independent variable being analyzed and causes the observed relationship to be obscured or accentuated (MacKinnon, Krull, & Lockwood, 2000). When the confounding variable is statistically

controlled for or removed, and the result is an increase in the relationship between the dependent and independent variable, the effect of that variable is known as suppressing (or negative confounding).

There are several methods by which confounding variables can be controlled. Controlling confounding variables can be accomplished through the measurement of known confounders and treatment as covariates. A second method for controlling confounding variables that the researcher used was the partial correlation analysis. Partial correlation analysis, as indicated in Chapter 3, was conducted to explore the possibility of confounding variables. The results will follow later in this chapter.

Table 8

Point-Biserial Correlation Analyses

Cohort (Fall)	<i>n</i>	<i>r_{pb}</i>	<i>r_{pb}²</i>	<i>p</i>	Statistically Significant
DistanceD					
2005	208	.157	.0246	.013	Yes
2006	327	.070	.0049	.102	No
2007	285	-.085	.0072	.076	No
2008	340	.025	.0006	.323	No
2009	356	.066	.0044	.109	No
2010	307	.003	.0000	.481	No
2011	338	-.055	.0030	.156	No
2012	322	-.031	.0010	.292	No
2013	354	-.080	.0064	.067	No
Cumulative	2837	.001	.0000	.470	No
DistanceF					
2005	208	.155	.0240	.012	Yes
2006	327	.044	.0019	.216	No
2007	285	-.090	.0081	.065	No
2008	340	.024	.0006	.332	No
2009	356	.072	.0052	.087	No
2010	307	.007	.0000	.451	No
2011	338	-.054	.0029	.159	No
2012	322	-.027	.0007	.316	No
2013	354	-.080	.0064	.066	No
Cumulative	2837	.000	.0000	.495	No
DistanceT					
2005	208	.159	.0253	.011	Yes
2006	327	.037	.0014	.251	No
2007	285	-.090	.0081	.064	No
2008	340	.024	.0006	.331	No
2009	356	.072	.0052	.087	No
2010	307	.005	.0000	.465	No
2011	338	-.056	.0031	.154	No
2012	322	-.018	.0003	.372	No
2013	354	-.079	.0062	.069	No
Cumulative	2837	.001	.0000	.479	No

Table 8 displays the results of the point-biserial correlation analysis by year showing that distance in miles (both straight-line and mapped) and distance in travel time were associated with attrition at a statistically significant level in 2005 only. The point-biserial correlation for fall of 2005 shows a statistically significant relationship at the $p < .05$ level (one-tailed) between attrition and the distance (straight-line miles,

recommended route in miles, and in terms of travel time) from home to the institution. The relationship between attrition and distance was positive (i.e., the incidence of attrition increased as distance increased). The effect size (r_{pb}^2) was weak at .0253 for DistanceT, .0246 for DistanceD, and .0240 for DistanceF. Distance accounted for only 2.4% to 2.53% of the variability in attrition in 2005.

The significance of a p-value of 0.05 is attributed to Fisher (1933). Although arbitrary in nature, it is commonly accepted and referred to (along with a p-value $< .01$) in statistical literature. However, for the purposes of exploring the results of the point-biserial correlation analyses in this case study, a p-value $< .10$ was also considered. Using the .10 level, one would assume that there is less than 1 in 10 odds that an event of interest occurred by chance alone. The point-biserial correlation for fall of 2007 and 2013 would show a statistically significant relationship at the $p < .10$ level (one-tailed) between attrition and the distance (straight-line miles, recommended route in miles, and in terms of travel time) from home to the institution. In 2009, a statistically significant relationship would be shown at the $p < .10$ level (one-tailed) between attrition and the distance (recommended route in miles, and in terms of travel time, excluding straight-line distance). However, the relationship in 2007 and 2013 between attrition and distance is negative or inverse (i.e., the incidence of attrition decreased as distance increased). The effect sizes (r_{pb}^2) in 2007 and 2013 were weak at .0064 and .0069 for DistanceT, .0072 and .0064 for DistanceD, and .0081 and .0062 for DistanceF.

For the purposes of exploring the data, and to better understand whether any relationship between distance and attrition might be positive or negative, the researcher then analyzed the data using several binary variables (dummy variables) that were created

artificially by grouping cases. Lani (2010) advised that researchers should avoid the creation of binary variables from ratio data because the original ratio data contains more variance information making correlation analysis more reliable. However, several other researchers including Huitema (2011) and Sheskin (2011) suggested that although the resulting dummy variables may be thought of as having an underlying continuity, grouping cases can provide some insight into a correlational relationship.

The researcher explored the data by artificially grouping cases based on DistanceT, DistanceD, and DistanceF based on several intervals of miles for DistanceD and DistanceF (i.e., 30, 60, 90, and 120-mile intervals) resulting in no improvement in the outcome of the correlation. He grouped DistanceT based on several intervals of minutes (i.e., 30, 60, 90, and 120-minute intervals). The cumulative results based on 30-minute intervals are shown in Table 9.

Table 9

Point-Biserial Correlation Analyses of DistanceT in 30-Minute Intervals (Cumulative)

Cohort (Fall)	<i>n</i>	<i>r_{pb}</i>	<i>r_{pb}²</i>	<i>p</i>	Statistically Significant
DistanceT					
0-30	426	.036	.0013	.029	Yes
31-60	479	-.015	.0002	.216	No
61-90	253	-.031	.0010	.048	Yes
91-120	217	.000	.0000	.497	No
121-150	153	.013	.0002	.238	No
151-180	163	-.040	.0016	.017	Yes
181-210	274	-.025	.0006	.087	No
211-240	118	.043	.0018	.010	Yes
241-270	133	.007	.0000	.362	No
271-300	114	.006	.0000	.384	No
301-330	206	-.014	.0002	.228	No
331-360	195	.026	.0007	.082	No
361-390	27	.008	.0001	.329	No
>390	79	.003	.0000	.428	No
Cumulative	2837	.001	.0000	.479	No

As shown in Table 9, the distance measured in travel time (minutes) was associated with attrition at a statistically significant level at intervals of 0-30, 61-90, 151-180, and 211-240 minutes. The point-biserial correlation for those intervals shows a statistically significant relationship at the $p < .05$ level (one-tailed) between attrition and the distance (in terms of travel time) from home to the institution. The relationship between attrition and distance was positive (i.e., the incidence of attrition increased as distance increased) for intervals 0-30 and 211-240 minutes. However, the relationship between attrition and distance was negative (i.e., the incidence of attrition decreased as distance increased) for intervals 61-90 and 151-180 minutes. The effect size (r_{pb}^2) was weak at .0013 for 0-30 minutes, .0010 for 61-90 miles, .0016 for 151-180 miles, and .0018 for 211-240 miles. The results suggested that a non-linear relationship, albeit weak, may exist.

Point-Biserial Correlation Summary

The null hypothesis for the central research question was stated as H_0 : There is not a statistically significant relationship between the institutional distance from home and first-year attrition for traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford. No statistically significant relationship was found between Attrition and DistanceT ($r_{pb} = .001$, $p = .479$), DistanceD ($r_{pb} = .001$, $p = .470$), or DistanceF ($r_{pb} = .000$, $p = .495$). The result is that we fail to reject the null hypothesis.

Further analysis showed that a statistically significant, but weak relationship was observed between attrition and the tertiary market, as defined by the institution ($r_{pb} = .038$, $p = .022$). Analysis by year resulted in a statistically significant, but weak

relationship observed in cohort year 2005 only: DistanceD ($r_{pb} = .157$, $p = .013$), DistanceF ($r_{pb} = .155$, $p = .012$), and DistanceT ($r_{pb} = .159$, $p = .011$).

Artificially grouping cases based on DistanceT in 30-minute intervals resulted in a statistically significant positive relationship at the $p < .05$ level for intervals 0-30 ($r_{pb} = .036$) and 211-240 minutes ($r_{pb} = .036$). However, the relationship between attrition and distance was negative for intervals 61-90 ($r_{pb} = -.031$) and 151-180 ($r_{pb} = -.040$) minutes. The results suggested that a non-linear relationship, albeit weak, may exist.

A statistically significant relationship was observed between SAT/ACT score and all other independent variables included in the study at the $p < .01$ level (except for secondary market designation, which was at the $p < .05$ level). This result led the researcher to conduct a partial correlation analysis to determine if the SAT/ACT score was acting as a confounding variable.

Partial Correlation

As described in Chapter 3, partial correlations were conducted to control for the possibility that the SAT/ACT score was serving as a confounding or suppressing variable as well as to explore the potential relationship between DistanceD, DistanceF, and DistanceT and other potential predictor variables that were described in the literature as having been found to have a statistically significant relationship with attrition (e.g., expected family contribution, residency status, housing status, etc.). If the relationship between the two variables of interest decreased after eliminating the effects of the third variable, then it is possible to conclude that the relationship between the variables of interest is apparently but not actually valid (Bay & Hakstian, 1972).

Partial correlations were conducted, and the results are listed in Table 10. Each independent variable was treated as a covariate and tested against the potential relationship between attrition and DistanceT, DistanceD, and DistanceF respectively.

First, when controlling for the effects of EFC, SAT/ACT score, first-generation college student status, housing status, and race and ethnicity on the relationship between attrition and the distance from home to the institution as measured in minutes (DistanceT), the researcher found that the correlation changed from $r = .001$ $p = .479$ to $r = .038$ $p = .022$. When controlling for the effects of EFC, SAT/ACT score, first-generation college student status, housing status, and race and ethnicity on the relationship between attrition and the distance from home to the institution as measured in straight-line-miles (DistanceD), the researcher found that the correlation changed from $r = .001$ $p = .470$ to $r = .033$ $p = .040$. Finally, when controlling for the effects of EFC, SAT/ACT score, first-generation college student status, housing status, and race and ethnicity on the relationship between attrition and the distance from home to the institution as measured in miles based on the recommended route (DistanceF), the researcher found that the correlation changed from $r = .000$ $p = .495$ to $r = .034$ $p = .037$.

Although the results were statistically significant at $p < .05$, and increases in the r values were observed, the effect size (r^2) for each relationship was very weak at .0014 for DistanceT, .0011 for DistanceD, and .0012 for DistanceF.

Table 10

Partial Correlation Analyses of Distance x Attrition

Correlated variables	Controlled variable	<i>r</i>	<i>r</i> ²	<i>p</i>
Attrition x DistanceT		.001	.000001	.479
Attrition x DistanceT	EFC	.008	.00006	.341
Attrition x DistanceT	SAT	.018	.00032	.163
Attrition x DistanceT	IS-RES	-.012	.00014	.262
Attrition x DistanceT	PRIM	-.013	.00017	.251
Attrition x DistanceT	SEC	.000	.00000	.500
Attrition x DistanceT	TERT	-.022	.00048	.122
Attrition x DistanceT	Male	.001	.000001	.483
Attrition x DistanceT	First Gen	.003	.00001	.434
Attrition x DistanceT	Campus	-.018	.00032	.172
Attrition x DistanceT	White-NH	.003	.00001	.442
Attrition x DistanceT	EFC, SAT, First Gen, Campus, White	.038	.00144	.022
Attrition x DistanceD		.001	.000001	.470
Attrition x DistanceD	EFC	.007	.00005	.351
Attrition x DistanceD	SAT	.018	.00032	.171
Attrition x DistanceD	IS-RES	-.012	.00014	.259
Attrition x DistanceD	PRIM	-.009	.00008	.313
Attrition x DistanceD	SEC	.000	.00000	.491
Attrition x DistanceD	TERT	-.017	.00029	.183
Attrition x DistanceD	Male	.001	.000001	.473
Attrition x DistanceD	First Gen	.003	.00001	.431
Attrition x DistanceD	Campus	-.014	.0002	.221
Attrition x DistanceD	White-NH	.003	.00001	.438
Attrition x DistanceD	EFC, SAT, First Gen, Campus, White	.033	.00109	.040
Attrition x DistanceF		.000	.00000	.495
Attrition x DistanceF	EFC	.006	.00004	.368
Attrition x DistanceF	SAT	.017	.00029	.185
Attrition x DistanceF	IS-RES	-.013	.00017	.247
Attrition x DistanceF	PRIM	-.012	.00014	.265
Attrition x DistanceF	SEC	-.001	.000001	.483
Attrition x DistanceF	TERT	-.021	.00044	.134
Attrition x DistanceF	Male	.000	.00000	.498
Attrition x DistanceF	First Gen	.002	.000004	.456
Attrition x DistanceF	Campus	-.015	.00023	.215
Attrition x DistanceF	White-NH	.002	.000004	.460
Attrition x DistanceF	EFC, SAT, First Gen, Campus, White	.034	.00116	.037

Partial Correlation Summary

As was previously stated, initial correlational analysis led to the conclusion that we fail to reject the null hypothesis. However, controlling for confounding (or suppressing variables) resulted in a statistically significant, albeit extremely weak observed relationship.

Partial correlation was used to evaluate the null hypothesis that there is not a statistically significant relationship between the institutional distance from home and first-year attrition for traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford while controlling for sex, race and ethnicity, expected family contribution (EFC), parental education level, residency status, housing status, SAT and ACT score, and market segmentation ($n = 2837$). There was significant evidence to reject the null hypothesis and conclude that there was a weak but statistically significant positive, partial correlation between distance as measured in travel time (DistanceT) and attrition while controlling for race and ethnicity, expected family contribution (EFC), parental education level, housing status, and SAT and ACT score $r(2837) = .038, p < .05$. Results of the point-biserial correlation yielded a non-significant weak positive correlation between distance as measured in travel time (DistanceT) and attrition $r_{pb}(2837) = .001, p = .479$ indicating that controlling for race and ethnicity, expected family contribution (EFC), parental education level, housing status, and SAT and ACT score influenced the strength of the relationship between attrition and distance measured in travel time.

There was a weak but statistically significant positive, partial correlation between orthodromic distance as measured in miles (DistanceD) and attrition while controlling for

race and ethnicity, expected family contribution (EFC), parental education level, housing status, and SAT and ACT score $r(2837) = .033, p < .05$. Results of the point-biserial correlation yielded a non-significant weak positive correlation between distance as measured in travel time (DistanceT) and attrition $r_{pb}(2837) = .001, p = .479$ indicating that controlling for race and ethnicity, expected family contribution (EFC), parental education level, housing status, and SAT and ACT score influenced the strength of the relationship between attrition and orthodromic distance as measured in miles.

There was a weak but statistically significant positive, partial correlation between the recommended route distance, as measured in miles (DistanceF), and attrition while controlling for race and ethnicity, expected family contribution (EFC), parental education level, housing status, and SAT and ACT score $r(2837) = .034, p < .05$. Results of the point-biserial correlation yielded a non-significant weak positive correlation between distance, as measured in travel time (DistanceF), and attrition $r_{pb}(2837) = .000, p = .495$ indicating that controlling for race and ethnicity, expected family contribution (EFC), parental education level, housing status, and SAT and ACT score influenced the strength of the relationship between attrition and distance, as measured in miles, based on the recommended route.

Binary Logistic Regression

As was originally proposed, a statistically significant relationship would be further examined conducting additional analyses while attempting to control for other variables that, according to a review of the literature, have been shown to influence attrition and retention. No statistically significant relationship was found following the analysis of the data using the point-biserial correlation. However, results from a follow-

up partial correlation suggested a statistically significant, albeit extremely weak, observed relationship. Furthermore, the point-biserial correlation and the partial correlation assume a linear relationship; any non-linear relationship would not be detected using either method. The researcher chose to continue analysis using the binary logistic regression.

The alpha level was set at .05. The alpha level is the value for the probability of a type-I error (i.e., the researcher rejects the null hypothesis when it is, in fact, true—an effect that is not present is detected). A 0.05 alpha level means that the test will falsely reject the null hypothesis once in every twenty tests.

The model that was tested included the variables of interest in this study: distance from home (as measured in travel time, straight-line miles, or recommended route), sex, race or ethnicity, expected family contribution (EFC), parental education levels (specifically, first-generation college student status), residency status, housing status, SAT or ACT score, and market segmentation as defined by the institution.

Multicollinearity is a statistical phenomenon in which a perfect or exact relationship between the predictor variables exist making it difficult for the researcher to determine reliable estimates of individual predictor variable coefficients. None of the correlation coefficients exceeded .7 indicating multicollinearity was not a factor (Hosmer & Lemeshow, 2000). However, a certain degree of multicollinearity was assumed between the independent variables measuring distance (in travel time and mileage). Therefore multicollinearity was tested by a second method: Tolerance and the Variance Inflation Factor (VIF) (Pallant, 2007).

Tolerance indicates the degree of variability of the independent variable that is not explained by the other independent variables in the model. The tolerance measure should

be greater than .10. Otherwise, the result would indicate that the multiple correlations with other variables are high and would suggest the possibility of collinearity. Tolerance “is calculated using the formula $1-r^2$ for each variable” (Pallant, p. 150). The Variance Inflation Factor (VIF) is the inverse of the Tolerance Factor. The inverse is calculated as 1 divided by Tolerance. A VIF value above 10 indicates multicollinearity. The results of both the Tolerance and the VIF confirmed the likelihood of multicollinearity between DistanceT, DistanceD, and DistanceF as shown in Table 11. As a result, separate binary logistic regression models were built incorporating DistanceT, DistanceD, and DistanceF respectively.

Table 11

Tolerance and Variance Inflation Factor

Predictor	<i>Tolerance</i>	<i>VIF</i>
DistanceT (time)	.004	266.114
DistanceD (straight-line)	.039	25.945
DistanceF (recommended route)	.003	286.259
EFC	.890	1.123
SAT/ACT	.901	1.109
In-State Residency	.753	1.328
Primary Market	.299	3.347
Secondary Market	.588	1.700
Tertiary Market	.513	1.948
Male	.964	1.037
First-Generation Status	.930	1.075
On-campus Housing	.693	1.442
White-Non-Hispanic	.805	1.234

Overall Evaluation of Models: The Maximum Likelihood model is a prediction of outcomes if only the independent variable of attrition is included in the model. In this data set, there were 836 occurrences of attrition and 2,001 occurrences of non-attrition (retention). The initial base model always assumes a prediction based on the larger

number of observations. Therefore, it correctly predicts that attrition will not occur and does so 2,001 times. Conversely, it incorrectly predicts attrition all 836 times. Therefore, the base model (no predictors) had an overall accuracy of 70.5% (2,001/2,837). After the independent variables had been added, the resulting models each had a lower overall accuracy of 70.4%. For the model using DistanceT, there were 7 accurate predictions of attrition as well as 1,989 accurate predictions of non-attrition. There were 12 incorrect predictions of attrition and 829 incorrect predictions of non-attrition. For the model using DistanceD, there were 8 accurate predictions of attrition as well as 1,989 accurate predictions of non-attrition. There were 12 incorrect predictions of attrition and 828 incorrect predictions of non-attrition. For the model using DistanceF, there were 8 accurate predictions of attrition as well as 1,988 accurate predictions of non-attrition. There were 13 incorrect predictions of attrition and 828 incorrect predictions of non-attrition.

As stated above, the base model (no predictors) had an overall accuracy of 70.5% (2,001/2,837). Forward Selection (Wald) was also tested. This stepwise selection method adds predictor variables based on the significance of the score statistic and removes them based on the probability of the Wald statistic. Using Forward Selection (Wald), the resulting models each had a lower overall accuracy of 70.4% and the measurements of distance incorporated in each respective model (DistanceT, DistanceD, and DistanceF) were identified as non-statistically significant predictors and were removed.

Goodness-Of-Fit Statistics: Each binary logistic regression model was evaluated and Hosmer-Lemeshow (H-L) Test values were calculated. Values of $p > 0.05$ would indicate that we fail to reject the null hypothesis that the model is not significantly different from

observed values. The resulting value for the first Hosmer-Lemeshow (H-L) Test incorporating DistanceT into the model was $p = .958$ with eight degrees of freedom. Therefore, we reject the null hypothesis. The resulting value for the second Hosmer-Lemeshow (H-L) Test incorporating DistanceD into the model was $p = .938$ with eight degrees of freedom. Therefore, we reject the null hypothesis. The resulting value for the final Hosmer-Lemeshow (H-L) Test incorporating DistanceD into the model was $p = .812$ with eight degrees of freedom. Therefore, we reject the null hypothesis.

A true R^2 value does not exist in logistic regression as it does in ordinary least squares regression analysis. However, there are measures intended to mimic the R-squared analysis. These pseudo R-squared values can be interpreted as an approximate variance in the outcome accounted for by the independent variables. SPSS provides output for two versions of pseudo R-squared values: the Cox & Snell R-Square and the Nagelkerke R-Square. Both take the approach of a measurement of the improvement from the null model to the proposed model and are interpreted as the smaller the ratio, the greater the improvement. The Nagelkerke R-Square is a modified version of the Cox & Snell R-Square that has been adjusted so as to extend the range of possible values to 1 (Hosmer & Lemeshow, 2000; O'Connell, 2006).

The resulting values for each of the respective models (i.e., incorporating DistanceT, DistanceD, and DistanceF respectively) were .031 for the Cox and Snell R-Square and .045 for the Nagelkerke R-Square. These values were close to 0 and indicated that the model is not a notable improvement over the null model, which contained no predictors. Therefore, the explained variation in the dependent variable

based on our models ranged from 3.1% to 4.5%, depending on whether the Cox & Snell R^2 or the Nagelkerke R^2 method were referenced.

The Chi-Square Goodness of Fit: The resulting Chi-square values implied that all three models are predictive of attrition: (*chi-square* = 90.363, $p < .001$ with $df = 10$) for the model that incorporated DistanceT, (*chi-square* = 90.333, $p < .001$ with $df = 10$) for the model that incorporated DistanceD, and (*chi-square* = 90.442, $p < .001$ with $df = 10$) for the model that incorporated DistanceF. These results allowed the researcher to reject the null hypothesis that the variables included in the models utilizing all predictor variables were not predictive of attrition.

Kendall's Tau-b, Goodman-Kruskal's Gamma, Somer's D: If Kendall's Tau-b, Goodman-Kruskal's Gamma, and Somer's D values > 0 , then the degree, measured as a percentage, to which the model results in fewer false predictions than chance, can be reported. For DistanceT, Kendall's Tau-b was .001, Goodman-Kruskal's Gamma was .002, and Somer's D .001 with $p = .933$. For DistanceD, Kendall's Tau-b was .002, Goodman-Kruskal's Gamma was .003, and Somer's D .002 with $p = .891$. For DistanceF, Kendall's Tau-b was .000, Goodman-Kruskal's Gamma was .000, and Somer's D .000 with $p = .984$.

Receiver Operating Characteristic (ROC) Curve: The ROC curve shows the tradeoff between sensitivity and specificity. As the cutoff or threshold changes, the increase in sensitivity will be accompanied by a decrease in specificity. The Area Under a Curve (AUC) (i.e., c statistic), is an additional overall measure of the model performance and is used to report the validation of the predicted probabilities. An AUC = 1 is a perfect model. An AUC = 0.50 is a model that is no better than chance. The predictive validity of

DistanceT demonstrated nearly no improvement over chance (AUC= .501, $p = .932$). The predictive validity of DistanceD also demonstrated nearly no improvement over chance (Area Under the Curve = .502, $p = .882$). Finally, the predictive validity of DistanceF demonstrated no improvement over chance (AUC = .500, $p = .984$).

Odds Ratios: The odds ratio (i.e., the SPSS output of $\text{Exp}(B)$) can be interpreted either as a relative measure of effect allowing for the comparison of the intervention group of a study relative to the comparison group, or as the multiplicative adjustment to the odds of the outcome given a unit of change in the independent variable. An odds ratio of one indicates that there is little to no difference between the intervention group and the comparison group, or that the multiplicative adjustment is very small compared to the size of a meaningful change. The resulting odds ratios for DistanceT, DistanceD, and DistanceF were all displayed as 1.000 in SPSS. The actual values were .999907 for DistanceT, .999915 for DistanceD, and .999901 for DistanceF. An odds ratio of 1 corresponds to an independent variable that does not affect the dependent variable: distance in travel time or miles is not associated with attrition. As would be expected, conversion of DistanceT (time) from minutes to miles affected the resulting coefficient and $\text{Exp}(B)$ or odds ratio ($B=-.006$, $\text{OR}=.994$).

Table 12

Binary Logistic Regression Results Incorporating DistanceT

Predictor	<i>B</i>	<i>SE</i>	Wald X^2	<i>df</i>	Sig.	OR	Lower	Upper
DistanceT (time)	.000	.000	.151	1	.697	1.000	.999	1.000
EFC	.000	.000	6.315	1	.012	1.000	1.000	1.000
SAT/ACT	-.002	.000	41.606	1	.000	.998	.997	.999
In-State Residency	-.314	.115	7.389	1	.007	.731	.583	.916
Primary Market	-.592	.139	18.063	1	.000	.553	.421	.727
Secondary Market	-.252	.116	4.706	1	.030	.777	.619	.976
Tertiary Market	.252	.116	4.706	1	.030	1.287	1.025	1.615
Male	.098	.085	1.337	1	.248	1.104	.934	1.304
First-Generation Status	.108	.087	1.526	1	.217	1.114	.939	1.322
On-campus Housing	-.541	.134	16.314	1	.000	.582	.448	.757
White-Non-Hispanic	.278	.113	6.061	1	.014	1.321	1.058	1.648
Constant	1.925	.364	28.014	1	.000	6.856		

Note: *B* = beta weight, *SE* = Standard error, *df* = degrees of freedom; Sig. = Significance; OR= Odds Ratio

Table 13

Binary Logistic Regression Results Incorporating DistanceD

Predictor	<i>B</i>	<i>SE</i>	Wald X^2	<i>df</i>	Sig.	OR	Lower	Upper
DistanceD (straight-line)	.000	.000	.119	1	.730	1.000	.999	1.000
EFC	.000	.000	6.332	1	.012	1.000	1.000	1.000
SAT/ACT	-.002	.000	41.675	1	.000	.998	.997	.999
In-State Residency	-.316	.116	7.379	1	.007	.729	.581	.916
Primary Market	-.585	.135	18.837	1	.000	.557	.428	.726
Secondary Market	-.248	.114	4.727	1	.030	.781	.624	.976
Tertiary Market	.248	.114	4.727	1	.030	1.281	1.025	1.602
Male	.098	.085	1.333	1	.248	1.103	.934	1.304
First-Generation Status	.108	.087	1.523	1	.217	1.114	.939	1.322
On-campus Housing	-.543	.134	16.447	1	.000	.581	.447	.756
White-Non-Hispanic	.280	.113	6.169	1	.013	1.323	1.061	1.650
Constant	1.918	.362	28.086	1	.000	6.810		

Note: *B* = beta weight, *SE* = Standard error, *df* = degrees of freedom; Sig. = Significance; OR= Odds Ratio

Table 14

Binary Logistic Regression Results Incorporating DistanceF

Predictor	<i>B</i>	<i>SE</i>	Wald X^2	<i>df</i>	Sig.	OR	Lower	Upper
DistanceF (recommended route)	.000	.000	.228	1	.633	1.000	.999	1.000
EFC	.000	.000	6.316	1	.012	1.000	1.000	1.000
SAT/ACT	-.002	.000	41.567	1	.000	.998	.997	.999
In-State Residency	-.316	.116	7.460	1	.006	.729	.581	.915
Primary Market	-.594	.137	18.730	1	.000	.552	.422	.723
Secondary Market	-.254	.115	4.854	1	.028	.776	.619	.972
Tertiary Market	.254	.115	4.854	1	.028	1.289	1.028	1.617
Male	.099	.085	1.340	1	.247	1.104	.934	1.304
First-Generation Status	.108	.087	1.533	1	.216	1.114	.939	1.323
On-campus Housing	-.541	.134	16.359	1	.000	.582	.448	.757
White-Non-Hispanic	.277	.113	6.042	1	.014	1.320	1.058	1.647
Constant	1.927	.363	28.217	1	.000	6.871		

Note: *B* = beta weight, *SE* = Standard error, *df* = degrees of freedom; Sig. = Significance; OR = Odds Ratio

Statistical Tests of Individual Predictors: Significant Wald Chi-Square Statistic values

at $p < 0.05$ indicate that the predictor is significant, and the null hypothesis can be rejected. As shown in Tables 12, 13, and 14, in all three models, the predictor variables EFC, SAT/ACT, In-State Residency, Primary Market, Secondary Market, Tertiary Market, On-campus Housing, and White-Non-Hispanic, had values that were significant at $p < 0.05$. The independent variables, SAT/ACT, In-State Residency, Primary Market, and On-campus Housing had values that were significant at $p < 0.01$. For each of these independent variables, the null hypothesis that the coefficient equals 0 is rejected. However, the null hypothesis (that the coefficient equals 0) is accepted for each of the remaining independent variables: DistanceT, DistanceD, DistanceF, EFC, Sex-Male, and First-Generation Status.

Binary Logistic Regression Summary

Mixed results from the point-biserial correlation and partial correlation led the researcher to explore the data further for the possible existence of a relationship between distance and attrition. A binary logistic regression analysis was conducted to examine the potential relationship further between distance and attrition and to determine the effect, if any, on the relationship when controlling for other independent variables. A test of each of the full models (utilizing DistanceT, DistanceD, and DistanceF, respectively) against the constant only models was statistically significant, indicating that the predictors reliably distinguished between cases of attrition and retention. The Chi-square values were 90.363 for the model incorporating DistanceT, 90.333 for the model incorporating DistanceD, and 90.442 for the model incorporating DistanceF ($p < .001$ with $df = 10$).

Nagelkerke's R^2 values of .045 for all three models indicated that the model is not a notable improvement over the null model containing no predictors. Prediction success overall was 70.4% (99.3% for retention and 1.2% for attrition). The model's predictive success was lower than the base model (no predictors), which had an overall accuracy of 70.5%. Therefore, the predictive model incorporating all variables was less accurate than the base model. The Wald criterion indicated that the predictor variables SAT/ACT, In-State Residency, Primary Market, Secondary Market, Tertiary Market, On-campus Housing, and White-Non-Hispanic had values that were significant at $p < 0.05$. SAT/ACT, In-State Residency, Primary Market, and On-campus Housing had values that were significant at $p < 0.01$. However, distance, as measured in travel time (DistanceT) or miles (DistanceD and DistanceF), was not a significant predictor.

Exp(B) or odds ratio values for DistanceT, DistanceD, and DistanceF were .999907 for DistanceT, .999915 for DistanceD, and .999901 for DistanceF. An odds ratio of 1 indicates an independent variable that does not affect the dependent variable. The results indicated that distance in travel time or miles is not associated with attrition.

Results of the Research Questions

The primary reason for this case study was to determine if a relationship existed between attrition and the institutional distance from students' permanent home addresses for first-time, full-time, traditional-aged students at the University of Pittsburgh at Bradford. The following research question was developed to address this question:

R1: Is there a statistically significant relationship between the institutional distance from home and first-year attrition for traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

Following the correlational analysis, further analysis was conducted to determine if the relationship was non-linear and if it existed between the institutional distance from home and attrition of first-year students prior to their second year when controlling for demographic, socioeconomic, and academic variables that have been routinely analyzed as possible factors and predictors of attrition and retention as was outlined in the literature review. To address this question, the following research sub-questions were developed:

R1a: Will controlling for sex affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

R1b: Will controlling for race or ethnicity affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

R1c: Will controlling for expected family contribution (EFC) affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

R1d: Will controlling for parental education levels (specifically, first-generation college student status) affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

R1e: Will controlling for residency status affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

R1f: Will controlling for housing status affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

R1g: Will controlling for the SAT composite score affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

R1h: Will controlling for market segmentation as defined by the institution affect the relationship, if any, between the institutional distance from home and attrition of traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford?

Results of the point-biserial correlation showed no statistically significant relationship between attrition and DistanceT ($r_{pb} = .001, p = .479$), DistanceD ($r_{pb} = .001, p = .470$), and DistanceF ($r_{pb} = .000, p = .495$). The result is that we fail to reject the null hypothesis.

However, results from the partial correlation showed a weak, but statistically significant relationship between distance and attrition when controlling for the combination of race and ethnicity, expected family contribution (EFC), parental education level, housing status, and SAT and ACT score $r(2837) = .034, p < .05$. The corresponding r and r^2 values were small, indicating that the effect of distance on attrition was weak, and only a fraction of the percentage of variance could be attributed to distance.

A binary logistic regression analysis was then conducted to examine further the potential relationship between distance and attrition and to determine the effect, if any, on the relationship when controlling for other independent variables. A test of each of the full models (utilizing DistanceT, DistanceD, and DistanceF, respectively) against the constant only models was statistically significant, indicating that the predictors as a set reliably distinguished between cases of attrition and retention. The Chi-square values were 90.363 for the model incorporating DistanceT, 90.333 for the model incorporating DistanceD, and 90.442 for the model incorporating DistanceF ($p < .001$ with $df = 10$). No significant relationship between distance and attrition was observed even when controlling for all other independent variables. Therefore, the answer to the research question is that a statistically significant relationship between the institutional distance from home and first-year attrition for traditional-aged, first-time, full-time students at the

University of Pittsburgh at Bradford does not exist. We fail to reject the null hypothesis for the central research question and all research sub-questions.

Chapter 5

Discussion of Findings and Implications

Introduction

Institutions face increasing demands to maintain or increase enrollments. As a result, colleges and universities have begun to recruit students from greater distances. Anecdotal evidence including exit interviews conducted with attriting students, reviews of current methodologies employed by predictive modeling and econometrics consultants, discussions with enrollment management professionals, first-year experience design and programming, and homesickness studies all suggested that distance plays a critical role in the decision-making process affecting student attrition and retention. This anecdotal evidence led the researcher to postulate that students at greater distances were more prone to attrition before their second year of study. If this conjecture were true, it would have significant implications for rural or geographically isolated institutions in terms of recruitment and retention strategies. Institutions might consider modifying their levels of admissions staffing, adjusting geographic territories and travel plans, modifying marketing plans, enhancing communications, and could change offers of merit and need-based, institutional aid based on location. Student-success services, first-year experience planning, and student affairs and activities programming might be changed to better address the needs of students who come from greater distances.

The purpose of this *ex-post facto* case study was to determine if a relationship existed between the institutional distance from home and the attrition of traditional-aged, first-time, full-time students prior to the second year at the University of Pittsburgh at Bradford, one of the four-year campuses of the University of Pittsburgh. Following the

correlational analysis, further analysis was conducted to determine if the relationship was non-linear and if it existed between the institutional distance from home and attrition of first-year students prior to their second year when controlling for sex, race or ethnicity, expected family contribution (EFC), parental education levels (specifically, first-generation college student status), residency status, housing status, SAT or ACT score, or market segmentation as defined by the institution. The literature shows that prior exploration of this relationship has been sporadic, and the methodology has been incomplete.

The central research question that this case study aimed to answer is, does a relationship exist between the institutional distance from home and the attrition of traditional-aged, first-time, full-time students prior to the second year at the University of Pittsburgh at Bradford?

Discussion of Results

The main findings of this study were non-significant; I failed to reject the null hypothesis and all sub-null hypotheses. Non-significant results do not imply an inferior study. Rudestam and Newton (1992) stated that nonsignificant results may stem from methodological or theoretical shortcomings. The researcher believes that the methodology was sound, both in terms of statistics used, but even more so, in terms of how distance was calculated. Recent strides in technology and software enabled the researcher to calculate each respective case's distance from home to the institution in miles based on the great-circle or orthodromic distance, and the distance and travel time of the recommended route from home to the institution based on the Google mapping APIs. The empirical research conducted in this study followed the scientific method and

sought to test nullifiable hypotheses based on quantifiable evidence in a systematic manner (Goodwin, 2009). The study was designed to verify whether a justified hypothesis could be backed up by evidence. However, in each case, the researcher failed to reject the null hypothesis: There is not a statistically significant relationship between the institutional distance from home and first-year attrition for traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford.

Research cited in the conceptual framework and literature review sections of this study had a general purpose: to describe the patterns of student retention and attrition. The majority of these studies had a secondary purpose: to develop predictive models that could lead an institution to implement specific intervention strategies that would mitigate those detrimental factors correlated with attrition or to enhance those factors found to enhance retention. Predictive modeling involves the creation of statistical models of future behavior based on known data and often involves choosing a model based on detection theory in which one tries to determine the probability of an outcome given specific input data. Research devoted to determining the reasons for student persistence and retention or determining causes of attrition has informed the development of predictive models. If students at risk for attrition are identified prior to attrition taking place, then remediation and services designed to increase student success can be applied. Conversely, institutions might use the results of predictive modeling to modify the profile of students who are recruited (i.e., to recruit students not exhibiting characteristics associated with attrition).

Pre-matriculation data (i.e., data that may be known about a student prior to enrolling in an institution) has been studied and incorporated into models developed by

Spady, 1970; Tinto, 1975; Bean, 1980; Astin, 1984; Pascarella, 1985; Cabrera, Nora, and Castaneda, 1992 and others. Pre-matriculation data includes prior academic achievement data, socioeconomic status, sex, age, race/ethnicity, parents and family data, and student commitment to a degree (Crissman Ishler & Upcraft, 2005). This study sought to explore the possibility that a relationship existed between the institutional distance and attrition. That relationship was tested while controlling for several variables. The variables used in this study that are considered pre-matriculation data have been routinely tested in previous attrition and retention studies and include sex, race and ethnicity, expected family contribution (EFC), parental education level, residency status, and SAT and ACT score.

The variable of market segmentation was defined by the institution, and although it, along with DistanceD, DistanceF, and DistanceT, is also considered pre-matriculation data, the literature shows that prior utilization of these kinds of variables describing distance has been sporadic, and the methodology has been incomplete. The variable, housing status (living on-campus or commuting), is considered a post-matriculation environmental variable, and has been incorporated into numerous studies and models.

The results of the binary logistic regression analyses showed significant statistical relationships between some pre-matriculation variables as well as the environmental variable indicating housing status.

SAT/ACT score. In each model, for each point increase on SAT/ACT score, the odds of attrition decreased from one to 0.998: DistanceT $b = -.002$, Wald $X^2(1) = 41.606$, $p < .01$, DistanceD $b = -.002$, Wald $X^2(1) = 41.675$, $p < .01$, and DistanceF $b = -.002$, Wald $X^2(1) = 41.567$, $p < .01$.

Research suggests that the most powerful predictor of student retention from the first year to the second year is the pre-matriculation or prior academic achievement data including high school grades and SAT or ACT scores (Hossler, 2000; Astin, 1993; Pascarella & Terenzini, 1991). Astin (1993) stated, “Hundreds of studies using various measurements and methodologies have yielded similar results: college grade point averages can be predicted with modest accuracy (multiple correlation around .55) from admissions information. The two most-potent predictors are the student’s high school grade point average and scores on college admissions tests” (p. 187). Although Schwartz and Washington (1999) concluded that standardized tests have less predictive power than the use of high school grades or rank, they also noted that standardized results may be biased toward or against sex and ethnic groups. However, researchers may have to continue to use standardized test scores.

The National Association for College Admission Counseling (2014) reported that more than half of all high schools no longer report high school rank. Many elite colleges and universities select students from only the top 10% of their respective classes. High schools are concerned that reporting class rank creates a disadvantage for good students who fall below the 10% category. Furthermore, an increasing number of high schools are reporting weighted grades and weighted averages, making the use of this data in predictive comparisons less reliable.

The SAT/ACT score was positively correlated with DistanceT ($r_{pb} = .142, p < .01$), DistanceD ($r_{pb} = .133, p < .01$), and DistanceF ($r_{pb} = .135, p < .01$) suggesting that SAT scores increased with the institutional distance from home. This relationship can be attributed to a specific university-wide recruitment program. Students who apply to the

Pittsburgh campus of the University of Pittsburgh and are subsequently rejected can have their applications referred to one of the four regional campuses of the University. The SAT/ACT scores of the rejected applicants tend to be 75 to 100 points higher than the scores of students who apply directly to one of the regional campuses. At UPB, approximately 25% of the incoming first-time, full-time cohort each year are students who had applied and been rejected by the Pittsburgh campus. The researcher considered the possibility that this referral status might affect the outcomes of the models developed through the binary logistic regression analyses. Therefore, the researcher subsequently added the variable of referral status to each of the three models (incorporating DistanceT, DistanceD, and DistanceF). There were no resulting changes. The predictability of each model (70.4%) remained below the Maximum Likelihood model. There were no improvements in the predictive validity of DistanceT, DistanceD, or DistanceF; the results demonstrated nearly no improvement over chance. Wald X^2 values continued to be nonsignificant with $p > .650$.

Expected Family Contribution (EFC). In each model, for each dollar increase in EFC, the odds of attrition decreased from one to 0.99999: DistanceT $b = -.00001$, Wald $X^2(1) = 6.315$, $p < .05$, DistanceD $b = -.00001$, Wald $X^2(1) = 6.332$, $p < .05$, and DistanceF $b = -.00001$, Wald $X^2(1) = 6.316$, $p < .05$. In each model, for each 10,000 dollar increase in EFC, the odds of attrition decreased from one to 0.90: DistanceT $b = -.1$, Wald $X^2(1) = 6.315$, $p < .05$, DistanceD $b = -.1$, Wald $X^2(1) = 6.332$, $p < .05$, and DistanceF $b = -.1$, Wald $X^2(1) = 6.316$, $p < .05$.

Research has demonstrated strong positive correlations between adjusted gross income and education levels (Baum & Ma, 2007), and adjusted gross income is one of

the primary variables used in calculating the EFC. Furthermore, research has shown that student success (a predictor and contributing factor to retention) is significantly correlated with sex, race, and family income (Betts & Morell, 1999; Tinto, 1993).

The binary logistic regression analysis conducted resulted in models such that for each 10,000 dollar increase in EFC, the odds of attrition decreased from one to 0.90 ($p < .05$). However, the results of each model including *B* and the odds ratio does not account for either Federal Title IV, institutional, or private aid received. Future retention or attrition studies incorporating EFC should also include aid received both as a total and as separate subtotals based on source.

EFC was positively correlated with DistanceT ($r_{pb} = .114, p < .01$), DistanceD ($r_{pb} = .098, p < .01$), and DistanceF ($r_{pb} = .104, p < .01$): EFC increased as distance increased. There is also verifiable by looking at the correlation between EFC and market segmentation. EFC was negatively correlated with the Primary Market ($r_{pb} = -.155, p < .01$) and EFC was positively correlated with the Tertiary Market ($r_{pb} = .138, p < .01$) (the relationship with the secondary market was non-significant). This relationship may be attributed to the fact that McKean County, in which the University of Pittsburgh at Bradford is located, and surrounding counties have some of the lowest median household incomes in the region (U.S. Census Bureau, 2012). The median household income increases as distance increases and as a result, EFC increases as distance increases because EFC is calculated using the Federal Methodology, which takes into account income, some assets, expenses, family size, and other factors.

In-state residency. In each model, in-state residency (i.e., being coded as a Pennsylvania resident) reduced the odds of attrition from one to .731 in the model

incorporating DistanceT and from one to .729 in the models incorporating DistanceD and DistanceF: DistanceT $b = -.314$, Wald $X^2(1) = 7.389$, $p < .01$, DistanceD $b = -.316$, Wald $X^2(1) = 7.379$, $p < .01$, and DistanceF $b = -.316$, Wald $X^2(1) = 7.460$, $p < .01$.

For many institutions, residency (state, county, or even local municipality) determines the tuition rate that is assessed. For example, in 2013, the in-state tuition per semester at the University of Pittsburgh at Bradford was \$6,104. The out-of-state tuition per semester was \$11,406. Out-of-state tuition at the University of Pittsburgh at Bradford is charged to residents of Limestone, New York, who live approximately 7.6 miles from campus. Conversely, students from Philadelphia, Pennsylvania are charged the in-state tuition rate, yet they live 332 miles away. Although in-state residency was negatively correlated with DistanceT ($r_{pb} = -.281$, $p < .01$), DistanceD ($r_{pb} = -.293$, $p < .01$), and DistanceF ($r_{pb} = -.284$, $p < .01$), the strength of the relationship was affected due to the location of the institution relative to the New York border.

Researchers including Astin (1975), Ramist (1981), and Pascarella and Terenzini (2005), all included residency in their studies of the relationship between individual student characteristics noting that residency, affecting cost, significantly affected retention. However, like EFC, future research incorporating residency as a predictor should also account for the amount of Federal Title IV, institutional, and private aid received as some institutions provide higher institutional awards to out-of-state students to offset the residency differential while other institutions may have policies or practices that result in smaller awards of financial aid. Furthermore, a number of states, such as Pennsylvania, allow state aid to be used by students who are attending out-of-state

institutions in reciprocating states. Other states, such as New York, do not allow students to use state aid outside the state of New York.

Domestic underrepresented minority students. In each model, domestic underrepresented minority student status (i.e., not being coded as white, non-Hispanic) reduced the odds of attrition from one to .757 in the model incorporating DistanceT, from one to .756 in the model incorporating DistanceD, and from one to .758 in the model incorporating DistanceF: DistanceT $b = -.278$, Wald $X^2(1) = 6.061$, $p < .05$, DistanceD $b = -.280$, Wald $X^2(1) = 6.169$, $p < .05$, and DistanceF $b = -.277$, Wald $X^2(1) = 6.042$, $p < .05$.

Allen (1999) reported differences in the effects of certain variables in predicting retention, noting that a student's high school rank, first-year college GPA, and the reported desire to finish college were significant variables that accounted for 68% of the variance in first-to-second year retention of underrepresented minority students. However, for White students, high school rank, first-year college GPA, and parental education were significant variables that accounted for 38% of the variance in first-to-second year retention.

The resulting models, developed as part of the binary logistic regression analysis, showed that domestic underrepresented minority student status (i.e., not being coded as white, non-Hispanic) reduced the odds of attrition. This outcome was unexpected and inconsistent with studies that were examined as part of the development of the conceptual framework and literature review that followed.

Significant but weak negative correlations were observed between domestic underrepresented minority student status and EFC ($r_{pb} = -.070$, $p < .01$), SAT/ACT score

($r_{pb} = -.051, p < .01$), and in-state residency ($r_{pb} = -.219, p < .01$). The results of other studies suggest that each of those relationships might contribute to attrition. However, there was also a significant but weak negative correlation between domestic underrepresented minority student status and first-generation college student status ($r_{pb} = -.015, p < .01$). Studies have shown that retention rates are higher among students with one or more parents who have completed a baccalaureate degree.

Many studies have included results showing that race and ethnicity is a factor related to retention. The effect is most notable at institutions that lack diversity in their student body, faculty, and institutional leadership (Swail, 2004). Similar work by Kuh and Love (2004) found that persistence was higher for those students who made social connections reflecting their culture of origin. Schwitzer and Thomas (1998) suggested that underrepresented minority students at predominately White institutions benefit from programs and resources that assist underrepresented minority students in connecting with the institution, resulting in a positive, productive experience. Recent changes in student population demographics at the University of Pittsburgh at Bradford may help to account for the results of the point-biserial correlation, partial correlation, and binary logistic regression analysis.

Between 2005 and 2014, the period represented in this study, the campus has seen a change in the number of underrepresented minority students as a percentage of the student body. In 2005, the percentage of underrepresented minority students comprising the entering first-time, full-time cohort was 8.0%, and the percentage of underrepresented minority students that comprised the entire student body was 5.9%. In 2013, the percentage of underrepresented minority students that comprised the entering first-time,

full-time cohort was 31.2%, and the percentage of underrepresented minority students that comprised the entire student body was 21.2%. These increases were the result of strategic planning, including a new vision by the president of the institution, who assumed his role in 2004. These increases in the diversity of the student body were shortly followed by increases in faculty and staff diversity, and diversity in student programs, activities, clubs, and organizations, all supporting the findings of Swail (2004), Kuh and Love (2004), and Schwitzer and Thomas (1998) as previously mentioned.

On-campus housing status. In each model, on-campus housing status (i.e., not commuting) reduced the odds of attrition from one to .582 in the model incorporating DistanceT, from one to .581 in the model incorporating DistanceD, and from one to .582 in the model incorporating DistanceF: DistanceT $b = -.541$, Wald $X^2(1) = 16.314$, $p < .01$, DistanceD $b = -.585$, Wald $X^2(1) = 18.837$, $p < .01$, and DistanceF $b = -.541$, Wald $X^2(1) = 16.359$, $p < .01$.

This may suggest that a relationship between distance and attrition is observable among commuting students, as was reported by Johnston (2013), and that on-campus housing (and tangible and intangible effects associated) may mitigate distance as it affects attrition. In developing a Theory of Commuter Student Departure, Braxton, Hirschy, and McClendon (2004) extended the work of Tinto by incorporating findings from research that looked at organizational, psychological, sociological, and economic factors affecting attrition. The resulting commuter model emphasized psychosocial factors such as motivation, control issues, self-efficacy, empathy, affiliation needs, and anticipatory socialization, and provided a clear alternative to previous student departure theory, which had primarily focused on the attrition observed by residential students.

Market segmentation. In each model, statistically significant relationships were observed between attrition and each of the three market segmentations: primary, secondary, and tertiary.

Primary market status reduced the odds of attrition from one to .553 in the model incorporating DistanceT, from one to .557 in the model incorporating DistanceD, and from one to .552 in the model incorporating DistanceF: DistanceT $b = -.592$, Wald $X^2(1) = 18.063$, $p < .01$, DistanceD $b = -.585$, Wald $X^2(1) = 18.837$, $p < .01$, and DistanceF $b = -.594$, Wald $X^2(1) = 18.730$, $p < .01$.

Secondary market status reduced the odds of attrition from one to .777 in the model incorporating DistanceT, from one to .781 in the model incorporating DistanceD, and from one to .776 in the model incorporating DistanceF: DistanceT $b = -.252$, Wald $X^2(1) = 4.727$, $p < .05$, DistanceD $b = -.248$, Wald $X^2(1) = 4.727$, $p < .05$, and DistanceF $b = -.254$, Wald $X^2(1) = 4.854$, $p < .05$.

Tertiary market status increased the odds of attrition from one to 1.287 in the model incorporating DistanceT, from one to 1.281 in the model incorporating DistanceD, and from one to 1.289 in the model incorporating DistanceF: DistanceT $b = .252$, Wald $X^2(1) = 4.727$, $p < .05$, DistanceD $b = .248$, Wald $X^2(1) = 4.727$, $p < .05$, and DistanceF $b = .254$, Wald $X^2(1) = 4.854$, $p < .05$.

Reiterating that distance from home to the institution, as measured in travel time (DistanceT) or miles (DistanceD and DistanceF), was not a significant predictor, the researcher assumed that market segmentation, which was thought to be acting as a proxy variable for distance, would also not be a significant predictor. In addition, statistically significant correlations were observed between EFC and Primary Market ($r_{pb} = -.155$, $p <$

.01), and EFC and Tertiary Market ($r_{pb} = .138, p < .01$). EFC increased as distance increased (i.e., the degree to which families could assist with paying for college increased as distance increased). Furthermore, statistically significant correlations were observed between SAT/ACT score and Primary Market ($r_{pb} = -.150, p < .01$), and SAT/ACT score and Tertiary Market ($r_{pb} = .109, p < .01$). SAT/ACT score increased as distance increased. Although both family income and SAT/ACT scores were higher among those students coded as originating from the tertiary market, their likelihood of attrition was greater.

The researcher has speculated that the location of the University of Pittsburgh at Bradford may play a role. The institution is the only four-year, degree granting institution in a geographic area that is equivalent in size to the state of Connecticut. Regional, post-secondary education options are limited. Therefore, do limited options and a student's sense (or the reality) of being place-bound reduce the degree to which other predictors affect predictive models of attrition?

Sex and first-generation status. Statistically significant relationships with attrition at the University of Pittsburgh at Bradford were not observed for sex ($p = .247$ and $.248$) or first-generation college student status ($p = .216$ and $.217$). Both are variables that are commonly cited in retention and persistence research and are often used in predictive modeling.

Limitations

This study has several limitations requiring review and discussion. As all research has limitations and threats to both internal and external validity, these limitations and threats need to be delineated so that the true value of the results can be determined. First,

this case study utilized a quantitative non-experimental design. The researcher was interested in studying naturally occurring events. Therefore, there was no manipulation of the independent variables. As such, this study explored how variables were related, but it could not determine cause and effect.

Another limitation is that this case study focused on a single population: all cases were domestic, traditional-aged, first-time, full-time students at the University of Pittsburgh at Bradford from 2005 to 2013. While the results may be of general interest and have some broad-based application, it can only be specifically interpreted to the University of Pittsburgh at Bradford.

As was outlined in chapter 1, confirmation bias is a potential limitation (Sandelowski, 1986; McMillan, 2008). There is the possibility that the author searched for or interpreted data and information in a way that confirmed his preconceptions, leading to statistical errors.

A large sample size (in fact, the entire population of qualifying cases) was used and may identify significant relationships for statistical reasons and not because the relationships actually exist (McMillan, 2008).

Data was recoded to assure that data fields and definitions were consistent, and this may have introduced errors that threaten reliability and validity.

There are multiple causes affecting a student's decision to persist or leave, and resulting independent variables outside of the focus of this case study were not explored. The variables studied were predominately individual characteristics, but variables such as institutional practices, student behaviors, affiliations, and student utilization of services were not studied.

Implications for Practice

Of those students who are least likely to be at risk for attrition based on the analyses of other variables that have been shown to have a significant relationship with attrition at the University of Pittsburgh at Bradford, distance does not appear to be a factor at all.

As was previously noted, Tinto (1993) warned that his theory “is not a systems model of departure” (p. 112). Tinto and other researchers have cautioned against the use of these theories and models in post-facto reviews of departure at other institutions. Retention and attrition models should be uniquely designed based on longitudinal institutional-specific data. Not all colleges and universities are the same; therefore, there cannot be a single, successful retention strategy (Hossler, 1991; Tinto, 1993; Milron, 2006). However, there is a core set of variables (demographic, socioeconomic, and academic) with which an exploration of the causes of attrition can begin. This case study examined distance as a potential variable that should be examined by other institutions of higher education when analyzing retention and attrition factors affecting the profile of a student body that is geographically heterogeneous.

However, the results of this case study have implications for the practitioners engaged in areas affecting enrollment management at the University of Pittsburgh at Bradford. Recommendations for practice are presented for the following areas: dissemination, marketing and advertising, admissions and recruitment, advising, and student affairs and student activities.

Dissemination. The broad circulation of important information is critical in organizational planning and affects institutional success. Dissemination, as a

recommended practice, refers to the communication of research results to specific members of various teams within the institution in order for these stakeholders to make better-informed decisions that ultimately lead to improved student and institutional outcomes. In this case, the areas recommended for the dissemination of the results are admissions, communications and marketing, financial aid, faculty, advisors, members of the enrollment management planning committee, and the president's leadership team. A good dissemination strategy will result in increased awareness of the research, maximizing the impact that the research can have in informing planning, decision-making, and improving outcomes. Regardless of how innovative or transformative the results of research are, they will not make an impact unless they are communicated to the various audiences who will most benefit from them (e.g. those engaged in institutional planning and decision-making). At the University of Pittsburgh at Bradford, it is important to disseminate the findings of this case study, which contradict the assumption held by most faculty and staff at the campus that distance plays a significant role in a student's decision to leave the institution.

Marketing and advertising. The results of this study have implications for various aspects of marketing and advertising. For the purposes of this discussion, marketing is the process of identifying and developing potential products or services for the marketplace and convincing potential students and their families (i.e., because of the potential influence of family on a student's selection of institution) that the products or services that you are offering are the right ones for them. It involves understanding students who are currently enrolled and identifying potential new students and their families in terms of sources, profiles, expectations, access, and behaviors. It also involves

an awareness of the overall marketplace in terms of access, supply, demand, and competition. Finally, it includes the development of an image and brand for the institution and the selection or creation of design elements, taglines, and other factors affecting that image and brand (Lovering, n.d.). Advertising is a single component of marketing: It is the process of creating awareness about the organization and its product or services to potential and existing customers. Advertising includes the placement and management of ads in newspapers, on billboards, at various events, and on television and radio. It also includes various aspects and means within the Internet (e.g., web page and website development, search engine optimization, placement of ads on sites, social media presence and management) as well as direct mailings to prospective students and their families. It can also include the management of information and messages to those audiences influencing students' decisions regarding choice of institution (e.g., extended family, guidance counselors, teachers, coaches). The results of this study could influence many aspects of marketing and advertising. Recommendations for the utilization of the findings of this case study are grouped into two categories: message and market.

The findings of this case study indicate that distance does not play a role in student retention at the University of Pittsburgh at Bradford. Those findings can influence the messages that are conveyed as part of developing the institutional brand and the resulting advertisements. For example, the brand promise of the University of Pittsburgh at Bradford is, "The University of Pittsburgh at Bradford is a safe, friendly institution for students who want to earn a world-renowned degree in a personalized environment" (University of Pittsburgh at Bradford, 2007). A recommendation might be to review this message, test it in focus groups, and consider amending this statement to emphasize that

the institution is suited to students from any region who wish to earn a world-renowned degree in a personalized environment.

The institution may want to consider new marketing initiatives targeting prospective students in the tertiary market. The tertiary market, as defined by the institution, comprises all other potential recruitment areas beyond the primary and secondary markets. New marketing initiatives could include print, media, or Internet campaigns in areas currently underserved by institutions of higher education as well as regions in which there is a significant number of students who might match the profile of students currently enrolled. The messages communicated in new advertisements placed in markets furthest from the institution might include language that refers to the successes of students who come from the greatest distances. Messages could include statistics as well as personal testimonials from current students and alumni who have been successful at the institution. An emphasis might be placed on securing testimonials from students from urban or semi-urban areas, and the focus could be on how they made the adjustment from living in the city to joining a community of learners in the Allegheny Mountains of Pennsylvania. Another recommendation is to review recruitment materials and advertisements (e.g., view books, flyers, open house or other event invitations, letters and emails that are part of the admissions communications sequence, advertisements) and to strategically develop language that links success with students who live several hours away. As part of this review, communications and advertisements that feature majors, services, and activities on campus could be developed that specifically target students from greater distances.

How might the institution utilize the findings of this study with regards to website and web page design? Similar new messages utilized in marketing, advertising, and recruitment materials should be adapted to the institution's website and web page design. In light of the findings, the institution may want to modify or develop new messages (i.e., text and pictures) designed to engage and to create interest in the university from the perspective of a prospective student who comes from a significant distance. Additionally, website and web page design present unique opportunities for incorporating what has been learned from this case study.

For example, utilizing JavaScript code, the institution could design and implement dynamic content on those web pages identified through analytics as likely first landing points for first-time visitors to the website. Specific messages or redirects to specific pages that are designed to engage and encourage prospective students from greater distances to apply could be designed to change depending on the geolocation of the visitor. For example, any HTML content (e.g., pictures, rich text, tables, and video) could be tagged to display only when a visitor is from a chosen country, state, or city. Conversely, content can be displayed when the visitor is not from the chosen location. A more specific example would be a testimonial from a successful, current student or recent alumnus, regarding distance and making the transition from their home to their "home away from home." This testimonial would appear and change based on the location of the IP address that was utilized by the website visitor.

Admissions. Recommendations concerning language and messages that would be incorporated into new and existing recruitment materials were previously mentioned. However, the findings of this study could also influence other areas of the recruitment

process: the purchasing of names of prospective students, territory management, travel, and messaging at events.

The University of Pittsburgh at Bradford currently purchases the names and contact information for college-bound seniors from several different sources including College Bound Selection Service (CBSS), The College Board, National Research Center for College & University Admissions (NRCCUA), and ACT. Requests for names from these organizations are based on student profile, intended major, and location. The highest percentage of names purchased (i.e., as a percentage of the possible total number of college-bound high school graduates in each county) are those from the primary market, followed by the secondary market, and lastly, the tertiary market. This practice exists due to historical precedence as well as the assumption, based on anecdotal evidence, suggesting that distance played a role in students' decisions to leave the institution. However, the findings of this case study suggest that the current criteria used for purchasing names should be reexamined, and that distance (accounting for residency) should not play a factor in the selection process. The institution may want to consider purchasing a greater number of names of potential students from greater distances. To support this initiative, previously recommended changes to recruitment materials and increased advertising in the areas in which additional names were purchased are required.

The findings of this case study suggest that residency does play a role in attrition. Pennsylvania residency reduced the odds of attrition from one to .731 in the model incorporating DistanceT and from one to .729 in the models incorporating DistanceD and DistanceF ($p < .01$). Pennsylvania residency or out-of-state status determines the tuition rate that is assessed. In 2013, the in-state tuition per semester at the University of

Pittsburgh at Bradford was \$6,104. The out-of-state tuition per semester was \$11,406. Therefore, it is recommended that during the review of the process and criteria used to purchase the names of prospective students, the practitioners remain mindful of the implication associated with the state of origin. This, in turn, could result in further exploration of the potential variables associated with attrition at this institution, and specifically, the potential relationship that might exist between attrition and residency, EFC, and institutional financial aid. It is possible that the effect could be lessened through the use of institutional financial aid.

As part of a holistic recommendation that includes review and changes to the messages being utilized, advertising, and name purchasing, I recommend that the institution also review its current practices concerning territory management and travel. Territory management is a process by which customer accounts (i.e., prospective students) are grouped based on a defined set of criteria (normally, but not always geographic location). Territory management can result in travel and logistical efficiencies as admissions counselors can better plan college fairs and high school visits in ways that reduce travel time and distances as well as the number of trips that are made, thereby saving the institution money in hotel accommodations, car rentals, and per diem charges. Territory management can also create opportunities for more efficient advertising in regions prior to visits by admissions counselors. I recommend that the institution review its current territory management practices, including the way in which geographic regions are assigned to admissions counselors, how travel is planned and managed (the number of consecutive days of travel, the utilization of a rental fleet, opportunities for several visits within one trip, the potential for using part-time counselors located in areas of interest),

how high school visits and college fairs are prioritized, when advertising occurs prior to an admissions counselor's visit, and the potential for new territories as new markets are identified and additional names are purchased.

Finally, I recommend that a small team review materials, messages, videos, and the slide decks used at open houses, high school visits, college fairs, orientation, and other recruitment events and activities. The major finding of this study is that there is not a relationship between the distance from home to the institution and attrition at the University of Pittsburgh at Bradford. This finding can be used to promote the positive message that students, whether local or from far-away, can be equally successful. Emphasizing this message may positively affect the attitudes and perceptions of students coming from greater distances (either actual or perceived). Although the findings of this study show no relationship between distance and attrition, the relationship between attrition and matriculation was not examined. Paulsen (1990) noted that the attractiveness of colleges, in general, as well as a specific institution, increased as the distance from home decreased. The importance of this research is that distance not only influences choice of institution, but it also can be seen as influencing a baseline perception of the attractiveness or fit of the institution even prior to matriculation and initial enrollment. An exploration of this premise should be conducted at the University of Pittsburgh at Bradford. However, until that is done, and if the previous recommendations are implemented, then clear messages about the successes of students who come from greater distances should be communicated within admissions materials and during recruitment events; if the question emerges during student inquiries or visits to campus, it can be addressed with data from this study.

Advising. It is important to disseminate the findings of this case study, which contradict the assumption held by most faculty and staff at the campus that distance plays a significant role in a student's decision to leave the institution. Advisors (both faculty advisors and staff who serve as full-time advisors) need to be made aware that there is not a relationship between attrition and distance at the University of Pittsburgh at Bradford. This paradigm represents a significant change in advisors' current belief system. It is important that advisors internalize this change in understanding so that it is reflected in their interactions with students. When advisors meet with students who may be having difficulty with "separation, transition, and incorporation" (Van Gennep, 1960, p. 11) due to perceived distance, the advisor can refer to data from this study that shows that there is no relationship between distance from home and leaving the institution (prior to the student's second year). The advisor can engage in a positive application of self-fulfilling prophecy in which they can state (with a certain level of confidence) that their advisee will not leave the institution because it has been found that other students do not leave due to distance.

Student Affairs and Student Activities. Like the advisors previously mentioned, staff in the areas of student affairs and student activities need to be made aware that there is not a relationship between attrition and distance at the University of Pittsburgh at Bradford. Resident assistants and resident directors, staff in Counseling Services, orientation leaders, and faculty who teach the first-year experience course (Freshman Seminar) present and discuss (one-on-one and during group presentations) the topics of homesickness and making the transition from the home environment to college. Also, like the advisors, it is important that staff members within the areas of student affairs and

student activities, and instructors in the first-year experience modify their approaches in discussing homesickness and students' transitioning to the institution. Presentations about these topics should be made during orientation sessions, residence hall programming, the first-year experience class, and in other student life programming. The staff can point to the findings of this case study to disassociate distance from students' difficulty with "separation, transition, and incorporation." Once again, we can engage students using a positive application of the self-fulfilling prophecy in which we state that students do not leave the institution due to the distance from home. This would imply that all students have an equal opportunity for success at the institution, regardless of distance from home.

Recommendations for Further Research

As was previously noted, retention and attrition theories and models should be uniquely designed based on longitudinal institutional-specific data. Colleges and universities differ in many different ways. Some examples include student profile, institutional selectivity, size, setting, location, mission, affiliation, culture, price, size of endowment, institutional aid, instructional programming, facilities, curriculum, policies and procedures, student life and activities, town-gown relationship, and athletics. Therefore, there cannot be a single, successful retention strategy (Hossler, 1991; Tinto, 1993; Milron, 2006). The literature review included studies conducted at institutions that represent many of the classification categories of the Carnegie Classification of Institutions of Higher Education.

Questions that could generate future research include:

- Have studies of attrition, leading to the development of predictive models, been conducted at institutions representing all institutional types and classifications?
- Has a set of predictive variables emerged, which can be utilized across institutional types and classifications?
- Are there specific or unique institutional characteristics or types that affect the outcomes of studies or lead to the utilization of certain predictive variables not commonly used?
- If this study was replicated at other institutions (both similar and dissimilar), would the results be the same: that the variable of distance does not enhance the predictive modeling of attrition?

Exit interviews are conducted with each attriting full-time student before or just after they leave the University of Pittsburgh at Bradford. A quantitative review of the data suggested that the institutional distance from home often played a role in students' decision-making processes. Other studies have cited similar conclusions. Cobb (2001) noted in a study of first-time, full-time students at the University of Oklahoma that academic difficulties and distance from home were the top two reasons cited for students' non-persistence. However, perhaps the perception of distance is serving as a proxy for a student's inability or difficulty in making a social, cultural, or geographic transition. The definition of distance may be more complex and may include social distance (Cash, 1990) and cultural distance (Malocsay, 2004). Distance may refer to a student's perception of the degree to which the setting is different from their home. Therefore, distance can include a perceived integrational "distance" based on differences in the

locale, population, education levels, and culture, as well as actual distance. Davis (2010) noted that the closer a student lived to the institution, the greater the degree to which the student was confident in the decision to attend. Furthermore, the greater the level of confidence in that decision, the higher the likelihood of persistence. Johnson (2010) noted that culture shock and feelings of isolation were two potential results of homesickness that could vary depending on the distance from home, but he also stated that a student's self-reported perception of distance was a better predictor of attrition than the actual distance.

Questions that could generate future research include:

- What do the results of exit interviews at the University of Pittsburgh at Bradford suggest—is distance a factor in a student's decision to leave the institution or is distance serving as a proxy for another variable or reason?
- Is there a relationship between the *perceived* distance from home to the institution, as experienced by students, and attrition at the University of Pittsburgh at Bradford (and at other institutions, both similar and dissimilar)?
- Is there a relationship between the perceived distance from home to the institution and the degree to which students report making social, cultural, and geographic transitions at the University of Pittsburgh at Bradford (and at other institutions, both similar and dissimilar)?
- Conversely, does a sense of being place-bound affect the degree to which other predictors affect predictive models of attrition?

As was stated previously, research involving homesickness is relatively new.

During the last two decades, research has been conducted in describing what

homesickness is, its causes and effects, and ways in which it might be mitigated. Only a few recent studies of homesickness have focused on students in the United States of America. Previous studies have focused on students in Europe and New Zealand.

Questions that could generate future research include:

- Is homesickness a measurable variable that is a significant predictor of attrition at the University of Pittsburgh at Bradford (and at other institutions, both similar and dissimilar)?
- How might homesickness be affected by greater distances between home and institution as experienced by students in American colleges and universities?
- Does homesickness affect the perception of distance, or does one's perception of distance influence homesickness?
- Is homesickness an indication of difficulty in making a social, cultural, or geographic transition?
- If homesickness is an underlying cause of attrition, does distance or perceived distance serve as its proxy at the University of Pittsburgh at Bradford (and at other institutions, both similar and dissimilar)?

Adaptability is the degree to which a student can adjust readily to different conditions. Adaptability is a student's ability to change something or oneself to fit into occurring changes (i.e. the change from the student's culture of origin to the dominant culture of the institution) (Andresen & Gronau, 2005). Research on adaptability, as it relates to the global business environment, has been quite extensive and has resulted in the development of several commonly used assessment instruments, and a commonly referred to set of attitudes, traits, and skills: conscientiousness and action oriented,

flexibility, emotional stability, openness and open-mindedness, sociability and extroversion, and cultural sensitivity and cultural intelligence (Ang, Dyne, & Koh, 2006).

Research on adaptability, as it pertains to the general measurement of the ability of students to bridge the gap between their personal cultures and the culture of the institution, has been limited. One specific focus of research has been the adaptability of students who engage in study abroad experiences (i.e., the degree to which they can be successful in a foreign environment).

Perhaps these areas are all related: perceived distance; social, cultural, and geographic transition; homesickness; and adaptability. These concepts each require greater research, as the literature suggests.

Questions that could generate future research include:

- Is adaptability (self-reported or otherwise measured) a variable that is a significant predictor of attrition at the University of Pittsburgh at Bradford (and at other institutions, both similar and dissimilar)?
- Does adaptability affect the perception of distance, or does one's perception of distance influence adaptability?
- If adaptability is an underlying cause of attrition, does distance or perceived distance serve as its proxy at the University of Pittsburgh at Bradford (and at other institutions, both similar and dissimilar)?
- Does one or more relationships exist between perceived distance; social, cultural, and geographic transition; homesickness; and adaptability and attrition at the University of Pittsburgh at Bradford (and at other institutions, both similar and dissimilar)?

Conclusion

This *ex-post facto* case study examined the attrition from the first to the second year for 2,837 domestic, traditional-aged, first-time, full-time students (freshmen) matriculated and enrolled during the fall semesters of 2005 through and including 2013 at the University of Pittsburgh at Bradford. The literature review showed that prior exploration of this relationship has been sporadic, results were varied, and the methodology used was incomplete. However, anecdotal evidence suggested that a relationship existed. Point-biserial correlation, partial correlation, and binary logistic regression analyses were conducted. The results of the analyses determined that the relationship between the institutional distance from home and the attrition of traditional-aged, first-time, full-time students prior to the second year at the University of Pittsburgh at Bradford was weak and nonsignificant. Although the findings were nonsignificant, the methodology involved the use of the Google Maps Distance Matrix Application Programming Interface (API) and the Google Maps API to accurately measure the institutional distance (in miles and minutes) from home. This approach enhances the literature on the subject of attrition and retention studies. Previous studies of the potential attributes related to attrition have rarely included a measurement of the distance from the student's permanent home address to the post-secondary institution. In those few studies that incorporated distance, the distance was calculated based on a straight line from a student's permanent home address to the post-secondary institution. In this case study, utilization of the Google Maps Distance Matrix API and the Google Maps API allowed the researcher to accurately measure and incorporate the orthodromic distance, distance based on the Google recommended route, and distance, as measured in travel time based

on the Google recommended route, as independent variables. In October of 2015, the researcher conducted a search of all dissertations in ProQuest's academic databases. No dissertations focusing on any education-related topics were found to have utilized the Google Maps Distance Matrix API or the Google Maps API.

Five major trends are influencing student enrollments across higher education: national enrollment growth is slowing, many regions of the country are facing significant demographic shifts, substantial changes have occurred in the way higher education is funded and the way that families are paying for higher education, a substantial intensification of political, societal, and institutional pressures to increase retention and completions, as well as changes resulting from new modalities in learning (BPU's Center for Online & Digital Learning, 2015). These five trends are creating new challenges for enrollment management at institutions across the country. The competition among institutions will increase. In response, institutions will have to allocate more money to staffing, travel, marketing, advertising, merit and need-based institutional aid, additional student-success services, and student programming and activities just to maintain the size and quality of their incoming first-year class and overall enrollments. Due to these increased costs and pressures, institutions will pay greater amounts of attention to the recruitment of students who are likely to persist and complete a credential or degree. As a result, it will be important for similar studies to be conducted on an institution-by-institution basis, and for all researchers to disseminate their findings to continue to build the body of literature on factors that influence student attrition, retention, and persistence.

References

- ACT. (2007). *ACT-SAT concordance tables*. Retrieved from <http://www.act.org/aap/concordance/pdf/report.pdf>
- ACT. (2008). *2008 retention/completion summary tables*. Retrieved from http://www.act.org/research/policymakers/pdf/retain_trends.pdf
- ACT. (2010). *2010 retention/completion summary tables*. Retrieved from http://www.act.org/research/policymakers/pdf/10retain_trends.pdf
- ACT. (2013a). *2012 retention/completion summary tables*. Retrieved from http://www.act.org/research/policymakers/pdf/12retain_trends.pdf
- ACT. (2013b). *College choice report—part 2: Enrollment patterns*. Retrieved from <http://www.act.org/collegechoice/13/pdf/CollegeChoiceReport2013-2.pdf>
- Aiken, L. R., Jr. (1964). The prediction of academic success and early attrition by means of a multiple-choice biographical inventory. *American Educational Research Journal, 1*, 127-135.
- Aldrich, J., & Forrest, D. N. (1984). *Linear probability, logit, and probit models*. Beverly Hills, CA: Sage.
- Allen, I. E., & Seaman, J. (2013) *changing course: Ten years of tracking online education in the United States*. Wellesley MA: Babson College/Quahog Research Group.
- Andresen, K., & Gronau, N. (2005). *An Approach to Increase Adaptability in ERP Systems*. In: *Managing Modern Organizations with Information Technology*. Proceedings of the 2005 Information Resources Management Association International Conference, 2005.

- Ang, S., Van Dyne, L., & Koh, C.K.S. (2006). Personality correlates of the four factor model of cultural intelligence. *Group and Organization Management*, 31, 100-123.
- Arreguin, M. (2008). *Spatial interaction models of freshman recruitment and retention at Stephen F. Austin State University for the years 2001-2005*. (Masters dissertation, Austin State University, 2008). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 1463288).
- Astin, A. W. (1975). *Preventing students from dropping out*. San Francisco: Jossey-Bass.
- Astin, A. W. (1984). Student involvement: A development theory for higher education. *Journal of College Student Personnel*, 25, 297-308.
- Astin, A. W. (1991). *Assessment for excellence*. New York: Macmillan.
- Astin, A. W. (1993). *What matters in college?* San Francisco: Jossey-Bass.
- Astin, A. W. (1997). How “good” is your institution’s retention rate? *Research in Higher Education*, 38(6), 647-658.
- Astin, A. W., & Oseguera, L. (2005). College and institutional influences of degree Attainment. In A. Seidman (ed.), *College student retention: Formula for student Success*. Westport, CT: Praeger Publishers.
- Atkinson, L. (2008). *Factors impacting student retention on the regional campuses and centers of Ohio University*. (Doctoral dissertation, Ohio University, 2008). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3319028).
- Baldwin, J. L. (2013). [First-time, full-time retention data of the University of Pittsburgh]. Unpublished raw data.

- Bank, B. J., Slavings, R. L., & Biddle, B. J., (1990, July). Effects of peer, faculty, and parental influences on students' persistence. *Sociology of Education*, 63(3), 208-225.
- Baum, S., & Ma, J. (2007). *Education pays: The benefits of higher education for individuals and society*. New York, NY: College Board.
- Bauman, P. (2002). Student Retention: What You Can Control, & How. *Distance Education Report*, 6(16).
- Bay, K. S., & Hakstianz, R. (1972). Note on the equivalence of the significance test of the partial point-biserial correlation and the one-factor analysis of covariance for two treatment groups. *Multivariate Behavioral Research*, 7(3), 391-395.
- Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12, 155-187.
- Bean, J. P. (1985). Interaction effects based on class level in an explanatory model of college student dropout syndrome. *American Educational Research Journal*, 22, 35-64
- Bean, J. P. (1990). Why students leave: Insights from research. In D. Hossler, J. P. Bean, & Associates, *The Strategic Management of College Enrollments* (pp. 147-169). San Francisco: Jossey-Bass.
- Bean, J. P., & Metzner, B. S. (1985, Winter). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research*, 55, No. 4, 485-540.
- Bebergal, J. (2003). *Predicting retention of first-year college students*. (Doctoral

- dissertation, Florida Atlantic University, 2003). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3080484).
- Berger, J. B., Ramirez, G. B., & Lyons, S. (2012). Past to present: A historical look at retention. In A. Seidman (Ed.) *College student retention 2nd Edition: Formula for student success*. New York. ACE/Rowman & Littlefield.
- Betts, J. R., & Morell, D. (1999). The determinants of undergraduate grade point average: the relative importance of family background, high school resources, and peer group effects. *The Journal of Human Resources*, 34(2), 268-293.
- BPU's Center for Online & Digital Learning (Producer). (2015, August 27). Foundations of strategic enrollment planning [Video file]. Retrieved from <http://www.youtube.com/watch?v=pzZ1fp2hzZY>
- Braxton, J. M., Hirschy, A. S., & McClendon, S. A. (2004). *Understanding and reducing college student departure*. ASHE-ERIC Higher Education Report, Vol. 30, No. 3. San Francisco: Jossey-Bass.
- Burns, R. B., & Burns, R. A. (2009). *Business research methods and statistics using SPSS*. Thousand Oaks, CA: Sage Publications, Inc.
- Cabrera, A.F., Nora, A., & Castañeda, M.B. (1992). The role of finances in the persistence process: A structural model. *Research in Higher Education*, 33(5), 571-93.
- Calvert, M. (2011). *GeodesiXA: Geocoder Addin for Excel*. Retrieved from <http://www.calvert.ch/geodesix>
- Campbell, J. D. (2008). *Analysis of institutional data in predicting student retention*

- utilizing knowledge discovery and statistical techniques.* (Doctoral dissertation, Northern Arizona University, 2008). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3318499).
- Cash, R. W. (1990). *Freshmen retention at two Seventh-day Adventist higher education institutions.* (Doctoral dissertation, Andrews University, 1990). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 9031251).
- Chambers, J. M., Cleveland, W. S., Kleiner, B., & Turkey, P. A. (1983). *Graphical methods for data analysis.* Boston: Duxbury Press.
- Cobb, C. M. S. G. T. (2001). *The concept of brand identity in relation to students' intent-to-persist.* (Doctoral dissertation, The University of Oklahoma, 2001). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3006667).
- Cohen, J., & Cohen, P. (1983) *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum
- The College Board. (2009). *ACT and SAT concordance tables* [Office of Research and Development RN-40]. New York: NY: Author.
- Coolican, H. (2014). *Research methods and statistics in psychology* (6th ed.). New York, NY: Psychology Press
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five traditions.* Thousand Oaks, CA: SAGE Publications.
- Creswell, J. (2009). *Research design: Qualitative, quantitative and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: SAGE Publications.
- Crissman Ishler, J. L., & Upcraft, M. L. (2005). The keys to first-year student persistence. In Upcraft, Gardner & Barefoot's (Eds.) *Challenging and supporting the first-*

- year student: A handbook for improving the first year of college*. San Francisco: Jossey-Bass.
- Cuff, N. (1929). Problem of elimination from college. *School and Society*, 30, 550-558.
- Dahl, R. W. (1982). *College attendance and institutional choice: Results from the Kentucky longitudinal study*. Paper presented at the Annual Forum of the Association of Institutional Research, Denver, June. Retrieved from ERIC database. (ED 220227).
- Davis, C. (2010). *Noncognitive predictors of academic performance and persistence in horizontal and vertical transfer students by academic level*. (Doctoral dissertation, Old Dominion University, 2010). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3407399).
- Dorans, N. J., Lyu, C. F., Pommerich, M., & Houston, W. M. (1997). Concordance between ACT assessment and recentered SAT I sum scores. *College and University*, 73, 24-31
- Dorans, N. (2000). *Distinctions among classes of linkages* [RN-11]. New York, NY: College Board.
- Dorans, N. (2008). *The practice of comparing scores on different test* [ETS Report No. 6]. Retrieved from http://www.ets.org/Media/Research/pdf/RD_Connections6.pdf
- Durkheim, E. (1951). *Suicide: A Study in Sociology*. New York: The Free Press.
- FAFSA.org. (2011). Free application for federal student aid. Retrieved from <http://www.fafsa.ed.gov/>
- Feder, D. (1940). Factors which affect achievement and its prediction at the college level. *Journal of the American Association of Collegiate Registrars*, 15, 107-117.

- Fisher, R. A. (1933). The arrangement of field experiments. *Journal of the Ministry of Agriculture, 9*, 503-513.
- Fisher, S. (1989). *Homesickness, cognition, and health*. Hove, East Sussex UK: Lawrence Erlbaum Associates.
- Fisher, S., Frazer, N., & Murray, K. (1986). Homesickness and health in boarding school children. *Journal of Environmental Psychology, 6*, 35-47.
- Fisher, S., Murray, K., & Frazer, N. (1985). Homesickness, health and efficiency in first year students. *Journal of Environmental Psychology, 5*, 181-195.
- Fredericksen, N., & Schrader, W. B. (1951). *Adjustment to college*. Princeton: Educational Testing Service.
- French, A. J. (1982). *Transition from school to university an environmental approach: a comparative study of students at the universities of Durham and Loughborough*. (Doctoral thesis, Durham University, 1982). Retrieved from Durham E-Theses Online: <http://etheses.dur.ac.uk/7392>.
- Fry, R. (2013, August). *A rising share of young adults live in their parents' home*. Washington, DC: Pew Research Center
- Garson, G. D. (2010). Logistic regression: Statnotes, from North Carolina State University, Public Administration Program [Online]. Retrieved from http://estudijas.lu.lv/file.php/747/Jaunais_satars/LogReg/Logistic_Regression_1.pdf
- Gillock, K. L. (1998). *Factors affecting students' adjustment and attrition during the transition to college*. (Doctoral dissertation, University of Illinois at Chicago, 1998). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 9934761).

- Gilmour, J. E., Spiro, L. M., & Dolich, I. J. (1978). *How High School Students Select a College*. Retrieved from ERIC database. (ED 208705).
- Gravetter, F. J., & Wallnau, L. B. (2009). *Statistics for the Behavioral Sciences* (8th ed.). Belmont, CA: Cengage Learning.
- Goodman, P. E. (1999). *Pre- and post-matriculation correlates of student retention within a community college setting*. (Doctoral dissertation, East Tennessee State University, 1999). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 9955486).
- The Google Maps Distance Matrix API. (n.d.). Retrieved from <http://developers.google.com/maps/documentation/distancematrix/intro>
- Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods*, 2 (3), 235-247.
- Gomm, R. Hammersley, M. and Foster, P. (2000) Introduction, in R. Gomm, M. Hammersley and P. Foster (Eds.), *Case Study Method*, London: Sage Publications
- Goodwin, C. J. (2009). *Research in Psychology: Methods and design*. Hoboken, NJ: Wiley.
- Habley, W. R., Bloom, J. L., & Robbins, S. (2012). *Increasing persistence: Research-based strategies for college student success*. San Francisco. Wiley & Sons, Jossey-Bass.
- Hagedorn, L.S. (2012). How to define retention: A new look at an old problem. In A. Seidman (Ed.), *College Student Retention: Formula for Student Success*. Updated second edition. ACE/Praeger. Rowman & Littlefield Publishers, Lanham, MD. 81-99.

- Hatch, C. W. (1983). *A procedure for predicting freshman retention: an institution-specific approach*. (Doctoral dissertation, University of South Carolina, 1983). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 8319294).
- Hearn, J. C. (1984). The relative roles of academic, ascribed, and socioeconomic characteristics in college destinations. *Sociology of Education*, 57, 22-30.
- Higher Education Research Institute (2007, May). *First in my family: A profile of first-generation college students at four-year institutions since 1971*. Retrieved from <http://www.heri.ucla.edu/pdfs/pubs/briefs/firstgenresearchbrief.pdf>.
- Hines, J. D. (1999). *Variables that discriminate between persisting and nonpersisting, first-year undergraduate students in an open-admissions, multicampus, commuter college system*. (Doctoral dissertation, Andrews University, 1999). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 9929069).
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*, (2nd ed.). New York, USA: John Wiley and Sons.
- Hossler, D. (2000). Effective admissions recruitment. In G. H. Gaither, *Promising Practices in Recruitment, Remediation, and Retention: New Directions for Higher Education*, No. 108. San Francisco: Jossey-Bass.
- Hossler, D. (Ed.). (1991). *Evaluating student recruitment and retention programs*. New Directions for Institutional Research, no. 70. San Francisco: Jossey-Bass.
- Hossler, D., Bean, J. P. (1990). *The strategic management of college enrollments*. San Francisco: Jossey-Bass.
- Howe, N., & Strauss, W. (2000). *Millennials rising: The next great generation*. New

York: Vintage Books.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression (2nd Edition)*. New York: Wiley.

Huitema, B. (2011). *The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies* (Vol. 2). Hoboken: Wiley.

Hussar, W. J., & Bailey, T. M. (2013). *Projections of Education Statistics to 2021* (NCES 2013-008). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>.

Iffert, R. E. (1957). *Retention and withdrawal of college students* (Vol. 1). Washington, DC: U.S. Government Printing Office.

Jackson, A. L. (1998). *Factors related to freshman adjustment to college*. (Doctoral dissertation, The University of Southern Mississippi, 1998). Retrieved from ProQuest Dissertations and Theses Database. (AAT 9901339).

Jackson, G. A. 1978. Financial aid and student enrollment. *Journal of Higher Education*, 49(6), 548-574.

Jackson, N. H. (2008). *Factors related to nonacademic adjustment of freshmen students*. (Doctoral dissertation, The University of Southern Mississippi, 2008). Retrieved from ProQuest Dissertations and Theses Database. (AAT 3346533).

Johnson, J. (2010). *The effects of extracurricular activities on academic performance and retention in the Middle Tennessee State University Horse Science program*. (Doctoral dissertation, Middle Tennessee State University, 2013). Retrieved from ProQuest Dissertations and Theses Database. (1437225479).

- Johnson, L. H. (2013). *Predicting first-year grade point average and retention of student-athletes using demographic, academic, and athletic variables*. (Doctoral dissertation, Ball State University, 2010). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3403233).
- Johnson, R. (2007) *Academic characteristics of student retention in college using logistic regression analysis*. (Doctoral dissertation, Touro University International, 2007). Abstract from ProQuest Dissertations and Theses Database. (AAT 3269168).
- Jordan, K. (2013, July). MOOC completion rates: The data. *E-Learning & Innovative Pedagogies*. Retrieved from <http://ubi-learn.com/the-latest-news/mooc-completion-rates-the-data>
- Jones, M.P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association*, 91,222-230.
- Kaiser, C. A. (2005). *Assessing admission criteria (high school grade point average and SAT) as a means to understand the influence on retention at a North East university*. (Doctoral dissertation, Seton Hall University, 2005). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3226838).
- Kelly, L. J. (1996, May). Implementing Astin's I-E-0 Model in the study of student retention: A multivariate time dependent approach. Paper presented at the 36th Annual Forum of the Association for Institutional Research, Albuquerque, NM.
- Knapp, L.G., Kelly-Reid, J.E., and Ginder, S.A. (2012). *Enrollment in Postsecondary*

- Institutions, Fall 2011; Financial Statistics, Fiscal Year 2011; and Graduation Rates, Selected Cohorts, 2003–2008: First Look (Provisional Data)* (NCES 2012-174 rev). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>.
- Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses: In depth. *EDUCAUSE Review Online*. Retrieved from <http://www.educause.edu/ero/article/retention-and-intention-massive-open-online-courses-depth-0>
- Kotler, P. (1976). Applying marketing theory to college admissions. In *A Role for Marketing in College Admissions* (pp. 54-72). New York: College Entrance Examination Board.
- Koul, A. (October 21, 2012). *Find Distance and Time Between Two Cities Using Google API in Excel*. Retrieved from <http://forum.chandoo.org/threads/find-distance-and-time-between-two-cities-using-google-api-in-excel.7271/>
- Kuh, G. D., & Love, P. (2000). A cultural perspective on student departure. In J. Braxton (Ed.), *Reworking the student departure puzzle* (pp. 196-212). Nashville: Vanderbilt University Press.
- Lani, J. (n.d.). *Assumptions of logistic regression*. Retrieved from <https://www.statisticssolutions.com/academic-solutions/resources/directory-of-statistical-analyses/assumptions-of-logistic-regression/>
- Lani, J. (2010). *Conduct and interpret a point-biserial correlation*. Retrieved from <http://www.statisticssolutions.com/wp-content/uploads/kalins-pdf/singles/point-biserial-correlation.pdf>

- Lee, S. (2010). *The implementation of enrollment management at two public universities experiencing demographic and funding challenges*. (Doctoral dissertation, Indiana University of Pennsylvania, 2010). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3403195).
- Leedy, P. & Ormrod, J. (2001). *Practical research: Planning and design* (7th ed.). Upper Saddle River, NJ: Merrill Prentice Hall. Thousand Oaks: SAGE Publications
- Lenning, O. T. (1982). *Variable-Selection and Measurement Concerns*. In E. T. Pascarella (Ed.) pp. 35-53. *New Directions for Institutional Research: Studying student attrition*, no. 36. San Francisco: Jossey-Bass, December 1982.
- Litten, L. H., Sullivan, D., & Brodigan, D. L. (1983). *Applying market research in college admissions*. New York: College Entrance Examination Board.
- Lorenzetti, J. P. (2002). Before They Drift Away: Two Experts Pool Retention Insights. *Distance Education Report*, 6(8).
- Lovering, N. (n.d.). What is the difference between marketing and advertising? *The Houston Chronicle*. Retrieved from <http://smallbusiness.chron.com/difference-between-marketing-advertising-25047.html>
- Lovitts, B. E., (2001). *Leaving the ivory tower: The causes and consequences of departure from doctoral study*. Lanham, MD: Rowman & Littlefield.
- Lunt, M. (2013). Modeling binary outcomes [Class handout]. Manchester, England: The University of Manchester. Retrieved from http://personalpages.manchester.ac.uk/staff/mark.lunt/stats/7_Binary/text.pdf
- Lynch, S.M. (2003). *Missing data*. Retrieved from <http://www.princeton.edu/~slynch/soc504/missingdata.pdf>

- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1*, 173-181.
- Malocsay, K. W. (2004). *The effects of cultural distance on student socialization and departure decisions*. (Doctoral dissertation, Florida International University, 2004). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3165163).
- Mattox, V. R. (1983). *A study of a means for the early identification of potential college dropouts*. (Doctoral dissertation, Ohio University, 1983). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 8400254).
- McMillan, J.H. (2008). *Educational research fundamentals for the consumer* (5th ed.). Boston: Pearson Education, Inc.
- McNeely, J. H. (1940, January). College student mortality studies. *Journal of the American Association of Collegiate Registrars* Vol 15, 119- 124.
- Menard, S. (1995). Applied Logistic Regression Analysis. *Sage Publications Series: Quantitative Applications in the Social Sciences, No. 106*. Thousand Oaks, CA: SAGE Publications.
- Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. Thousand Oaks, CA: Sage Publications, Inc.
- Milon, R. A. (2006). *The relationship between institutional retention plans and retention rates: A correlational study of independent, public and proprietary institutions of higher learning in New York State*. (Doctoral dissertation, Capella University, 2006). Abstract from ProQuest Dissertations and Theses Database. (AAT 3229497).

- Mooney, S.P., Sherman, M.F., & LoPresto, C.T. (1991, May/June). Academic locus of control, self-esteem, and perceived distance from home as predictors of college adjustment. *Journal of Counseling & Development*, 69, 445-448.
- 445-448.
- Morgan, S. P., & Teachman, J. D. (1988). Logistic regression: Description, examples, and comparisons. *Journal of Marriage and the Family*, 50(4), 929-936.
- Morris, J. F. (1988). *An exploration of factors affecting student retention at an urban university*. (Doctoral dissertation, Temple University, 1988). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 8818823).
- Morrison, L., & Silverman, L. (2012). Retention theories, models, and concepts. In A. Seidman (Ed.) *College student retention 2nd Edition: Formula for student success*. New York. ACE/Rowman & Littlefield.
- National Conference of State Legislatures (NCSL). (2013). [Map illustration of the status of performance funding]. *Performance Funding for Higher Education*. Retrieved from <http://www.ncsl.org/issues-research/educ/performance-funding.aspx>.
- NCSS Statistical Software. (n.d.). *Point biserial correlation tests*. Retrieved from http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/PASS/Point_Biserial_Correlation_Tests.pdf
- Noel-Levitz. (2011). *2011 Cost of recruiting an undergraduate student: Benchmarks for four-year and two-year institutions*. Coralville, Iowa: Author. Retrieved from www.noellevitz.com/BenchmarkReports.
- Noel-Levitz. (2012). *Fall 2012 new student enrollment and retention outcomes at four-*

- year and two-year institutions*. Coralville, Iowa: Noel-Levitz. Retrieved from www.noellevitz.com/BenchmarkReports.
- Nora, A. and Plazas Snyder, B. (2008). Technology and higher education: The impact of e-learning approaches on student academic achievement, perceptions and persistence. *Journal of College Student Retention*, 10(1), 3-19.
- O'Connel, A.A. (2006). *Logistic regression models for ordinal response variables*. Thousand Oaks: Sage. QASS #146.
- Orazem, V. N. (2000). *Understanding why they stay and why they leave: A grounded theory investigation of undecided students at a rural land grant institution*. (Doctoral dissertation, University of Wyoming, 2000). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 9973568).
- Ousley, C. (2010). *A geographic-information-systems-based approach to analysis of characteristics predicting student persistence and graduation*. (Doctoral dissertation, The University of Arizona, 2010). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3403277).
- Pallant, J. (2007). *SPSS Survival Manual*, New York, New York: Open University Press.
- Pantages, T. J., & Creedon, C. F. (1978, Winter). Studies of college attrition: 1950-1975. *Review of Educational Research*, 48, No. 1, 49-101.
- Papa, J. J. (1996). *Academic preparedness as a predictor of undergraduate student retention at liberal arts colleges*. (Doctoral dissertation, University of Maryland College Park, 1996). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 9719802).
- Pascarella, E.T. (1985). College environmental influences on learning and cognitive

- development: A critical review and synthesis. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 1). New York: Agathon.
- Pascarella, E.T., & Terenzini, P.T. (1991). *How college affects students*. San Francisco: Jossey-Bass.
- Pascarella, E.T., & Terenzini, P.T. (2005). *How college affects students: A third decade of research* (2nd ed.). San Francisco: Jossey-Bass.
- Paulsen, M. B. (1990). College choice: Understanding student enrollment. *Association for the Study of Higher Education*. Washington, DC: ASHEERIC Higher Education Reports.
- Peltier, G. L., Laden, R., & Matranga, M. (1999). Student persistence in college: A review of research. *Journal of College Student Retention, 1*(4), 357-375.
- Peng, C. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research, 96*(1), 3-14.
- Peng, C. J., & So, T. H. (2002). Logistic regression analysis and reporting: A primer. *Understanding Statistics, 1*(1), 31-70.
- Pett, M. A. (1997). *Nonparametric Statistics for Health Care Research: Statistics for Small Samples and Unusual Distributions*. Thousand Oaks, CA: Sage Publications, Inc.
- Pommerich, M., Hanson, B. A., Harris, D. J., & Sconing, J. A. (January 2000). Issues in creating and reporting concordance results based on equipercentile methods. *ACT Research Report Series. 2000-1*. Iowa City, IA.
- Prescott, B. T., & Bransberger, P. (2012). *Knocking at the College Door: Projections of*

- High School Graduates* (eighth edition). Boulder, CO: Western Interstate Commission for Higher Education.
- Price, J. L. (1977). *The study of turnover*. Ames: Iowa State University Press.
- Radney, R. (2009). *Predicting first-time freshman persistence at California State University, Bakersfield: Exploring a new model*. (Doctoral dissertation, University of the Pacific, 2009). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3356563).
- Raisman, N. (2002). *Customer service in higher education*. Horsham: PA, LRP Publications.
- Ramist, L. (1981). *College attrition and retention* (College Board Rep. No. 81-1). New York: College Entrance Examination Board.
- Rubin, D. (1976). Inference and missing data. *Biometrika*, 63 (3), 581–592
- Rudestam, K. E., Newton, R. R. (1992). *Surviving your dissertation: A comprehensive guide to content and process*. Thousand Oaks, CA: SAGE Publications.
- Rueda, D. (2014). *Intermediate Social Statistics: Lecture 3*. Personal Collection of D. Rueda, Nuffield College, Oxford University. Retrieved from <http://users.ox.ac.uk/~polf0050/Lecture3.pdf>.
- Sandelowski, M. (1986). The problem of rigor in qualitative research. *Advances in Nursing Science*, 8(3), 27-37.
- Schuh, J. H., & Gansemer-Topf, A. (2012). Finances and retention: Trends and potential Implications. In A. Seidman (Ed.) *College student retention 2nd Edition: Formula for student success*. New York. ACE/Rowman & Littlefield.
- Schwartz, R. A., & Washington, C. M. (1999). Predicting academic success and retention

- for African-American women in college. *Journal of College Student Retention, 1*, 177-191.
- Schwitzer, A. M., & Thomas, C. (1998). Implementation, utilization, and outcomes of a minority freshman peer mentor program at a predominantly white university. *Journal of the Freshman Year Experience, 10*(1), 31-50
- Scopelliti, M., & Tiberio, L. (2010). Homesickness in university students: The role of multiple place attachment. *Environment and Behavior, 42*(3), 335-350.
- Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures*, (5th ed.). Chapman and Hall/CRC, London.
- Simpson, O. (2004). The impact on retention of interventions to support distance learning students. *Open Learning, 19*(1), 79-95.
- Sivak, M., & Schoettle, B. (2012) Update: Percentage of young persons with a driver's license continues to drop. *Traffic Injury Prevention, 13* (4), 341.
- Smith, B. W., Spinelli, J. G., & Zhou, Y. (2002, June). Geographic patterns of student enrollment in Ohio's state-assisted universities. *The Ohio Journal of Science, 102*(3), 34-39.
- Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange, 1*, 64-85.
- Spady, W. G. (1971). Dropouts from higher education: Toward an empirical model. *Interchange, 2*(3), 38-62. Retrieved from <http://link.springer.com/content/pdf/10.1007%2F02282469.pdf>
- State Higher Education Executive Officers Association (SHEEO). (2013). *State Higher Education Finance FY 2012*. Retrieved from <http://www.sheeo.org/node/631>.

- Strang, R. (1937). *Behavior and background of students in college and secondary school*. New York: Harper Brothers.
- Stillman, M. (2007). *A study of factors related to freshman year to sophomore year retention at Southern Oregon University*. (Doctoral dissertation, The University of Nebraska - Lincoln, 2007). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3275063).
- Stork, D., & Berger, P. D. (1978). Attrition in the liberal arts college of a major metropolitan university. *Research in Higher Education, 9*, 281-289.
- Strobe, M., Van Vliet, T., Hewstone, M., & Willis, H. (2002). Homesickness among students in two cultures: Antecedents and consequences. *British Journal of Psychology, 93*, 147-168.
- Summerskill, J. (1962). Dropouts from college. In N. Sanford (Ed.), *The American college* (pp. 627-637). New York: John Wiley and Sons.
- Swail, W. S. (2004). *The art of student retention: A handbook for practitioners and administrators*. Austin, TX: Educational Policy Institute.
- Terenzini, P. T., Pascarella, E. T., Theophilides, C., & Lorang, W. G. (1985). A replication of a path analytic validation of Tinto's theory of college student attrition. *Review of Higher Education, 55*, 621-636.
- Thurber, C.A. (1995). The experience and expression of homesickness in preadolescent and adolescent boys. *Child Development, 66*, 1162-1178.
- Thurber, C.A. (1999). The phenomenology of homesickness in boys. *Journal of Abnormal Child Psychology, 27*, 125-139.
- Thurber, C.A., & Walton, E.A. (2012). Homesickness and adjustment in university

- students. *Journal of American College Health*, 60, 1-5.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45, 89-125.
- Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press.
- Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition*. (2nd edition.) Chicago: University of Chicago Press.
- Tinto, V. (2000). Linking, learning, and leaving: Exploring the role of the college classroom in student departure. In J. M. Braxton (Ed.), *Reworking the Student Departure Puzzle*. Vanderbilt University Press.
- Tufte, E. R. (2003). *The Cognitive Style of PowerPoint*. Cheshire, Connecticut: Graphics Press.
- University of Pittsburgh. (n.d.). Common Data Set 2012-2013. Retrieved from http://www.ir.pitt.edu/cds/documents/CDS2012_13PghCampus.pdf
- University of Pittsburgh at Bradford. (n.d.). Voluntary Accountability Report. Retrieved from <http://www.upb.pitt.edu/uploadedFiles/About/VSA2011.pdf>
- University of Pittsburgh at Bradford. (2007, March). News and Notes. Retrieved from http://www.upb.pitt.edu/templates/Beyond.aspx?menu_id=1368&ID=6561
- University of Pittsburgh at Greensburg. (n.d.). Voluntary Accountability Report. Retrieved from <http://www.greensburg.pitt.edu/sites/default/files/About/2011%20Accountability%20Report%20Greensburg%20REVISED.pdf>
- University of Pittsburgh at Johnstown. (n.d.). Voluntary Accountability Report. Retrieved

from <http://www.upj.pitt.edu/34675.pdf>

U.S. Census Bureau. (2012). Median household income in 2012 by county. Retrieved

from

[https://upload.wikimedia.org/wikipedia/commons/thumb/9/9e/US_county_household_median_income_2012.png/1280px-](https://upload.wikimedia.org/wikipedia/commons/thumb/9/9e/US_county_household_median_income_2012.png/1280px-old_median_income_2012.png)

[US_county_household_median_income_2012.png](https://upload.wikimedia.org/wikipedia/commons/thumb/9/9e/US_county_household_median_income_2012.png/1280px-old_median_income_2012.png)

U.S. Department of Education, National Center for Education Statistics. (2010). IPEDS Data Center. <http://nces.ed.gov/ipeds/datacenter>.

U.S. Department of Education, National Center for Education Statistics. (2011). IPEDS Data Center. <http://nces.ed.gov/ipeds/datacenter>.

U.S. Department of Education, National Center for Education Statistics. (2013). *The Condition of Education 2013* (NCES 2013-037). “Institutional Retention and Graduation Rates for Undergraduate Students.” Retrieved from http://nces.ed.gov/programs/coe/indicator_cva.asp.

Van Gennep, A. (1960) *The Rites of Passage*. (M. B. Vizedom & G. L. Caffee, Trans.). London, Routledge & Kegan Paul. (Original work published in 1908).

Van Tilburg, M. A. L., Vingerhoets, A. J. J. M., & Van Heck, G. L. (1996). Homesickness: A review of the literature. *Psychological Medicine*, 26, 899-912.

Van Tilburg M. A. L., & Vingerhoets, A. J. J. M. (Eds.). (2007). *Psychological aspects of geographical moves. Homesickness and acculturation stress*. Tilburg: Tilburg University Press.

Walke, J. (2010). *Choosing to succeed: An exploration of the relationship between*

- college choice and freshman retention*. (Doctoral dissertation, The College of William and Mary, 2010). Abstract retrieved from ProQuest Dissertations and Theses Database. (AAT 3405470).
- Wheeler, D. J. (2004). *Advanced Topics in Statistical Process Control: The Power of Shewhart's Charts* (2nd ed.). Knoxville, Tennessee: SPC Press.
- Wilcox, S., Estes, T., & Buter, G. (2010, December). *Report #420: Retention and attrition of undergraduates at UC Davis, 2000 to 2006*. Retrieved from www.sariweb.ucdavis.edu.
- West, R. M. (1928). Student mortality, student survival, and student accounting. In E. Hudelson (Ed.), *Problems of college education: Studies in administration, student personnel, curriculum, and instruction*. Minneapolis: University of Minnesota Press.
- Wojciechowski, A., & Palmer, L. B. (2005). Individual student characteristics: Can any be predictors of success in online classes? *Online Journal of Distance Learning Administration*, 8(2). Retrieved from <http://www.westga.edu/%7Edistance/ojdl/summer82/wojciechowski82.htm>
- Zemsky, R., & Oedel, P. (1983). *The structure of college choice*. New York: College Entrance Examination Board.