Lindenwood University
Digital Commons@Lindenwood University

Dissertations | Theses & Dissertations

Fall 9-2018

Understanding the Relationship Between Student Demographic, Attribute, Academic, and Social Integration Factors with Retention

Landon Keefer Adams
Lindenwood University

Follow this and additional works at: https://digitalcommons.lindenwood.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
Adams, Landon Keefer, "Understanding the Relationship Between Student Demographic, Attribute, Academic, and Social Integration Factors with Retention" (2018). *Dissertations*. 87.
https://digitalcommons.lindenwood.edu/dissertations/87

This Dissertation is brought to you for free and open access by the Theses & Dissertations at Digital Commons@Lindenwood University. It has been accepted for inclusion in Dissertations by an authorized administrator of Digital Commons@Lindenwood University. For more information, please contact phuffman@lindenwood.edu.

Understanding the Relationship Between Student Demographic, Attribute, Academic, and Social Integration Factors with Retention

by

Landon Keefer Adams

September 2018

A Dissertation submitted to the Education Faculty of Lindenwood University in partial fulfillment of the requirements for the degree of

Doctor of Education

School of Education

Understanding the Relationship Between Student Demographic,

Attribute, Academic, and Social Integration Factors

with Retention

by

Landon Keefer Adams

This Dissertation has been approved as partial fulfillment

of the requirements for the degree of

Doctor of Education

Lindenwood University, School of Education

Rhonda L Bishop	9-25-18
Dr. Rhonda Bishop, Dissertation Chair	Date
Sherry DeVore	9-25-18
Dr. Sherry DeVore, Committee Member	Date
Thomas H V IV	9/25/18
Dr. Thomas Vernon, Committee Member	Date

Declaration of Originality

I do hereby declare and attest to the fact that this is an original study based solely upon my own scholarly work at Lindenwood University and that I have not submitted it for any other college or university course or degree.

Full Legal Name: Landon Keefer Adams

Signature: Date: 10-1-18

Acknowledgements

I would like to begin by thanking my dissertation committee. Each has played a crucial role in allowing me to take advantage of this enormous opportunity. Thank you Dr. Sherry DeVore, Dr. Alex Vernon, and Dr. Rhonda Bishop for your assistance through this process. Thank you to my Missouri Southern University family who have been incredibly encouraging and supportive. Thank you to Holley Goodnight and Julie Wengert. We began the program together, and we are now at the end of our journey.

Finally, I would like to thank my family. To my parents, Lance and Jill Adams, and my brothers Jansen and Kellen, I cannot describe how blessed I am to have shared a home with the four of you. Additionally, thank you to the rest of my family specifically Jeff, Carol, and Barbara Smith. I am so grateful for all of support you have provided me.

To my daughter, Annabelle, I adore you. I cherish all of the moments I was able to hold you in my arms as I highlighted articles and typed papers. Eventually, those moments turned into you sitting on my lap as I crunched data on my computer. I am so thankful to complete this degree while you were so young. Last, but not least, my biggest thank you is owed to my sweet, supportive, and beautiful wife Katie. I am blessed beyond measure to experience my life with you. Without your encouragement I am not sure I would have ever taken the plunge into this program. When I went class, you provided me support. When I had a stack of articles, you were my cheerleader. When I needed to refocus, you were my motivator. Every step along the way, you were there for me. Now my journey has come to an end, and I know you will be prouder than anyone. I do not know how I came to be so blessed, but I thank the Lord to have you. I am so thrilled to be able to invest my time back into you and our sweet Annabelle.

Abstract

Student retention has been studied more than any higher education subject (Vlanden & Barlow, 2014). Attempts to better understand the retention process through predictive modeling have become more common (Bingham & Solverson, 2016). However, modeling efforts have failed to properly account for elements of social integration and sense of belonging, both of which serve as key tenants in Astin's (1975, 1999) theory of student involvement and Tinto's (1982, 1993) model of college dropout and theory of student departure (Bingham & Solverson, 2016). In this study, social integration was evaluated in isolation using *z*-tests. Several forms of social integration were found to have a statistically significant difference in the proportion of retained participants versus non-participants including campus fitness programs, fraternity or sorority programs, recreation facilities, and student activities. Participants in intramural sports and on-campus living were not found to have statistically significant results. Additionally, binary logistic regression was used to analyze how social integration variables interplayed with demographic, student attribute, and academic performance inputs. The model produced through the analysis successfully met previous goodness-of-fit standards established in prior research (Bingham & Solverson, 2016; Jia & Maloney, 2014). Findings of this research are especially relevant to higher education administrators. A key method to the promotion of persistence and student retention is the ability to predict attrition (Harvey & Luckman, 2014). By including social integration data, higher education leaders could seize upon the opportunity to more accurately identify those students who are less likely to persist than their peers (Bingham & Solverson, 2016).

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

In recent times, there has been an increase in the level of attention universities have received regarding retention and graduation rates (Bingham & Solverson, 2016). The increased focus is the result of several factors including national policies, a shift toward a more customer-oriented mentality in higher education, and concerns about quality of the student experience (Denson & Bowman, 2015). As the student recruitment process becomes more expensive, institutions must retain more students to be sustainable (Ekowo & Palmer, 2017). After all, it is much less expensive to retain a student than recruit a new one (Vlanden & Barlow, 2014).

There has been much research on student retention (Vlanden & Barlow, 2014). A considerable portion of the research has been dedicated to determining which demographic, student attribute, and academic performance indicators are predictive of student retention or departure from an institution (Bingham & Solverson, 2016; Branand, Mashek, Wray-Lake, & Coffey, 2015; Jia & Maloney, 2014). However, it has long been theorized social integration or engagement is critical to student success outcomes such as GPA, retention, and graduation (Astin, 1999; Tinto, 1982, 2006). In this study, various types of a student's social integration were tested to determine if the measurements carry statistically significant predictive qualities. Social integration variables which were studied included student-activity-fee funded events, student organization membership, college athletics, fraternity or sorority membership, honors program membership, student government, recreation programs, recreation facilities, and student residency.

Background of the Study

Throughout the last 45 years, a wealth of research has been dedicated to the topic of college student retention (Belch, Gebel, & Mass, 2001). According to researchers Vlanden and Barlow (2014), student retention has been studied more than any other topic in American higher education. Originally, research on retention was framed from a psychological perspective (Tinto, 2006). At the time, attrition was viewed as a negative reflection of students who did not persist (Tinto, 2006). In the 1970s, the view of retention moved toward a more systematic assessment of why students were not retained (Tinto, 2006). Instead of a focus on why a student failed to persist, attention shifted to why institutions were not retaining students (Tinto, 2006). Despite a large amount of research, substantial improvements in retention rates have been difficult to obtain (Tinto, 2006).

According to Astin (1999), about half of the variability in student retention can be explained by student attributes and demographic data. As such, there is still a large amount of variability, which has not been explained (Astin, 1999). Researchers Bingham and Solverson (2016) also analyzed student attributes and demographic information, but like Astin (1999) were not able to account for a large portion of the variability. Bingham and Solverson (2016) went on to theorize the addition of more variables could boost the variability of existing retention models. It has been posited by many that social integration into a university's experience is a key piece in projecting a student's likelihood to be retained by an institution (Tinto, 1982, 2006; Vlanden & Barlow, 2014). Co-curricular engagement opportunities such as student organizations, living on-campus, or campus recreational activities are also important in projecting student success outcomes like retention and graduation (Tinto, 1993).

Activities on college campuses can vary (Astin, 1999). Belch et al. (2001) found participation in recreation programs provided a positive correlation with student retention. Recreation programs include fitness classes, intramural sports programs, and recreation facility usage (Belch et al., 2001). Recreation programs are an excellent example of how membership in campus groups, such as being a part of an intramural sports team or attending a fitness class with the same group, develops a connectedness to an institution (Branand et al., 2015). A connection to an institution is integral in developing a sense of belonging for a student (Branand et al., 2015; Jacoby, 2015; Masika & Jones, 2016; Tinto, 2017). Concerning recreation facility usage, Belch et al.'s (2001) research determined recreation facility users were retained from the first-year of college to the second-year 71% of the time (p. 261). User retention compares positively to the 64% retention rate of recreation facility non-users (Belch et al., 2001, p. 261). Higher retention rates for recreation facility users could be because recreation facilities provide the opportunity to connect socially with fellow students as well as university faculty and staff (Belch et al., 2001).

Another way students can integrate socially is through membership in groups and organizations (Tinto, 1988). In 2015, Branand et al. found membership in a Greek organization had a positive impact on a student's likelihood of retention. Membership in groups such as fraternities or sororities are an excellent way to build student integration into a campus community (Tinto, 1988). Another way to build relationships is through on-campus living (Tinto, 1988). Bronkema and Bowman (2017) concluded living on-campus is a positive indicator for stronger retention rates. While the development of a sense of belonging and social integration would seem more likely to occur in campus housing, this is a variable which has always been accounted for in retention models (Bingham & Solverson, 2016). Tinto (1993) found participating in student activities increased a student's sense of engagement to a university. In the same work, Tinto (1993) explained social integration to a campus community increased students' likelihood to persist at the institution.

Retention is a complex subject, which requires a complex answer (Belch et al., 2001). Before improvements to an institution's retention rate can be made, it must first be examined what is influencing retention (Bingham & Solverson, 2016). To understand why students chose to remain at an institution, universities have become among the worldwide leaders in the use of predictive analytics and big data (Gill, 2017). By quantifying variables, which demonstrate social integration on a college campus and combining those with student demographic and attribute data, there is the opportunity to add to the understanding of student retention in higher education (Bingham & Solverson, 2016). Analytic system implementation could fundamentally alter the way in which higher education professionals connect with students to meet each student's personal needs (Page & Gehlbach, 2018). An individualized analytic system makes it more likely the student will be proactive in pursuit of his or her own educational interests (Page & Gehlbach, 2018).

Theoretical Framework

Theories chosen for this research were selected because they provide a theoretical basis regarding student retention in higher education and the importance of social integration in the process (Astin, 1975, 1993, 1999; Tinto 1982, 1988, 1993, 2001, 2007, 2017). Gennep's rites of passage theory was formulated in 1960, and served as the basis for Tinto's (1993) theory of student departure. Additionally, Astin (1975, 1999) has been an industry leader in retention and persistence theory.

Institutions of higher education face more scrutiny than ever concerning student retention and graduation metrics (Bingham & Solverson, 2016). Over two dozen states, in some way, now appropriate funds based on enrollment, retention, and/or graduation metrics (Ekowo & Palmer, 2017). Additionally, state appropriations for higher education have only incrementally increased over the last 12 months (Kelderman, 2018). Miniscule appropriation increases, along with the growing portion of those funds being directly tied to performance measures, have caused uncertainty regarding the financial commitment of state governments to higher education (Kelderman, 2018). As such, there has been a developing focus in higher education to increase retention and graduation rates (Tinto, 2006). Before institutions can target strategies for change, it is critical to understand relevant theories regarding student retention so better focus on relationships can be obtained (Jacoby, 2015). Utilization of theories, models, and frameworks are essential in leading practice down a productive path (Jacoby, 2015).

The use of theory is the best way to implement beneficial systems and perform best practice methods in student affairs (Jacoby, 2015). It is important to note, however, theories and models are not the final stage in the process (Jacoby, 2015). Students and learning environments are complex systems, which interact in a countless variety of ways (Jacoby, 2015). Theories and models are a means by which educators can better understand the way systems interface with one another (Jacoby, 2015). In 1988, Tinto theorized the framework of Arnold Van Gennep's rites of passage theory could apply to the process of student departure.

Gennep's (1960) rites of passage theory is an analysis of the way in which relationships develop within groups. Tinto (1988) theorized, similar to the rites of passage theory, there is a process in student persistence and retention which students must advance through. Rites of passage theory includes three stages which are separation, transition, and incorporation (Gennep, 1960). Each of these stages include specific challenges, ceremonies, and rituals (Gennep, 1960). Experiences in Gennep's stages are similar to situations students face as they attempt to integrate themselves with fellow students and the institution (Tinto, 2006). Integration hinges on students engaging with campus and developing a sense of belonging to the institution (Tinto, 2017).

The first stage in Gennep's (1960) rites of passage theory is separation. During this stage, individuals process leaving former social groups and communities in favor of new ones (Tinto, 1988). Leaving social groups can be difficult and disorienting for students (Tinto, 1988). It is essential during the separation stage for students to commit to disassociating themselves physically and socially from past communities so they may successfully integrate into the college community (Tinto, 1988). Students experiencing separation often encounter academic and social difficulties as they attempt to balance the challenges of two communities (Tinto, 1988). During the separation stage, it is critical for institutions to provide timely support and engagement opportunities to students to assist in the separation process (Tinto, 2017).

The second stage in the process is transition, when individuals move from old to new communities (Gennep, 1960). It is important to note, transition from one community to another is not always as obvious as a student moving into a residence hall (Jacoby, 2015). Transition into college provides students with a unique set of opportunities and challenges (Page & Gehlbach, 2018). Some more noticeable challenges include the application and financial aid process, as well as integrating oneself within the institution (Page & Gehlbach, 2018). Transitions also occur socially and academically (Jacoby, 2015). In fact, transitions can be any event or non-event which affects a student's social relationships, life roles, or routines (Jacoby, 2015). Regardless of the type of transition, it must be noted, transitions are challenging, as students have not yet integrated themselves into their new community (Tinto, 1988). If students do not successfully integrate themselves into their new community, failure to integrate can lead to withdrawal from the university (Tinto, 1988). For these reasons, engagement opportunities are critically important during a student's first six weeks of his or her first-year (Tinto, 1988).

The final stage in Gennep's (1960) rites of passage theory is incorporation. During the incorporation stage, individuals embrace new communities, rituals, and routines (Gennep, 1960). While many institutions have implemented orientation programs to expedite the incorporation process, institutions often come up short and fail to provide the long-term integration needed to establish membership in a community (Tinto, 1988). It is paramount student support and engagement opportunities are provided early and often to ensure they are effective (Tinto, 2017). Not all students can make their own social connections, so structured opportunities incorporated into their first-year can be instrumental in providing integrative contacts (Tinto, 2017).

Also, utilized in this study was Astin's (1999) theory of student involvement. Astin (1999), in his theory of student involvement, posited social integration and involvement in the college community fosters positive outcomes and student satisfaction. When this concept is coupled with Tinto's (1982) work, which stated interaction with other students and faculty outside the classroom is especially important in retention outcomes, one can conclude social integration is a key piece in retention analysis.

The level at which a student is integrated within an institution can vary, so researchers must consider the appropriate type of outreach (Tinto, 1988). According to Tinto (1988), the most important time in influencing student retention is a student's first semester of college. The critical nature of a freshman's first semester could be explained by Gennep's (1960) work, which described the importance of social integration when one transitions from one community to another. According to Astin (1999), virtually every type of involvement during a student's first semester of college positively affects student outcomes.

More so, than other points in the educational journey, first-year student success is dependent on the institution rather than the student's self-efficacy (Tinto, 2017). However, according to Astin (1999), depending on the student outcome, a student's level of involvement is sometimes more important than the student's pre-college characteristics. The essential nature of student involvement is why Tinto (1988) advocated universities should consider front-loading institutional action. During the first-year of college, student involvement is vital; as such, university administrators must take social integration and engagement seriously (Tinto, 2006).

University campuses are both academic and social (Tinto, 1988). Students' decisions to persist or withdraw are impacted by the way they are academically and socially integrated into the institution (Tinto, 1982). In the study of retention, it has been determined there is a direct link between student success outcomes and involvement in areas such as academic experience, as well as social and residential life (Carter & Yeo, 2016). Through engagement, students can develop a sense of social belonging, which facilitates other forms of engagement and improves student success outcomes like retention and graduation rates (Tinto, 2017). Often students' sense of belonging can develop through engagement opportunities (Fernandes, Ford, Rayner, & Pretorius, 2017). It is essential for institutions to identify the highest impact engagement opportunities and create time for students to participate in those activities (Astin, 1999). Programs such as fraternities, sororities, campus housing, extracurricular programs, intramural athletics, or campus recreation can provide high impact engagement opportunities (Tinto, 2017)

Engagement is not the factor which ultimately influences student retention, but the perception is engagement and a sense of belonging influence retention (Tinto, 2017). Institutional administrators should strive to develop a sense of belonging in all students (Jacoby, 2015). When administrators are successful in the development of students' sense of belonging, students will demonstrate enhanced motivation to persist, resulting in higher retention rates for the institution (Tinto, 2017). By engaging students and developing an academic and social belonging at an early stage in a student's first-year, learning, persistence, and completion become more likely (Tinto, 2017).

Datasets for established retention models do not include all necessary variables to account for student retention (Bingham & Solverson, 2016). Tinto (1975) proposed both demographic and student attributes, as well as university experiences, are what influence the level in which first-year students integrate socially and academically into the university. With first- to second-year retention rates so critical in projecting long-term student success, as well as retention's role in performance funding, what else could be gained by expanding upon university inputs (Bingham & Solverson, 2016)? By adding a measure of social integration to existing models, opportunity exists for researchers to obtain a better grasp as to how social integration impacts student retention (Bingham & Solverson, 2016).

Statement of the Problem

The Missouri Department of Higher Education (2017) evaluates four-year public institutions on six performance measures. Each institution's leadership selects three measures on which the institution is evaluated (Missouri Department of Higher Education, 2017). These measurements determine the proportion of the institution's performance funding appropriated to the university (Missouri Department of Higher Education, 2017). One of the Missouri Department of Higher Education (2017) performance measures is retention, as defined by the Integrated Postsecondary Data System. Many view retention as the main indicator of student success, as it demonstrates a positive relationship with graduation (Bingham & Solverson, 2016). According to the Integrated Postsecondary Education Data System (IPEDS) glossary, retention is the percentage of first-time, full-time, degree seeking students retained from one fall semester to the subsequent fall semester (IPEDS 2016-17 Glossary, 2017). Due to the stress of performance measures and fear of losing attached funding, colleges and universities face intense pressure to improve student retention and persistence metrics (Bingham & Solverson, 2016). Financial concerns at Missouri colleges and universities were amplified due to a state appropriation increase of less than the consumer price index from 2017 to 2018 (Seltzer, 2018). Between 2017 and 2018, more than one-third of states in the nation faced similar decreased appropriation levels (Seltzer, 2018).

There has been much research to determine how student attributes and demographic profiles influence retention (Bingham & Solverson, 2016). However, there are still elements of the student experience which remain anonymous to retention modeling (Bingham & Solverson, 2016). According to Tinto's (1993) theory of student departure, social integration with the campus community positively affects retention. By adding variables, which capture social integration and a sense of campus belonging, the predictive power of retention modeling could be increased (Bingham & Solverson, 2016). Administrators in higher education must consider whether variables which quantify social integration impact a student's likelihood to be retained (Bingham & Solverson, 2016). Additionally, research needs to determine how social integration variables interact with demographic, student attribute, and academic performance variables in the student retention process (Bingham & Solverson, 2016).

Purpose of the Study

The purpose of this study was to analyze student demographic, attribute, academic performance, and social integration data at a Midwestern four-year public institution. The objective of the analysis was to add to existing knowledge on student retention. Specifically, the researcher attempted to build upon existing retention models by quantifying student social integration using student identification card swipe data. Currently, retention models rely on student demographic, attribute data, and in some cases, surveys to capture social integration (Bingham & Solverson, 2016; Jia & Maloney, 2014). However, authors of retention theories postulate social integration is a key component of the retention process of first-year students (Branand et al., 2015; Tinto, 1975, 1982, 2007). Bridging the gap between those two worlds by integrating the framework of retention theory in the retention modeling process through other quantitative means of social integration tracking was a primary goal. Results of the research could serve to either validate the value of social integration in retention theory or be inconclusive.

The researcher began by determining which social integration inputs are statistically significant in projecting the likelihood of retention for an individual. By identifying which types of student social integrations are predictive, the researcher then utilized regression techniques to develop a model. Variables established in previous retention modelling efforts such as gender, ethnicity, high school grade point average, ACT score, and major were combined those variables with social integration variables theorized to positive impact student success outcomes.

Research questions and hypotheses. The following questions were addressed in this study:

1. Does student participation, minimum one class attended, in campus fitness programs have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H1_0$: There is no statistically significant difference in first-year to second-year retention between those students who participate in campus fitness programs and those who do not participate in campus fitness programs at a Midwestern four-year public institution.

$H1_a$: There is a statistically significant difference in first-year to second-year retention between those students who participate in campus fitness programs and those who do not participate in campus fitness programs at a Midwestern four-year public institution.

2. Does student membership in a fraternity or sorority have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H2_0$: There is no statistically significant difference in first-year to second-year retention between those students who are members of a fraternity or sorority and those who are not at a Midwestern four-year public institution.

$H2_a$: There is a statistically significant difference in first-year to second-year retention between those students who are members of a fraternity or sorority and those who are not at a Midwestern four-year public institution.

3. Does student participation, minimum one intramural event attended, in intramural sports programs have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H3_0$: There is no statistically significant difference in first-year to second-year retention between those students who participate in intramural sports programs and those who do not at a Midwestern four-year public institution.

$H3_a$: There is a statistically significant difference in first-year to second-year retention between those students who participate in intramural sports programs and those who do not at a Midwestern four-year public institution.

4. Does student participation, minimum one check-in, at a university recreational

facility have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H4_0$: There is no statistically significant difference in first-year to second-year retention between those students who utilize university recreational facilities and those who do not at a Midwestern four-year public institution.

$H4_a$: There is a statistically significant difference in first-year to second-year retention between those students who utilize university recreational facilities and those who do not at a Midwestern four-year public institution.

5. Does student housing status, living on-campus or not, have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H5_0$: There is no statistically significant difference in first-year to second-year retention between those students who live on-campus and those who do not at a Midwestern four-year public institution.

$H5_a$: There is a statistically significant difference in first-year to second-year retention between those students who live on-campus and those who do not at a Midwestern four-year public institution.

6. Does student participation, minimum one event attended, in student-activity-fee-funded events have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H6_0$: There is no statistically significant difference in first-year to second-year retention between those students who participate in student-activity-fee-funded events and those who do not at a Midwestern four-year public institution.

$H6_a$: There is a statistically significant difference in first-year to second-year retention between those students who participate in student-activity-fee funded events and those who do not at a Midwestern four-year public institution.

7. Does the inclusion of social integration variables, in association with already established variables, which account for demographics, student attributes, and academic performance, produce a statistically significant model which can be used as an instrument for projecting a student's likelihood to be retained?

$H7_0$: The combination of variables, which account for demographics, student attributes, academic performance, and social integration do not establish a statistically significant model.

$H7_a$: The combination of variables, which account for demographics, student attributes, academic performance, and social integration establish a statistically significant model.

Significance of the Study

There are several reasons why a recent uptick in attention devoted to student retention has occurred (Denson & Bowman, 2015). First, in the world of higher education, resources have been diminishing for some time (Tinto, 2006). Considering costs of recruitment and education of students have become more expensive, it is essential for students to be retained following a successful recruitment (Ekowo & Palmer, 2017). In fact, according to De Freitas et al. (2015), every 50 undergraduate students retained from year two until graduation are worth approximately $2 million dollars to the institution (p. 1177). In addition, affecting finances of institutions is the rise in performance funding measures at the state level (Fain, 2017). As of September 2017, 35 states had implemented performance funding measures, which tie state appropriations to an institution's performance on student outcome criteria (Fain, 2017, p. 1). Several states have implemented retention into some sort of performance funding system (Tinto, 2006). Missouri is one of those states (Missouri Department of Higher Education, 2017).

In addition to obvious financial motivations, there are other reasons why student retention is now more important than ever (Tinto, 2006). Recent research indicates a link between student equity and retention (Denson & Bowman, 2015). Studies have indicated the modern student views himself or herself as a consumer of a service, and, as a result, enters college with certain quality of life expectations (Blumenthal, 2009). With higher education so competitive for students, it is essential for institutions to provide a consumer-oriented approach to meet not just academic needs but also the social needs of the modern student (Denson & Bowman, 2015).

Retention is one way in which institutions measure student achievement for the first-year (Bingham & Solverson, 2016). Retention of first-year students, as a measure of student achievement, is in large part due to first-year retention being so critical in a student's path to graduation (Bingham & Solverson, 2016). In previous retention models, high school grade point average, ACT score, gender, and race, have been identified as significant predictors of student retention (Bingham & Solverson, 2016). However, Bingham and Solverson (2016) speculated some critically predictive variables are missing from previous models projecting student retention. As Tinto theorized in 1982, factors which affect student retention and gauge student-to-student interaction are particularly critical. Tinto's (1982) research is significant because it can bridge the gap between existing models and long held retention theories, which argue social integration is essential in developing a student's sense of belonging and positively impacting retention.

Definition of Key Terms

For the purpose of this study, the following terms are defined:

Loyalty. A student's persistence from his or her first-year through graduation, as well as students' contributions to the institution as an alumnus (Carter & Yeo, 2016).

Persistence. Reflection of a student's motivation to continue to pursue a degree (Tinto, 2017).

Retention. A percentage of first-time, full-time, degree seeking students retained from one fall semester to the subsequent fall semester (IPEDS 2016-17 Glossary, 2017). For colleges and universities, this is a key performance indicator (Bingham & Solverson, 2016).

Sense of belonging. A psychological feeling in which one is a part of, and a valued member of, a college's community (Branand et al., 2015).

Social integration. Engagement with the campus community, which both increases a student's likelihood to be retained and a student's commitment to the institution (Tinto, 1993).

Student engagement. A measurement of the degree to which a student is invested into his or her college experience (Jacoby, 2015). Student engagement captures both the student's involvement in academic and non-academic pieces of the institution (Jacoby, 2015).

Student satisfaction. A measurement to define a student's contentment with a course, program, or university (Carter & Yeo, 2016).

Student success. Positive results such as retention, student satisfaction, persistence, and GPA (Branand et al., 2015; Denson & Bowman, 2013).

Limitations

The following limitations were identified in this study:

Instrument. To collect data on social integration, the study relied on the use of student identification card swipes. As a result, there were likely incidents in which students were permitted to engage in a social integration opportunity without swiping their student identification card. A student without an identification card would result in the student not getting credit for participation in the given event or utilizing the facility on the occasion in which the card was not swiped. However, it should be noted, no student touchpoint was applied to a student's profile unless the touchpoint was gathered through the card swipe system.

Sample demographics. The sample was limited because the researcher only extracted data from one university's student population (Institutional Data, 2018). The institution was predominately commuter-based (Institutional Data, 2018). Demographics of the study were limited by demographics of the sample institution from which deidentified student data were acquired (Institutional Data, 2018).

Sample size. While the sample of participants in the study was adequate, it is possible the quantity of individuals who participate in each social integration opportunity was insufficient to determine the statistical connection, or lack thereof, between those opportunities and student retention (Bluman, 2017). If the sample size is not sufficient, the test for proportions will not yield statistically significant results (Bluman, 2017).

Summary

An exploration into the background of why increasing student retention rates is critical for higher education was provided in this chapter. Additionally, the foundation for a theoretical framework built upon Tinto's (1993) theory of student departure and Gennep's (1960) rites of passage theory was provided. A gap in current research was identified and a problem statement was formulated. Finally, research questions were presented, along with definitions of key terms, and the determination of limitations within the research.

In Chapter Two, further analysis on the theoretical framework of this research is conducted. Furthermore, a literature review is presented on the rationale behind social integration variables, discussed in the Background of Study section of Chapter One. Finally, an analysis on the use of predictive analytics to better understand student retention is offered.

Chapter Two: Review of Literature

Higher education administrators have the ability and responsibility to improve the institution's student retention (Tinto, 1982). However, most institutions, despite paying lip service to the importance of retention, have failed to implement necessary strategies to move the needle in a positive way (Tinto, 2006). Not enough higher education administrators have shown a willingness to invest the required time and finances to improve retention rates (Tinto, 2006). Even with a vast amount of research on the subject, nationwide retention rates of students in higher education has remained stagnant (Vlanden & Barlow, 2014). Predictive analytics could provide an answer to this challenge (De Freitas et al., 2015). Analytics have come to the forefront of higher education over the past 10 to 20 years (De Freitas et al., 2015). Now, universities are taking the lead in the use of big data and predictive analytics (Gill, 2017). Just as colleges and universities invest in campus infrastructure in the form of buildings, they should do the same in predictive analytics (Blumenstyk, 2016).

Chapter Two of this research is dedicated to the review of relevant literature. Chapter Two is divided into four sections. In the first section, further analysis of relevant theory, Tinto's (1993) theory of student departure and Astin's (1999) theory of student involvement, are reviewed. In the second section, an in depth look into the concepts of student engagement, social integration, and sense of belonging are offered. In the third section, the use of predictive analytics in higher education is discussed. In the final section, an evaluation of variables, which should be reviewed within a retention model, is provided.

Theoretical Framework

Higher education as a field has transitioned from one of expansive resources to one of diminishing resources (Tinto, 2006). State revenues across the country have been in decline in recent years (Kelderman, 2018). Budget cuts to higher education have come as the result of state tax cuts, necessitated by the Great Recession (Kelderman, 2018). Indeed, the national funding landscape for higher education is undoubtedly causing worry among higher education professionals (Seltzer, 2018). Despite a national economy, which performed well in 2017, increased tax revenues which resulted from this upward swing did not translate into increased appropriations (Seltzer, 2018). In 2017, state appropriations were just 1.6% higher than in the previous fiscal year (Seltzer, 2018, p. 2). The low 2017 increase was a half percent lower than the consumer price index and was a smaller increase than any of the previous five years (Kelderman, 2018). Digging deeper into those numbers, if California, Florida, and Georgia are removed, the state appropriation increase was just two-tenths of a percent (Kelderman, 2018). Additionally, 19 state governments opted to decrease state appropriations for higher education from the previous year (Kelderman, 2018, p. 3).

When there are cuts, there comes an increased pressure for institutions to improve upon university retention and graduation rates (Tinto, 2006). Despite increased focus, university administrators have struggled to make large gains in student retention (Tinto, 2006). Originally, viewed as a reflection of the student's self-efficacy, retention was thought of as a student's inability to cope with the challenges of higher education (Tinto, 2006). The concept of putting retention on the shoulders of the student rather than the institution, could be interpreted as victim blaming (Tinto, 2006). Now, the focus is on the institution to retain students (Tinto, 2006).

Since the 1970s, when the view of retention flipped from holding students accountable to holding institutions accountable, a variety of theories have surfaced (Denson & Bowman, 2015; Tinto, 2006). Astin's (1999) theory of student involvement accounts for most of the empirical knowledge regarding environmental factors on student success. The theory of student involvement's can be traced to Astin's (1975) work on college dropouts, as well as Tinto's (1982) model of dropout. Both Astin (1999) and Tinto (1982) analyzed the retention process in a longitudinal manner. A longitudinal approach to examining student retention is a more effective and frequently used approach in successful programs (Tinto, 1982). With a longitudinal approach, it was found student involvement is critical during a student's first year of college (Tinto, 2006).

Using the theory of student involvement, higher education administrators can better understand and design programs which foster student success (Astin, 1999). First and foremost, administrators must accept responsibility for student retention (Tinto, 1982). Higher education administrators must question what institutions can do structurally to retain students (Tinto, 1982). Student dropout and retention are both products of the system of higher education (Tinto, 1982).

A student's identity is the result of how the student interacts with his or her environment (Branand et al., 2015). Every higher education institution's environment is different (Denson & Bowman, 2015). It is the combination of the student's environment and personal pre-college characteristics, which merge to determine whether a student will be retained or drop out (Tinto, 1982). Astin's (1999) theory of student involvement also indicated students learn best when they are involved. There is a proportional relationship between learning and student involvement (Branand et al., 2015). When institutional leaders implement new programs or policies, they must remember the success of the program or policy will require an increase in student involvement (Branand et al., 2015). In addition, to improve learning outcomes, studies which have utilized the theory of student involvement have also shown links between involvement and the student's institutional satisfaction (Astin, 1999).

Involvement refers to the amount of energy a student expounds, both physically and psychologically, into the university experience (Astin, 1999). Highly involved students dedicate time to studying, being on campus, participating in extracurricular opportunities, and interacting with fellow students, as well as campus faculty and staff (Astin, 1999). Uninvolved students do not dedicate time to these things (Astin, 1999). As a result, uninvolved students do not develop a sense of belonging (Tinto, 2017). If students' experiences are negative, they will not connect with the university and will ultimately withdraw (Tinto, 1993, 2012). When student experiences are good, the opposite happens, and a sense of belonging is established (Branand et al., 2015). The more groups a student is a part of, the greater chance the student develops meaningful relationships at the institution to further his or her identity development (Branand et al., 2015). Relationships are a critical piece to a strong college experience (Supiano, 2018). When students build identities, they integrate deeper into college communities and become more connected to the university (Branand et al., 2015).

Tinto (2006) explained student integration is the by-product of a strong relationship between the institution and the student. Integration is a critical component of the retention process (Tinto, 1982). Moreover, additional research is needed to better understand how students become engaged, and, in some cases, disengaged (Tinto, 1982). As the understanding of student retention and dropout is refined, higher education leaders should take a moment to consider how much ability they have to influence retention (Tinto, 1982). The most effective retention strategies are those which successfully integrate individuals into both the academic and social life of the institution (Tinto, 1982).

To be involved at an institution, students must invest physical and psychological energy into both academic and social integration (Astin, 1999). According to Tinto (2017), first-year student success is often more the result of a student's early experiences than a reflection of self-motivation. Involvement opportunities present themselves uniquely to every student (Astin, 1999). Involvement manifests itself in different ways and at different times (Astin, 1999). According to retention theory, involvement is most critical during a student's first year of college (Astin, 1999; Tinto, 1982, 2007).

Involvement can manifest itself in a variety of ways, whether a dedication to academics, participation in extracurricular activities, or interactions with university personnel (Astin, 1999). When Tinto (1982) developed the model of dropout, one objective was to enhance knowledge of social interactions within academic and social systems. A student's social integration can be impacted in a variety of ways, but information interaction with fellow students and faculty is particularly influential (Tinto, 1982). When faculty spend more time interacting with students, and the more students interact with one another, the more likely a student will persist to graduation (Tinto, 1982). Informal interactions, academically and socially, are critical to the social and intellectual development process (Tinto, 1982).

Higher education leadership should pay careful attention to structures in place which could be leveraged to foster more interactions between students and faculty (Tinto, 1982). One example is developing a culture in which faculty eat in the same dining venues or utilize the same recreational facilities as students (Tinto, 1982). Utilizing student-centered facilities in a way which increases faculty usage, promotes student and faculty interactions to occur more regularly (Tinto, 1982). It is not just student and faculty interactions which enhance student success outcomes, there is also a positive relationship between those outcomes and student life engagement (Astin, 1999).

A student’s time is the most essential piece in the theory of student involvement (Astin, 1999). The more time faculty can dedicate to students and the more time students can devote to one another, the more successful students are going to be (Tinto, 1982). Administrators must remember, time is a finite resource, and every decision they make impacts a student’s time (Astin, 1999). For this reason, living and working on campus results in higher retention rates among those student populations (Astin, 1999; Bronkema & Bowman, 2017).

Regardless of the retention program developed by an institution, the implementation and management of each program determines its success or failure (Tinto, 1982). Successful retention programs often begin at the point of admission (Tinto, 1982). Additionally, successful retention programs involve a vast complement of university departments and personnel (Tinto, 1982). Institutions must seek to increase the likelihood of retention and graduation for students they serve (Tinto, 1982). The next step for research on this process is to test the theory of student departure (Tinto, 1993) and theory of student involvement (Astin, 1999). Researchers must seek to quantify effects of extracurricular activities related to persistence and retention processes (Astin, 1999).

The higher education process is impactful on students because of the number of ways the process develops the student (Branand et al., 2015). As such, it is essential to remember, student involvement can be achieved in many ways (Astin, 1999). For some students, this is accomplished through academic integration, as in absorption in studies or a connection with faculty (Astin, 1999). For other students, it comes through social integration via participation in extracurricular activities or relationships with other students (Astin, 1999). Higher education personnel are still working to figure out how to translate retention knowledge and theory into stronger retention numbers (Tinto, 2006). What theory has established is, the more involved a student is the higher likelihood the student will achieve academic and personal development (Astin, 1999). However, it should be noted, it is not the theory which is inherently valuable, it is implementation which impacts retention (Tinto, 2006). Finally, in implementation, theory must be translated into actionable items and successful implementation (Tinto, 2006). Student retention is as important now, as ever (Tinto, 2006).

Social Integration and Sense of Belonging

Student retention studies must consider how the campus environment affects student decisions whether to persist or drop out (Tinto, 2017). Student involvement has been widely shown as beneficial to increasing the likelihood a student is retained (Branand et al., 2015). Engaging students in the university community is the most important thing for higher education personnel (Fain, 2016). However, there is less information regarding how to foster student involvement to boost retention and graduation rates (Tinto, 2006). Despite the information gap, student retention theory has shifted to more largely consider the institution's role in student retention (Tinto, 2006). Research is still needed regarding the role of student involvement in the process (Tinto, 2006). Given the role of student involvement, higher education leadership must engage students to view themselves as members of a community and feel they belong (Tinto, 2017).

Involvement matters, and it is essential during the first year of a student's journey in higher education (Tinto, 1988). It is at this point when students are most likely to leave an institution (Tinto, 1988). A student's first semester is of critical importance to the student's chance of graduation (Tinto, 1988). For this reason, higher education leadership should consider strategies which front load student involvement opportunities and institutional action (Tinto, 1988). Some institutions have taken this consideration too far and bundled programs into a few days prior to the start of a semester (Tinto, 1998). Instead, university officials should spread these programs over the first six weeks of the student's first semester (Tinto, 1988).

First-year students have been the topic of abundant student retention research (Belch et al., 2001). Interest in the subject has only intensified since first- to second-year retention rates have increased in use as a performance funding measure (Bingham & Solverson, 2016). With the rise in the use of first- to second-year retention rates as a performance measure, many higher education personnel are wondering how much rates are influenced by institutional factors versus how much is determined by pre-college student characteristics and demographics (Bingham & Solverson, 2016).

Both prior research and anecdotal evidence from student advisors indicated different factors impact retention during a student's first year compared to subsequent years (Tinto, 1988). In the first year of a student's journey, success is less of a result of student self-efficacy than in other years (Tinto, 2017). First-year students also have been found to have more strength-based conversations with faculty and advisors than other students (Soria & Taylor, 2016). Additionally, prior research indicates involvement is critical to student success outcomes, and this is most true in a student's first year (Tinto, 2006).

In a study of student loyalty, researchers Vlanden and Barlow (2014) found students with positive attitudes regarding their institutions were more likely to be retained. It has been found variables representing behavior and attitudes are more predictive of student loyalty than pre-college variables (Vlanden & Barlow, 2014). Additionally, students who form the strongest bonds with college communities have stronger institutional satisfaction (Branand et al., 2015). Student relationships with fellow students, faculty, or staff are most impactful when they grow organically (Supiano, 2018). While organically-formed relationships are the most powerful, it does not mean higher education professionals cannot have an influence in fostering relationship growth (Supiano, 2018). Higher education officials can develop relationships with students in an assortment of ways (Branand et al., 2015). Scheduling events which create an opportunity for students and faculty to interact early in the students' journey is particularly impactful (Supiano, 2018). Additionally, systems which create a culture of quality advisement and mentorship can improve the connection between students and faculty (Supiano, 2018).

When students feel they are recipients of personal attention from university personnel, satisfaction and loyalty have been shown to increase (Vlanden & Barlow, 2014). Through the development of relationships with fellow students, faculty, staff, and the college itself, satisfaction of a student will improve (Branand et al., 2015). These findings provide statistical backing for decades-old retention theory, which indicates the more time faculty devote to students, the more likely students will achieve successful outcomes (Tinto, 1982). Sense of community and belonging is one of the most impactful experiences on a student’s life (Branand et al., 2015). As such, it is essential for institutional leadership to provide students with a variety of opportunities to integrate themselves within the college community (Branand et al, 2015). Thorough integration between a student and the college community results in a higher level of satisfaction for the student (Branand et al., 2015).

There are some common traits among students who drop out from an institution (Denson & Bowman, 2015). Students are often not involved in extracurricular activities and do not feel as though the university met expectations (Denson & Bowman, 2015). Students who persist successfully, transition into the university and make many connections to students, faculty, and staff inside and outside of the classroom (Denson & Bowman, 2015). As a result, persisting students have a higher sense of belonging and are more satisfied with their universities (Denson & Bowman, 2015). While campus life has shown to have a positive correlation with student satisfaction (Carter & Yeo, 2016), there are multiple ways in which students can mesh with the college community to increase satisfaction (Branand et al., 2015).

Tinto's (1993) theory of student departure made clear the importance of academic and social integration within an institution. Students do not have to be equally integrated on both fronts; however, the more integrated the student is the higher chance for retention within the university (Belch et al., 2001; Tinto, 1993). Students have ample opportunity to get involved on a college campus, both academically and socially (Branand et al., 2015). For each way to get involved, there is a unique opportunity for students to develop personal experiences and relationships (Branand et al., 2015).

The level by which a student is committed to a university is directly tied to the level of integration a student has within the university (Vlanden & Barlow, 2014). The more rewarding a student's integration is and the higher level of satisfaction a student feels, the greater the impact on student retention (Branand et al., 2015). Also, of importance are engagements, which occur outside the classroom, and have a greater impact than those which occur inside the classroom (Denson & Bowman, 2015). Given the importance of outside classroom integration, it could be concluded non-academic aspects of the college experience are integral to student success outcomes (Denson & Bowman, 2015). Since universities are particularly important in developing students personally, it is essential those students connect and integrate themselves into the campus community (Branand et al., 2015).

Sense of belonging is described as the way a student feels accepted, valued, and encouraged (Masika & Jones, 2016). Sense of belonging can come from a small group or class, or it can be applied more broadly to encapsulate the entire institution (Tinto, 2017).

Even if the student derives sense of belonging from what could seem as something small from an outsider's perspective, sense of belonging can still facilitate persistence and enhance the student's relationship with the institution (Tinto, 2017). Quite simply, student integration within an institution, regardless of the group, enhances sense of belonging (Branand et al., 2015). Due to this, sense of belonging is often described as a commitment, which binds the student and university (Tinto, 2017). Institutional bond serves to anchor the student when challenges arise in the educational journey (Tinto, 2017). Researchers have found even a moderate level of student involvement produces sense of belonging and is impactful on persistence (Branand et al., 2015).

Students must believe they are important to the university both inside and outside of the classroom to feel as though they belong (Masika & Jones, 2016). To enhance a student's engagement, higher education leaders must carefully improve campus processes and structure to aid in the creation of campus communities (Masika & Jones, 2016). As Astin (1999) described in his theory of student involvement, a student's most precious resource is time. If institutions can improve processes and procedures, students can connect with more communities in less time (Masika & Jones, 2016). When students connect with one group, they are more likely to integrate into other groups (Tinto, 2017). When a student integrates with more groups, the student further develops a sense of belonging, which in turn improves an institution's odds of retaining the student further (Tinto, 2017).

With research indicating a sense of belonging is linked to student retention outcomes, it bares to reason many universities have added a sense of belonging element to program objectives (Masika & Jones, 2016). Institutional leadership would also be wise to consider how academic curricula and teaching enhances a student's sense of belonging to the class, program, and institution (Masika & Jones, 2016). When university personnel value membership of students within the campus community, and emphasize this feeling to students, the students will feel a greater sense of belonging (Tinto, 2017).

There are no pre-college variables which project a student's retention rate as accurately as the extent to which a student is satisfied with his or her institution (Vlanden & Barlow, 2014). Institution satisfaction is influenced most strongly by the level of integration the student achieves academically and socially within the university (Vlanden & Barlow, 2014). Despite the importance of these items, there are elements of the student experience, which are still not being gathered by institutions (Bingham & Solverson, 2016). Work must be done to determine how social integration variables interact with retention in specific institutional settings and conditions (Tinto, 2006). Researchers must continue to improve measurements and analysis of the student experience on retention (Bingham & Solverson, 2016). Additionally, more research is needed to gain a broader understanding of the relationship between variables which foster and inhibit student engagement and sense of belonging (Masika & Jones, 2016).

Predictive Analytics

Success of higher education institutions has been found to be strongly tied to student attrition (Harvey & Luckman, 2014). To retain students, institutions must first prevent attrition (Harvey & Luckman, 2014). To accomplish attrition prevention, universities must find a way to predict which students are the least likely to be retained (Harvey & Luckman, 2014). Predictive analytics provide higher education administrators with the opportunity to forecast outcomes and behaviors of students (Calvert, 2014). Risk and opportunity can be determined regarding students when information is carefully collected, processed, and analyzed (Calvert, 2014). As analytic systems continue to be improved, the systems enable people to make more individualized complex judgements about students (Page & Gehlbach, 2018). The most successful of these systems are those which can handle a variety of needs (Page & Gehlbach, 2018).

Predictive analytics provide institutions with a more effective means to ensure students receive the highest possible quality learning experience (De Freitas et al., 2015). Institutions have already started utilizing these tools through the analysis of demographic and academic data to determine which students are low-risk and which are high-risk (Ekowo & Palmer, 2017). High-quality education, coupled with strong institutional support, has proven to be an effective combination for institutions as they seek to improve student success outcomes such as retention and graduation rates (De Freitas et al., 2015). Analytical tool usage has been a logical next step for higher education as the field shifts toward a more customer, service-oriented model (De Freitas et al., 2015). To satisfy constituents and retain customers, university administrators have implemented predictive analytics as a proactive measure to devote resources to areas and students with the most need (De Freitas et al., 2015).

Looking outside the education box. When attempting to solve challenges of student persistence, one method is to consider the perspective of another discipline (Vlanden & Barlow, 2014). Areas such as marketing, management, or public relations might shed light on a course of action which could improve student loyalty, and, by extension, student retention metrics (Vlanden & Barlow, 2014). Students and parents of the current generation view higher education as a more customer-oriented field than previous generations (Vlanden & Barlow, 2014). Perhaps, by no coincidence, the most meaningful student success management systems have been implemented in a manner which resembles customer relationship systems in the retail industry (Straumsheim, 2017).

Like higher education's counterparts in retail, student success management systems are designed to predict the behavior of students (Straumsheim, 2017). Predictions produced by retail systems are designed to develop relationships with consumers to increase the odds of the customer returning to buy another product (Vlanden & Barlow, 2014). The same concept has been applied to student success management systems, which attempt to build relationships to retain more students (Vlanden & Barlow, 2014). Customer retention cost is considerably lower than the cost of recruiting a new customer or student (Vlanden & Barlow, 2014). Due to the necessity in higher education for administrators to develop a more cost-efficient operation, one can understand why systems which are designed to retain customers would be desirable (Vlanden & Barlow, 2014).

There are several critical steps for higher education administrators to successfully implement a predictive analytics strategy (De Freitas et al., 2015). The first of these steps is to craft a learning analytics strategy which will fit the needs of the institution (De Freitas et al., 2015). One such strategy, population health management, utilizes analytics to disaggregate the overall population into smaller groups (Straumsheim, 2016). Each smaller group would then receive an outreach or service depending on the group's risk profile (Straumsheim, 2016). Whatever retention strategy chosen, it is of the utmost importance for the analytic system to provide early identification of vulnerable students (Márquez-Vera et al., 2016).

The next step is for an organization to develop infrastructure needed to integrate data into the decision-making process of the institution (De Freitas et al., 2015). One type of system, which can impact advising and can take the use of big data down to the student level, is a recommender system (Ekowo & Palmer, 2017). Recommender systems are structures which consider a student's plan of study and develop a roadmap to completion (Ekowo & Palmer, 2017). Recommender systems improve advising and can help a student decide everything from selecting what course is best for next semester's schedule, or helping a student near graduation decide if a different major would make more sense (Ekowo & Palmer, 2017).

The development of a service-focused campus is also critical in big data implementation (De Freitas et al., 2015). Many universities have already adjusted the campus culture toward a more consumer-oriented approach (Straumsheim, 2016). Those organizations which adapt to the changing student and parent mindset will be the most successful in improving retention rates of the institution (Straumsheim, 2016).

Predictive analytics, when implemented correctly, also provide administrators of higher education with a more dynamic perspective of the student learning process (De Freitas et al., 2015). Viewing a student's journey in a longitudinal manner is not a new concept among higher education leaders (Tinto, 2006). Longitudinal student analysis was presented by Tinto in his book, *Leaving College*, which was the first to examine relationships between environment, academic, and social systems of an institution and how they interplay with student retention (Tinto, 2006). Astin's (1975) theory of student involvement also considered student retention from a longitudinal perspective when he determined student environment was significantly influencing student retention.

Once previously identified steps have been completed, predictive analytic implementers can begin to model student behavior (De Freitas et al., 2015). Behavior modeling gets into the basic concept of predictive analytics, knowing more about the customer, or in this case, the student (O'Flaherty & Heavin, 2015). Predictive analytics accomplishes an increased knowledge via information mining from large databases (O'Flaherty & Heavin, 2015). Data mining yields valuable customer information, which enables organizations to be proactive in decision-making (O'Flaherty & Heavin, 2015).

Somewhat surprisingly, some higher education administrators have even praised the use of predictive analytics for making outreach efforts feel more personal, despite the use of big data (Straumsheim, 2017). The perception of personalized outreach is due to advisors, faculty, staff, and counselors, having access to more information about a student, allowing them to better understand the student's personal struggles (Straumsheim, 2017). Ability to predict the future behaviors of students and utilizing those predictions to increase the success of the institution is perhaps the most valuable tool of predictive analytics (O'Flaherty & Heavin, 2015). Even Tinto (2017) has complimented analytical tools to better understand the relationship between student behaviors and student success outcomes.

Another item administrators should consider when implementing predictive analytic systems is how the piece will fit within the institution's dynamics (De Freitas et al., 2015). It is critical for institutional personnel to be diligent in treating data and predictions of the systems in an ethical and appropriate manner (De Freitas et al., 2015).

Ethical data usage includes implementation of protocol and procedures which help to keep the institution in line with ethical standards (De Freitas et al., 2015). Ethical data procedures often include both external and internal review (De Freitas et al., 2015). Institutional stakeholders must think in advance how they will avoid unintended results and negative consequences, which could result from predictive analytic systems (Reed, 2017). These processes should always include a feedback loop, which enables quick communication, as well as effective and timely change when necessary (Reed, 2017).

Use of predictive analytics can vary depending on the institution, but there are many ways in which big data can be implemented in a positive and effective manner (Ekowo & Palmer, 2017). To determine the way in which predictive analytics should fit within the university structure, one must first develop a strategy for analytic implementation (De Freitas et al., 2015). Some purposes of predictive analytics include early warning systems, recommendation systems, adaptive technology, and enrollment management (Ekowo & Palmer, 2017).

One way predictive analytics have been implemented on higher education campuses is with early alert systems (Ekowo & Palmer, 2017). In such analytic systems, students have profiles which include both non-academic and academic data within a system (Ekowo & Palmer, 2017). If student data are triggered due to a certain variable or combination of variables, then outreach or student support is needed (Ekowo & Palmer, 2017). Student outreach could come from anywhere including tutoring, advisor contact, career services, a financial aid counselor, or student engagement personnel (Ekowo & Palmer, 2017). Administrators of schools across the country have determined which variables are warning signs for all types of students (Tinto, 2017). Sometimes systems are as simple as considering grades and attendance of first-year students, while other times, systems incorporate a vast array of institutional factors (Tinto, 2017).

Another method of use is the previously mentioned recommender systems (Ekowo & Palmer, 2017). These systems greatly assist students in developing degree paths, as well as plans of study (Ekowo & Palmer, 2017). These systems have also been referred to as course suggestion engines (Blumenstyk, 2016). Recommender systems can suggest an alternative time for a course, or even for a student to switch majors depending on results of early coursework in the field (Blumenstyk, 2016).

Adaptive teaching tools are also on the rise in higher education (Ekowo & Palmer, 2017). Adaptive tools analyze the way a student utilizes his or her learning environment, then adapts the environment to better suit the student's needs (Ekowo & Palmer, 2017). Adaptive tools also analyze student results for weaknesses and gaps, then provide additional support in those areas to ensure the student is grasping the entirety of a subject (Ekowo & Palmer, 2017). Targeted support approaches fits well with Astin's (1999) theory of student involvement, which emphasizes the importance of faculty's role in the learning design process to improve a student's capacity to learn by crafting a more effective learning environment.

Finally, is the use of predictive analytics in enrollment management (Ekowo & Palmer, 2017). Enrollment management analytics are designed to help admissions personnel and higher education administrators better utilize resources for recruitment purposes (Ekowo & Palmer, 2017). Additionally, these efforts help personnel determine the best manner to distribute financial aid in an impactful way on enrollment metrics (Ekowo & Palmer, 2017). For instance, some research indicates financial aid is more impactful the further a student is into the education journey (Tinto, 1988). Financial aid's importance could mean administrators decide to reserve more financial aid funds toward upper classmen rather than first-year students (Ekowo & Palmer, 2017). With the growing pressure to improve retention and graduation rates, some enrollment managers have also utilized these systems to target traditional-age, high-performing students who are more likely to be retained and graduate (Pike & Graunke, 2014). In the new world of higher education, a long-term view of student performance must be taken during student recruitment (Harvey & Luckman, 2014).

The various methods to utilize predictive analytics are a great way to impact student retention (Bingham & Solverson, 2016). Once a path has been chosen, the institution should develop a system to collect data and student inputs (De Freitas et al., 2015). Once data have been collected, institutions should develop the model and determine implications for the results of the modeling process (De Freitas et al., 2015).

Higher education is one of the nation's leaders in the use of predictive analytics and big data (Gill, 2017). It is estimated, at least 40% of institutions in the United States have tried predictive analytic implementation (Kamenetz, 2016, p. 1). Some universities have taken these attempts further than others (Biemiller, 2017). Institutions such as Arizona State, Georgia State, Michigan, Purdue, South Florida, and Texas have leveraged predictive analytic strategies into early warning systems to identify at-risk students early in the student life cycle (Biemiller, 2017; Gill, 2017; Vlanden & Barlow, 2014). The University of South Florida identifies students at the onset of college who might need additional intervention (Vlanden & Barlow, 2014). Early identification information is supplied to advisors, faculty, and residence hall staff so early intervention can occur on those students' behalf (Vlanden & Barlow, 2014). Some smaller institutions have utilized data to improve retention through the analysis of selection criteria, as well as being more thoughtful regarding which prospective students should be admitted (Biemiller, 2017). Other small universities have used data to improve retention through advising software systems (Biemiller, 2017).

The University of Arizona is another institution which has implemented a strategy of flagging students who demonstrate at-risk behavior (Papandrea, 2017). Flagged students are sent to department heads who can provide additional outreach (Papandrea, 2017). One finding uncovered by the University of Arizona administration is, students who obtained just a C in English 101 are 15% less likely to graduate, even though the grade did not indicate a lower likelihood to be retained (Papandrea, 2017, p. 3). To tackle a subject so complex, Arizona administrators simplified the impact of retention for faculty (Papandrea, 2017). To faculty, administrators explained even two more students retained would boost retention by 4% (Papandrea, 2017).

Thanks to the help of predictive analytics, faculty and staff at the University of Texas at Austin have grown the institution's graduation rate from 51% in 2011 to over 65% in 2017 (Williams, 2017, p. 1). Administrators attribute the university's success to predictive analytics, which has fed into student success programs and allowed the university to focus resources on students who need them the most (Williams, 2017).

The University of Texas features an admission process which allows for automatic acceptance of the top performing students in each Texas school district (Gill, 2017). Automatic acceptance of top performing Texas high school graduates resulted in some students being accepted who were less likely to be retained than other automatically accepted students (Gill, 2017). A barrier between students identified as likely to persist and those who were not was coined by University of Texas administrators as the *persistence gap* (Gill, 2017). In 2011, the University of Texas at Austin implemented a data tracking system which collected dozens of markers on student demographic and academic characteristics (Gill, 2017). Most of the variables in the system are sensitive personal information, which requires firm ethical standards (Gill, 2017). The system works by flagging variables known to correlate with a higher attrition rate (Gill, 2017). A combination of learning analytics and a clear intervention strategy enables the University of Texas to obtain impactful results quickly (Gill, 2017).

When implementing analytic systems, institutions must not stop at the analysis of variable relationships (O'Flaherty & Heavin, 2015). By determining which student characteristics are statistically significant predictors of student dropout vulnerability, the University of Texas provided better, more effective, services (Gill, 2017). The University of Texas's analytic system is an example of the next step in implementation, by utilizing data to make assessments about which students are vulnerable in the future (O'Flaherty & Heavin, 2015). Characteristics, which are predictive of dropout typically include first-generation students, Pell-eligible students, and minority students (Gill, 2017). Texas has not only aimed to move students through tracks quicker, but the university also believes the analytic systems have improved education quality as well (Gill, 2017).

Georgia State University is another institution which has implemented early-warning systems (Biemiller, 2017; Dimeo, 2017). The Georgia State system is an extremely robust analytic model, featuring over 800 possible red flags for over 50,000 students, all updated daily (Dimeo, 2017, p. 3). Georgia State has fully integrated the system with the institution's advising model (Kamenetz, 2016). Prior to implementation, a small advising staff conducted approximately 1,000 meetings with students per year (Dimeo, 2017, p. 3). Since implementation, the advising staff has experienced enormous growth and expansion (Kamenetz, 2016).

In 2016, the department conducted over 51,000 face-to-face meetings with students who were flagged by the analytic system (Kamenetz, 2016). Since implementation, Georgia State's graduation rate has risen over 6% with students graduating, on average, half a semester earlier (Kamenetz, 2016). Georgia State's result illustrates the importance of a commitment to developing system and culture which leverages analytics into the decision-making process (Kamenetz, 2016). Without a plan for how data will be used in high-impact practices, the value of data and predictive analytic systems is substantially lower (Kamenetz, 2016).

When it comes to the use of big data and predictive analytics, it is essential to protect students from pre-conceived judgements (Dimeo, 2017). Unfortunately, there are already many examples of institutions which have not been proactive in developing data-centric policies (Blumenstyk, 2016). Institutions must be vigilant in developing policies to address both privacy and technology concerns (Blumenstyk, 2016). Institutions have an ethical responsibility to share these policies with students so they can better understand how student data are being gathered and for what purpose (Blumenstyk, 2016).

Institutions can mine an incredible wealth of information on students (Blumenstyk, 2016). Using this wealth of information in an ethical way is a complicated task, and there is no perfect solution (Ekowo & Palmer, 2017). Higher education administrators must consider many points (Blumenstyk, 2016). For instance, should students be aware of every decision made about them based on their data (Blumenstyk, 2016)? Also, should institutions be able to boost revenue streams using this data (Blumenstyk, 2016)? Finally, should students have the option to decline the gathering of student-level data (Blumenstyk, 2016)?

A possibility always exists for data to be abused, especially in the case of students who could have opportunities limited based on pre-conceived judgements (Dimeo, 2017). Higher education administrators can guard against data misuse with a variety of methods (Blumenstyk, 2016). First, data should be viewed as shared between the student, institution, and instructor and each party should understand how data will be collected and used, as well as limits of the use (Blumenstyk, 2016). A data education process should include ample transparency, especially in circumstances when analytic systems determine what will happen for a student (Blumenstyk, 2016). All parties should understand, be able to explain why the decision was made, and discuss how it was justified (Blumenstyk, 2016). Finally, there should always be a desire to tweak and improve the analytic system (Blumenstyk, 2016).

In addition to ethical concerns at the student level, there are also concerns at the administration, state, and federal levels (Bingham & Solverson, 2016). With the rise in outcome-based funding measures, there is pressure to impact enrollment of students from low-income or certain demographic backgrounds (Bingham & Solverson, 2016). In some circumstances, these pressures have resulted in changes in the admissions standards of an institution, sometimes limiting student opportunities (Bingham & Solverson, 2016). One community college did not allow students in remedial classes to enroll in a full semester so they would not be considered first-time, full-time students for performance funding purposes (Reed, 2017). While data manipulation improved the institution's retention and graduation rates, it also resulted in retention and graduation rates of students enrolled in remedial courses to drop (Reed, 2017). As a result, the institution's retention and graduation rates rose superficially, but excluding the statistical manipulation, they retained and graduated a lower percentage of students (Reed, 2017). If institutions receive too much pressure to meet state and federal standards to obtain funding, there is a higher likelihood for predictive analytic systems to be used unethically to help institutions manipulate data (Reed, 2017). If one college took a stand against the unethical use of data, but some did not, the ethical institution would be at a competitive disadvantage despite a more ethical treatment of data (Reed, 2017).

In addition to ethical considerations, data analysis must be adapted in a way to model and predict the desired behavior (De Freitas et al., 2015). Most frequently retention studies have used a type of regression technique called logistic regression, which requires a wide array of student inputs to produce a probability of retention (Márquez-Vera et al., 2016). Logistic regression models are common because they can produce probabilities such as completion, retention, and graduation (Calvert, 2014). Other common modeling techniques include decision trees and neural nets, but logistic regression is the most prevalent (Calvert, 2014).

The capabilities of predictive analytic systems have rapidly improved over the last 10 years (De Freitas et al., 2015). Now universities are leading the charge (Gill, 2017). Abilities of analytics have enhanced the capabilities of student support services (De

Freitas et al., 2015). Institutions can now better understand a student's past, present, and future (De Freitas et al., 2015). With analytic opportunities, implementation of early interventions and ways to personalize students' higher education experiences are created (De Freitas et al., 2015). Predictive analytics are now perhaps the most economical way for higher education administrators to ensure actions provide an impact on student success outcomes (De Freitas et al., 2015). Institutions in which analytic systems are implemented in a smart and proactive manner will achieve performance measure objectives more effectively and efficiently than others (Page & Gehlbach, 2018).

Inputs and Variables

Astin (1999) described students as a black box. On one end are various inputs; while on the other end are student success outcomes (Astin, 1999). However, in current iterations, there are missing inputs which could better explain how inputs produce student success outcomes (Astin, 1999). Characteristics within a student provide a large effect on retention rates (Pike & Graunke, 2014). Many variables and inputs influence retention probabilities for students (Jia & Maloney, 2014). Each of these variables either provides a positive impact on student involvement and success outcomes or a negative factor (Astin, 1999).

There have been many studies which determined both student characteristics and institutional factors influence retention (Jia & Maloney, 2014). Pike and Graunke (2014) suggested student characteristics provide a strong influence on retention and graduation rates. By extension, Pike and Graunke (2014) also concluded retention and graduation rates receive a relatively minor influence from institutional characteristics. More research is needed to determine the extent to which individual characteristics, student educational backgrounds, and institutional factors influence student success outcomes like retention and graduation rates (Jia & Maloney, 2014).

Demographics. General demographics such as age, gender, and race have been utilized in previous retention studies (Belch et al., 2001). In recent research, demographic variables such as gender, race, and ethnicity have been found statistically significant in predicting student retention (Bingham & Solverson, 2016). Belch et al (2001) paired demographic and academic performance data in early efforts of student retention modeling. Pairing of demographic and academic data was a strategy replicated by Calvert (2014) using socio-demographic information, which in addition to age and gender, also studied occupational status and geographic location data. Soria and Taylor (2016) used demographic data to account for approximately 3.4% of the variance in student retention outcomes (p. 70). However, when demographic variables have been combined with other variable types, a greater degree of variance has been explained (Bingham & Solverson, 2016; Vlanden & Barlow, 2014). For example, Vlanden and Barlow (2014) combined pre-college characteristics with college attitudes and behaviors. Pre-college variables included in Vlanden and Barlow's (2014) study included sex, race, and college choice rank. The model also included six survey questions designed to capture student attitudes and behaviors including frequency of student engagement, intent to leave, satisfaction, institutional fit, initial impressions, and perceived skill development (Vlanden & Barlow, 2014).

Prior research on student retention has also examined the interplay of demographic variables (Tinto, 1982). Race and gender, for instance, are particularly illuminating when paired together (Tinto, 1982). For instance, African American females, African American males, white females, and white males all have different attainment processes (Tinto, 1982). Attainment process variance means, when possible, the interrelations of gender and race should be considered while studying student retention (Tinto, 1982). To account for this, some researchers have also examined descriptive statistics of student demographic variables (Vlanden & Barlow, 2014). Demographic and academic performance variables are the standard starting point in retention modeling efforts (Fain, 2016). Numerous studies have analyzed student demographics and demographic effect on retention (Jia & Maloney, 2014).

Age. Retention rates can vary depending on a student's age (Jia & Maloney, 2014). Students age 20 and 21 have a stronger chance of course completion than students of other ages (Jia & Maloney, 2014, p. 141). Students who begin college at 20 instead of 18 are 2.4% more likely to complete, while 21-year-old students are 2.9% more likely to complete (Jia & Maloney, 2014, p. 147). Students who are 25 years old and over are less likely to complete degrees than younger counterparts (Jia & Maloney, 2014, p. 141). Additionally, researchers have also found students who take at least one year off in between high school and college are more likely to be retained than students who enter higher education immediately (De Freitas et al., 2015).

Gender. Vlanden and Barlow (2014) found gender a statistically significant predictor of student loyalty. Vlanden and Barlow's (2014) finding collaborated Astin's (1975, 1997) findings, which indicated gender to be statistically significant even when used as a stand-alone variable in retention projection. However, other researchers have found gender only significant in projecting student retention when interfaced with the student's race (Bingham & Solverson, 2016). Females have a higher likelihood of completing courses by 2.7% compared to males (Jia & Maloney, 2014, p. 141). Despite a higher likelihood to complete courses, Jia and Maloney (2014) did not find females have a higher chance to be retained than males.

Race. In addition to sex, Vlanden and Barlow (2014) also found race to be a predictor of student loyalty within an institution. Researchers should be conscious of the way gender and race can interplay with one another in retention models (Tinto, 1982). When it comes to retention efforts for individuals from minority backgrounds, it is critical for institutions to provide orientation opportunities to integrate students within the campus community (Tinto, 1982). Students of minority backgrounds who possess or develop social skills at an early stage in the university experience are more likely to have a successful journey through higher education (Tinto, 1982).

Ethnicity. Ethnicity has been utilized in previous student retention studies (Belch et al., 2001). Ethnicity has been found to have a major influence on whether a student completes a given course (Jia & Maloney, 2014). Márquez-Vera et al. (2016) found ethnicity, along with course program and course block to be the three most critical variables in determining if a student was successful.

Student attributes. When student retention became a popular item for research 40 years ago, many believed it was a by-product of an individual's attributes and motivation (Tinto, 2006). While this view shifted to include other factors, student attributes still play a key role in student success outcomes (Tinto, 2006). Characteristics such as a student's enrollment status, financial situation, and program of study have been shown to be predictive of certain outcomes (Tinto, 1982, 2006). Additionally, whether a student is traditional or non-traditional affects both retention and graduation outcomes (Pike & Graunke, 2014).

Enrollment status. Student attributes have been established as reflective of student retention (Tinto, 2006). One attribute found to be predictive was enrollment status (Fain, 2016). Enrollment status reflects whether students are attending full-time or part-time and is one of the standard variables institutions use to predict student success outcomes (Fain, 2016). Jia and Maloney (2014) found attending part-time was detrimental to course completion. In fact, students enrolled part-time were found to be 17.2% less likely to complete a given course (Jia & Maloney, 2014, p. 141). Part-time students have also been found less likely to be retained (Calvert, 2014; Jia & Maloney, 2014). One study found full-time students were retained 70% of the time, while part-time students were retained at a rate of 57% (Calvert, 2014).

Finances. Finances are another factor which influence student retention outcomes (Tinto, 1982). However, not enough emphasis has been placed on this variable in previous retention models (Tinto, 1982). Research has shown students from poor socio-economic backgrounds have the lowest rates of course completion and retention (Jia & Maloney, 2014). Financial factors can clearly inhibit student success outcomes, as students eligible for federal grants have also been found to be retained at a lower rate than students who are not eligible (Pike & Graunke, 2014).

Domestic status. Another input which can influence the likelihood of a student's retention is domestic status (Jia & Maloney, 2014). International students have a higher chance to be retained than domestic students (De Freitas et al., 2015). Jia and Maloney

(2014) found international students were retained 4.4% more often than domestic students were (p. 141).

Program of study. Program of study also plays a central role in retention outcomes (Jia & Maloney, 2014). A student's personal educational and career objectives are factors which impact retention (Calvert, 2014). Whether the student has chosen a program for personal or career reasons has implications on student retention (Calvert, 2014). One study found program of study to be one of the three most critical factors in projecting the likelihood a student is retained (Márquez-Vera et al., 2016).

Traditional or non-traditional. The final student attribute to be included within the research is whether a student is traditional or non-traditional (Pike & Graunke, 2014). Pike and Graunke (2014) found an inverse relationship between the size of an institution's non-traditional student population and the institution's retention rate. Non-traditional student population size has been found to not only be impactful at the institutional level but also the cohort level (Pike & Graunke, 2014). Due to the negative impact of a large non-traditional student population, some researchers have recommended institutions focus recruitment on traditional-age, high-ability students (Pike & Graunke, 2014).

Academic performance. A student's satisfaction with his or her academic performance results in a student who is more likely to be retained at an institution (De Freitas et al., 2015). Additionally, Astin's (1999) research indicated past academic achievement in high school and college were predictive of student success outcomes. For this reason, most universities utilize grade point average and standardized test scores to predict success among students (Fain, 2016). In fact, grade point average, standardized test scores, student enrollment status, demographics, and students' academic standing are the most common inputs higher education administrators have utilized in student success predictions (Fain, 2016). Numerous studies have found elements of academic performance are positively related to retention and graduations rates (Jia & Maloney, 2014; Pike & Graunke, 2014). Students who have proven to be of high-ability are ideal students for institutions to recruit in a modern environment of federal and state accountability systems, which rate institutions based from retention and graduation metrics (Pike & Graunke, 2014).

Credit hours. Credit hours taken in a student's first and second semesters have been found to contribute to student success outcomes (Branand et al., 2015). When combined with honors program membership, and contact between students and faculty, the number of credit hours completed by a student has been shown to offer validity in retention models (Branand et al., 2015). For these reasons, some analytic systems aim to maximize the number of credit hours taken during the first year of college (Straumsheim, 2017).

First-year grade point average. A student's grade point average is strongly linked with the student's likelihood to persist and ultimately graduate (De Freitas et al., 2015). The grade point average and retention relationship are especially strong during a student's first year (Harvey & Luckman, 2014). The strongest correlation to be found by one research team between first-year grade point average and student retention was for students pursuing Bachelor of Arts degrees (Harvey & Luckman, 2014). The same research team found first-year grade point average to be one of the two primary factors in student retention at the institution studied (Harvey & Luckman, 2014). First-year grade point average carried more predictive value than any demographic or academic performance variable (Harvey & Luckman, 2014). Administrators at another institution found first-year students who did not get an A or B in a course in their program of study only graduated 25% of the time, while high-achieving peers were graduating at approximately 75% (Kamenetz, 2016, p. 3). Research has clearly shown a strong relationship between first-year grade point average and student retention (Márquez-Vera et al., 2016).

High school grade point average. In addition to first-year grade point average, high school grade point average is predictive of student success (Belch et al., 2001). High school grade point average was confirmed as a statistically significant predictor of student success outcomes by Bingham and Solverson (2016), as was student performance on standardized tests such as the ACT or SAT. Some predictive analytic systems have gone so far as to dictate alternate paths for students when a low grade is achieved for certain courses (Straumsheim, 2017).

Standardized tests. Pike and Graunke (2014) found ACT scores, as well as composite ACT scores to be positively related to retention rates. Standardized tests, which examine literacy or numeracy levels among a student population, have been used in retention modeling efforts (Jia & Maloney, 2014). Students with better results on national exams have a lower risk of course non-completion and are more likely to be retained into the second year (Jia & Maloney, 2014).

Tutoring. One reason students are not successful academically is due to a lack of necessary prerequisite skills (Copus & McKinney, 2016). A current national trend focuses on how to provide more opportunities for higher achieving students (Copus &

McKinney, 2016). However, challenges faced by less prepared, or low achieving, students are often overlooked (Copus & McKinney, 2016). Retention studies have found a positive relationship between tutoring sessions attended by a student and student retention (Bingham & Solverson, 2016). Frequent interactions between tutors and students are key for improving outcomes for those students who might otherwise fall behind (Copus & McKinney, 2016).

Social integration. The best retention programs involve a vast array of stakeholders (Tinto, 1982). According to Tinto, it is the meshing of institutional efforts and demographics which influence if a student persists at an institution or drops out (Tinto, 1975). Previous retention research has yielded statistically significant findings (Bingham & Solverson, 2016). However, with a limited number of variables available, there has not been sufficient opportunity to observe how social and climate characteristics interplay with student retention (Bingham & Solverson, 2016). With more variables examined, the predictive power of retention models could be increased (Bingham & Solverson, 2016). By adding variables, which account for Tinto's key tenets of retention theory regarding how the student and campus environment interact, models could be improved (Bingham & Solverson, 2016).

In 2006, Tinto's theory of student integration elaborated on his previous works and explained retention is the result of the relationship between the student and the institution. The more successful an institution is in integrating a student into the fabric of the university, the more likely the student will be retained at the institution (Belch et al., 2001). Establishment of a student's sense of belonging is a key component of student retention (Belch et al., 2001). A clear relationship exists between the social constructs of an institution, student habits, and the performance of the student (Márquez-Vera et al., 2016).

Campus activities. Tinto (1993) theorized two of the most critical ways for students to develop social integration were activities and campus organizations. Research has shown involvement in campus activities leads to higher rates of student satisfaction with the institution (Belch et al., 2001). Additionally, longitudinal studies have found involvement in activities to significantly impact retention in a positive manner (Belch et al., 2001). Some researchers have referred to this effect as *psychosocial integration* (Branand et al., 2015). Participation in student extracurricular activities is one measure of psychosocial integration (Branand et al., 2015).

Student organization membership. Another measure of psychosocial integration is student organization membership (Branand et al., 2015). General membership in groups has been found to have a positive impact on developing a sense of belonging and integration into the social element of an institution (Astin, 1999). Branand et al. (2015) verified these findings regarding student organization membership. Branand et al. (2015) also found grade point average positively correlated with organization membership. Organization membership positively tying to grade point average has been shown true for graduate students (Branand et al., 2015). If grade point average is a statistically significant predictor of student success (Harvey & Luckman, 2014), then campus organization membership could also be a statistically significant predictor of retention outcomes (Tinto, 1993).

The more a student is involved in student organizations, the more opportunities to interact with faculty, staff, and fellow students are available (Astin, 1999). Participation in student groups develops a stronger sense of community for students (Branand et al., 2015). The more groups a student is involved with, the more opportunities for the student to learn about campus resources, interact with unique perspectives, and develop relationships (Branand et al., 2015). Relationships a student develops leads to a stronger sense of belonging at the institution (Branand et al., 2015). A student's relationships are critical even if a student does not feel comfortable with the entire institution; the student then has a greater opportunity to develop a connection with a group of friends or student organization (Denson & Bowman, 2015).

College athletics. Participation on a collegiate athletic team has an exceptionally high impact on predicting retention outcomes (Astin, 1999). Student athletes have also been found to have higher levels of satisfaction with an institution's academic reputation, friendships, and administration (Astin, 1999). Student satisfaction in an institution increases the likelihood of student success outcomes, mainly retention and graduation (Branand et al., 2015).

Fraternity or sorority membership. Fraternities and sororities also provide students with the opportunity to develop socially with other students at the institution (Tinto, 1988). Opportunities to build social relationships with fellow students can lead to integration within the institution's community (Tinto, 1988). Predictability of retention outcomes is also present in fraternity or sorority membership (Branand et al., 2015). Additionally, Astin (1999) found not only do students involved in Greek life benefit from the increased social bonds and sense of belonging which comes from being in the fraternity or sorority, they are also more likely to participate in other extracurricular activities. Other extracurricular activities can also serve to boost students' likelihood to achieve positive student success outcomes (Astin, 1999).

Honors program. Branand et al. (2015) and Astin (1999) both found membership within honors programs was impactful on developing social integration between the student and institution. Researchers have found students who are members of an institution's honors program have more self-esteem and interpersonal self-esteem (Astin, 1999). In part, due to higher levels of social and intellectual self-esteem, students in honors programs are more likely to integrate themselves into the campus community (Astin, 1999). Honors program students are more likely than non-honors program students to be retained by an institution and ultimately graduate (Astin, 1999).

Student government. Similar findings have also found student government membership to be positively linked to student satisfaction and social integration (Astin, 1999; Branand et al., 2015). In addition to positive links in student satisfaction and social integration, students involved in student government experience an increase in political liberalism, artistic needs, and a greater satisfaction with campus relationships (Astin, 1999). A stronger sense of satisfaction with peer relationships develops because of better developed social skills and more frequent interaction with the institution's faculty, staff, and students (Astin, 1999).

Recreation facilities. Researchers have found students view recreational facilities as focal points to the student social experience (Belch et al., 2001). Recreation facilities can serve as a hub for both social and academic interactions for students (Belch et al., 2001). Recreation facilities provide first-year students with a place where they can develop a sense of belonging and a means to connect with the campus community (Belch et al., 2001). Through connections developed in recreation facilities, students can sometimes integrate at institutions they otherwise might have felt overwhelmed in (Belch et al., 2001). While it is important for students to perform well academically, it is also critical to student success outcomes they engage socially (Carter & Yeo, 2016). Recreation facility usage is one opportunity for this (Carter & Yeo, 2016). It is clear, usage of recreation facilities grant first-year students with enormous opportunity to interact with other faculty, staff, and students, leading to greater student satisfaction (Belch et al., 2001).

The results of Belch et al.'s (2001) research illustrated a large difference in retention rates of recreation facility users and non-users. In the study, first-year recreation facility users were retained 71% of the time from fall-to-fall, while just 64% of non-users were retained at the institution (Belch et al., 2001, p. 261). Retention numbers were higher for users than non-users even when various demographic sub-populations were analyzed (Belch et al., 2001). Additionally, recreation facility users achieved higher grade point averages and earned more credit hours during their critical first-years (Belch et al., 2001). Additionally, recreation facility users with lower than average grade point averages were more likely to be retained than low achieving students who did not utilize the recreation facility (Belch et al., 2001).

Recreation programs. Tinto (1993) established a relationship between student involvement in campus recreation programs and student success outcomes. Among these programs, intramural sports participation, fitness class attendance, and recreational facility usage have been included as variables, which might correspond with student retention (Astin, 1999; Belch et al., 2001). Astin (1999) made a clear connection between student participation in recreation programs such as fitness classes and intramural sports to student satisfaction and graduation. Astin (1999) also found participation in recreation programs has a positive relationship to students' physical health, alcohol consumption, institutional satisfaction, and other student success outcomes.

Tinto (1993) posited recreational activities such as intramural sports and fitness classes are important factors in getting students academically and socially engaged. In one study, participation in intramural sports along with a relationship with faculty were found to be the two most critical factors in predicting retention outcomes (Belch et al., 2001). Intramural sports programs provide students with an opportunity to make friends, form groups, and find study partners (Belch et al., 2001). Intramural sports participation has been considered another measure of psychosocial integration (Branand et al., 2015).

On-campus employment. Astin (1999) called holding an on-campus job one of the most interesting things which affect retention. On-campus employment is interesting because while the assumption is on-campus employment takes away from a student's academic efforts, retention studies have found on-campus employment has a positive relationship with student retention (Astin, 1999). Astin (1999) theorized working on campus creates more opportunities for students to interact with faculty, staff, and fellow students, thus integrating the student into the campus community. A greater sense of attachment to the institution could stem from a psychological element, as the work also serves as the student's source of income (Astin, 1999). A clear relationship exists between on-campus employment, which is often limited to an average of four hours per day, and retention (Márquez-Vera et al., 2016).

Residency. Living on campus, or residency status, is another form of psychosocial integration (Branand et al., 2015). Living on-campus is one variable found to be significant in contributing to student retention (Belch et al., 2001; Bronkema & Bowman, 2017). One study found residency status is the most critical variable in predicting retention outcomes (Astin, 1999). Astin's (1999) study also found residency status as a positive influence on a student's retention regardless of the student's sex, race, or academic ability (Astin, 1999). Additionally, students who live in a residence hall are more likely to graduate than students who live off campus (Bronkema & Bowman, 2017).

Like working on campus, living on campus increases the opportunity for students to develop relationships which integrate the student socially into the community (Astin, 1999). A student who lives on campus has more opportunity to participate in campus life (Astin, 1999). As is the case with receiving a paycheck from the institution, counting on the institution for shelter and meals likely increases students' sense of attachment to the school (Astin, 1999). When a student eats, sleep, studies, and spends free time on the college campus, he or she will develop a stronger identification with the institution (Astin, 1999).

It is worth considering the relationship between residency status and retention could be indirect in nature (Bronkema & Bowman, 2017). As the result of living on campus, students are more integrated and involved in student life opportunities, which results in higher retention (Bronkema & Bowman, 2017). Researchers have found a positive link between living on campus and participation in other student life programs (Carter & Yeo, 2016). Residency and student involvement also results in higher student satisfaction (Carter & Yeo, 2016).

There is good reason to believe being a part of a residence hall community improves personal development of the student (Manata, DeAngelis, Paik, & Miller, 2017). Residence hall students have been shown to become more liberal, artistic, and develop more interpersonal self-esteem (Astin, 1999). Residence hall students also develop more as leaders than commuter peers (Astin, 1999). Leadership development is in part the result of stronger relationships with students, and faculty, as well as a higher satisfaction with the institution, friendships, and the institution's campus life (Astin, 1999). Living on campus results in a higher likelihood for students to join student government or Greek life, thus giving them more opportunities to develop as leaders than commuter students (Astin, 1999).

Residence hall design has also been found impactful on student satisfaction (Bronkema & Bowman, 2017). Students living in suite-style rooms are found to have a lower feeling of community than students living in traditional housing (Bronkema & Bowman, 2017). However, lower achieving students with low grade point averages are often distracted from schoolwork in traditional units due to increased social opportunities (Bowman & Bronkema, 2017). Despite differences in residence hall design related to sense of belonging and community, there has not been a statistically significant difference between students in various housing designs and retention (Bronkema & Bowman, 2017).

When it comes to the impact of on-campus living on student success outcomes, the effect is most pronounced among first-year students (Bronkema & Bowman, 2017). Some institutions have attempted to extrapolate this positive relationship with the implementation of all first-year student dorms (Soria & Taylor, 2016). These housing arrangements have resulted in higher student satisfaction, grade point averages, and retention (Bronkema & Bowman, 2017).

Summary

Recruitment of students should not end upon enrollment (Harvey & Luckman, 2014). Institutions should instead provide ongoing support and do all they can to increase the likelihood of student success (Tinto, 1982). To this point, retention models have failed to account for all of the variance (Bingham & Solverson, 2016). Future research must attempt to account for more variables, which could increase variance of retention models (Pike & Graunke, 2014). By including variables which are designed to account for social integration within the campus environment, the known variance in student retention could be increased (Bingham & Solverson, 2016).

Theories of student departure and involvement have established academically and social engaged students are more likely to achieve student success outcomes, including retention (Astin, 1999; Tinto, 1982, 1993). Numerous studies, both quantitative and qualitative, have found campus engagement to be positively linked with retention outcomes (Bingham & Solverson, 2016; Branand et al., 2015; Fain, 2016; Tinto, 2017; Vlanden & Barlow, 2014). Higher education's success depends on the ability to retain and graduate students (Harvey & Luckman, 2014).

Institutions have turned to predictive analytics to identify and target at-risk students (Gill 2017; Kamenetz, 2016). However, models are currently lacking a full accounting of variables, which would explain a larger portion of variance than current modeling efforts (Bingham & Solverson, 2016). Many variables could be combined to account for variations in student retention, and as many as possible should be explored (Jia & Maloney, 2014).

In Chapter Three, a discussion is provided regarding the research design and methodology of this study. The population and sample are identified. Research questions and hypotheses are also provided. Study limitations are discussed, and the data analysis protocol is outlined.

Chapter Three: Methodology

The intent of this research study was to determine whether effects of social engagement during a student's first year of higher education had a statistically significant effect on the student's likelihood to be retained at the institution. Chapter Three begins with a summary of the problem and purpose of this research, as well as specific research questions and hypotheses addressed in the study. For the study, a quantitative approach was used as the research design. Also included in this chapter is discussion regarding the population and sample, instrumentation, data collection, data analysis procedure, and ethical considerations for this research.

Problem and Purpose Overview

Institutions of higher education face an intense amount of pressure to increase student retention and graduation rates (Bingham & Solverson, 2016). These metrics are common in accountability measures at the state and federal level (Pike & Graunke, 2014). In part due to these external pressures, there has been considerable analysis of institutional retention rates (Pike & Graunke, 2014). However, these studies have failed to provide a comprehensive analysis of student retention (Pike & Graunke, 2014). Quite clearly, there are still variables missing from analysis, which could better explain higher education retention (Pike & Graunke, 2014).

The most widely known student retention research is Tinto's 1975 theory on the subject (Márquez-Vera et al., 2016). Tinto's (1982) model of departure attempted to provide an understanding of the impact an institution has on a student's decision to drop out or persist. Researchers in past studies have indicated the more a student integrates into the fabric of an institution, the more likely the student will be retained and persist to graduation (Belch et al., 2001). Few studies on student retention have focused on institutional effects to retention rates (Pike & Graunke, 2014). A lack of research on institutional effects could in part explain why former attempts at developing predictive models have failed to completely explain the variance in higher education retention rates (Pike & Graunke, 2014). A baseline already exists in student retention modeling efforts, which establish student demographic, attributes, and academic performance variables as predictive of student retention (Bingham & Solverson, 2016; Márquez-Vera et al., 2016; Pike & Graunke, 2014). The intent of this research study was to provide additional data to bridge the gap between student modeling efforts and retention theory, which establishes social engagement as a critical component of student retention (Astin, 1975, 1993, 1999; Tinto, 1982, 1988, 1993, 2001, 2007, 2017).

Research questions and hypotheses. Through the research, the following questions were answered:

1. Does student participation, minimum one class attended, in campus fitness programs have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H1_0$: There is no statistically significant difference in first-year to second-year retention between those students who participate in campus fitness programs and those who do not participate in campus fitness programs at a Midwestern four-year public institution.

$H1_a$: There is a statistically significant difference in first-year to second-year retention between those students who participate in campus fitness programs and those who do not participate in campus fitness programs at a Midwestern four-year public institution.

2. Does student membership in a fraternity or sorority have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H2_0$: There is no statistically significant difference in first-year to second-year retention between those students who are members of a fraternity or sorority and those who are not at a Midwestern four-year public institution.

$H2_a$: There is a statistically significant difference in first-year to second-year retention between those students who are members of a fraternity or sorority and those who are not at a Midwestern four-year public institution.

3. Does student participation, minimum one intramural event attended, in intramural sports programs have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H3_0$: There is no statistically significant difference in first-year to second-year retention between those students who participate in intramural sports programs and those who do not at a Midwestern four-year public institution.

$H3_a$: There is a statistically significant difference in first-year to second-year retention between those students who participate in intramural sports programs and those who do not at a Midwestern four-year public institution.

4. Does student participation, minimum one check-in, at a university recreational facility have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H4_0$: There is no statistically significant difference in first-year to second-year retention between those students who utilize university recreational facilities and those who do not at a Midwestern four-year public institution.

$H4_a$: There is a statistically significant difference in first-year to second-year retention between those students who utilize university recreational facilities and those who do not at a Midwestern four-year public institution.

5. Does student housing status, living on-campus or not, have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H5_0$: There is no statistically significant difference in first-year to second-year retention between those students who live on-campus and those who do not at a Midwestern four-year public institution.

$H5_a$: There is a statistically significant difference in first-year to second-year retention between those students who live on-campus and those who do not at a Midwestern four-year public institution.

6. Does student participation, minimum one event attended, in student-activity-fee-funded events have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

$H6_0$: There is no statistically significant difference in first-year to second-year retention between those students who participate in student-activity-fee-funded events and those who do not at a Midwestern four-year public institution.

$H6_a$: There is a statistically significant difference in first-year to second-year retention between those students who participate in student-activity-fee funded events and those who do not at a Midwestern four-year public institution.

7. Does the inclusion of social integration variables, in association with already established variables, which account for demographics, student attributes, and academic performance, produce a statistically significant model which can be used as an instrument for projecting a student's likelihood to be retained?

$H7_0$: The combination of variables, which account for demographics, student attributes, academic performance, and social integration do not establish a statistically significant model.

$H7_a$: The combination of variables, which account for demographics, student attributes, academic performance, and social integration establish a statistically significant model.

Research Design

Several research approaches including quantitative, qualitative, and mixed methods were considered for this study. Ultimately, a quantitative approach was selected for several reasons. First, a quantitative method is ideal for studies examining variables to test a hypothesis (Creswell & Creswell, 2018). In quantitative design, the researcher attempts to find patterns or relationships between variables (Leavy, 2017). Additionally, since the data sample was large, a quantitative design was preferable (Leavy, 2017). A quantitative approach was appropriate for the study because an already developed theory was being tested (Creswell & Creswell, 2018). While a qualitative design would have been more appropriate for theory development, a quantitative design was ideal because the researcher was attempting to validate or disprove an existing theory (Leavy, 2017). In this case, student retention theory already existed and suggested social integration improves student commitment to an institution (Tinto, 1993).

In Chapter Two of this study, the use of statistical analysis in student retention and persistence was established. As is recommended for quantitative research, within the literature review, a basis for furthering the need to explore research questions and hypotheses was established (Creswell & Creswell, 2018). Demographic and student attribute variables were noted as key pieces in the analysis of student retention. Ground work was laid as to which social integration variables should be analyzed for predictive power. Variables, which were found to offer a statistically significant predictive power, were combined with established demographic and student attribute variables to create a new student retention model. Social integration variables were identified in the new model which impacted a student's likelihood of being retained.

For this research, a causal-comparative approach was conducted. A causal-comparative approach is utilized when two groups, in this case retained and not retained students, differ on some type of variable (Fraenkel, Wallen, & Hyun, 2014). A search for causal relationships within a large dataset is a common approach within the quantitative research framework (Leavy, 2017). A causal-comparative study is one in which there is an effort to determine the effect of variables on an outcome when data already exist (Fraenkel et al., 2014). When the effect of variables and the result have already occurred, it is considered causal-comparative (Fraenkel et al., 2014). It is for this reason, causal-comparative research is occasionally called ex post facto research (Fraenkel et al., 2014). Causal-comparative research contrasts a traditional experimental approach in which the researcher's objective is to determine how a treatment influences the outcome (Creswell & Creswell, 2018).

With the assistance of data analysis software and statistical tests, inputs were utilized to measure social integration. Inputs measuring social integration were combined with demographic and student attribute inputs, which have been shown to be predictive of student retention (Bingham & Solverson, 2016; Jia & Maloney, 2014). According to Fraenkel et al. (2014), quantitative researchers seek to better understand and establish relationships between variables, and in some cases, researchers are even able to provide commentary regarding causes of relationships between variables. Specifically, quantitative research calls for the use of dependent and independent variables (Creswell & Creswell, 2018).

According to Salkind (2016), a dependent variable "is the outcome variable, or what the researcher looks at to see if any change has occurred as a function of the treatment that has taken place" (p. 103). The treatment is referred to as the independent variable (Salkind, 2016). The dependent variable for this study is an ordinal variable. An ordinal level of measurement is one in which values are assigned to various categories the variable could fall into (Bluman, 2017). To be an ordinal variable, precise differences within the values of the variables must not exist (Bluman, 2017). In this instance, the ordinal dependent variable is also represented in a binary manner, meaning just two possible outcomes exist (Bluman, 2017). For each case within this study, the dependent variable was represented by a one, retained, or zero, not retained.

Archived institutional data at a Midwestern four-year public institution were gathered via the university's institutional effectiveness office (Institutional Data, 2018).

Data approval from the Midwestern four-year public institution in this study was obtained (see Appendix A). As is the case in quantitative research, numbered data were analyzed with statistical procedures (Creswell & Creswell, 2018). As a part of this quantitative study, the researcher tested the theoretical framework, which included theorizing the crucial nature of social integration for first-year students of the campus community and development of a sense of belonging (Tinto, 1993). The variables were identified, related to the hypotheses, and tested for validity and reliability. Then statistical procedures were implemented to determine the effect of independent variables on the dependent variable through a statistical modeling process (Creswell & Creswell, 2018).

While evaluating which research method was necessary for this study, other methods were considered. Typically, a qualitative method would be appropriate if a theory were being developed (Creswell & Creswell, 2018); however as previously mentioned, a theory was not developed for this study. Rather, the study was meant to conceptualize the relationship between variables (Punch, 2014). Additionally, in the case of this study, a dataset already existed for research. Collecting qualitative data through observations, interviews, focus groups, or collection strategies, which are common in qualitative methods of research, were not necessary (Creswell & Creswell, 2018). These reasons were also considered in the decision to not utilize a mixed method approach for this study. Qualitative and mixed methods research designs provide researchers with more open-endedness than quantitative close-ended approaches (Creswell & Creswell, 2018). Pros and cons exist for each type of research, however, in this case, the researcher determined quantitative as the most appropriate method.

Population and Sample

The Integrated Postsecondary Education Data System's (IPEDS) measure for retention is the percentage of a fall cohort retained to the subsequent fall (IPEDS 2016-17 Glossary, 2017). The IPEDS defines the relevant cohort as the number of first-time, full-time, bachelor degree-seeking students enrolled at an institution (IPEDS 2016-17 Glossary, 2017). As a measure, retention is demonstrated as a percentage of those cohort students measured at census of one fall semester who are enrolled at census of the subsequent fall semester (IPEDS 2016-17 Glossary, 2017). Census is the date at which an institution reports an official enrollment figure (IPEDS 2016-17 Glossary, 2017). As such, the population for the study was based on the IPEDS measure for retention. Sample and population were equal and included all first-time, full-time, bachelor's degree-seeking students enrolled at a Midwestern four-year public institution during the Fall 2016 semester. The population and sample for the study were the same for each research question.

When collecting data, the researcher ensured all members of the sample for the study had equal access to the variables analyzed. Additionally, with the assistance of descriptive statistics, the sample was disaggregated. Concerning gender, 56.7% of the sample was female (Institutional Data, 2018). White, non-Hispanic students made up the majority of the sample at 72.8% (Institutional Data, 2018). The second highest subpopulation for ethnicity was Black, non-Hispanic at 10.1%, while Hispanic or Latino students comprised 5.4% (Institutional Data, 2018). No other ethnicity registered higher than 4% of the sample (Institutional Data, 2018).

Another key point considered was the type of degree students within the sample were pursuing. The Midwestern four-year public institution studied was a Liberal Arts institution, and there was a possibility results would be different for other types of institutions. For the group sampled, 47.6% were pursuing a Bachelor of Science, 20.3% were undeclared, 11.4% a degree in Business Administration, 11.3% a degree in Education, 7.1% a degree in Arts, and the remaining 2.2% were dispersed among fine arts, general studies, and social work (Institutional Data, 2018). It should also be noted, the Midwestern four-year public institution was a commuter campus, despite 46% of the sample living on-campus (Institutional Data, 2018). Overall, 28.9% of students in the sample were from within a 10 mile radius, while an additional 69.5% were from within 605 miles of the institution (Institutional Data, 2018). In total, 97% of students in the sample paid in-state tuition rates (Institutional Data, 2018).

Validity and reliability. The model developed in this research needed to provide validity and reliability (Field, 2017). Validity and reliability are measures which provide confidence an instrument is properly doing its job (Field, 2017). If the researcher cannot be sure an instrument is measuring what it is intended to measure (Creswell & Creswell, 2018). Measurements could result in inaccurate conclusions regarding data generated (Creswell & Creswell, 2018). Validity is if an instrument measures what it is intended to measure (Field, 2017). Reliability is whether this instrument can be utilized in different circumstances (Field, 2017). Multiple goodness-of-fit assessments to gauge validity and reliability were applied to this research and are discussed later in this chapter.

Data Collection

Data collection for this research began following approval from the institutional review boards of Lindenwood University (see Appendix B) and the Midwestern four-year public institution where the study took place (see Appendix C). Data for the study were from the fall 2016, spring 2017, and fall 2017 semesters at the institution (Institutional Data, 2018).

For each student in the sample, demographic, academic performance, student attribute, and social integration data were requested (see Appendix D). In totality, the deidentified student data requested from the Midwestern four-year public institution included 24 variables (Institutional Data, 2018). Demographic variables requested were age, gender, race, and ethnicity (Institutional Data, 2018). Student attribute variables included enrollment status, student financial information, domestic status, program of student, and non-traditional student status (Institutional Data, 2018). Academic performance variables requested included credit hours, first-year grade point average, high school grade point average, standardized test scores, and tutoring information (Institutional Data, 2018). Finally, social integration variables requested were touchpoint data for campus activities, recreation facilities, recreation programs, as well as membership or status information for student organizations, college athletics, fraternities or sororities, the Honors program, student government, recreation facilities, and student residency (Institutional Data, 2018).

While previous models offered a narrow range of variables by adding measurements of the social and campus climate, a wider focus was made possible (Bingham & Solverson, 2016). A key element of Tinto's (1982) theory about retention was the way the student and campus interact with one another. With more variables to measure social integration, predictive power of the model may be increased (Bingham & Solverson, 2016). Once information was obtained from the campus, data analysis began.

Data Analysis

Research questions for the project involved two distinct methods of data analysis. For research questions one through six, a two-proportion *z*-test provided the determination on whether the null hypotheses was not rejected or rejected (Bluman, 2017). For research question seven, a binary logistic regression analysis was conducted to construct the best possible model given the independent variables and method utilized (Bingham & Solverson, 2016). It is important to note, while findings of research questions one through six could, in some cases, provide a road map toward model construction, research questions one through six were distinct from research question seven. Research questions one through six were designed to examine various measures of social integration and the effect on retention in a vacuum. However, analysis for research question seven was intended to determine effects of social integration measurements when they interact with a student's demographics, attributes, and academic performance.

Research questions one through six. To test the first six hypotheses of the study, Microsoft Excel's data analysis capabilities were utilized. Through Microsoft Excel, a two-proportion *z*-test was conducted on each social engagement variable. A two-proportion *z*-test was appropriate because there were only two outcomes and data were binomial, retained or not retained (Bluman, 2017). According to Bluman (2017) "the *z*-test is a statistical test for the mean of a population" (p. 411). In each *z*-test, individuals which demonstrated the act of social engagement were placed into group one. Individuals who did not partake in social engagement were placed into group two.

Research question seven. Following *z*-tests, a binary logistic regression analysis was used to evaluate retention rates with deidentified institutional data. Regression is a statistical technique utilized to determine how a set of independent variables can predict a dependent variable (Punch, 2014). A binary logistic regression approach was established as good practice in retention analysis and modeling by Bingham and Solverson (2016). Logistic regression was used as a method to modeling has grown in popularity over the last 10 years (Belhekar, 2016). Logistic regression as a technique is especially popular in the medical research field, and its growing popularity is due to the method's ability to make predictions for analysis, which involve a binary target variable (Belhekar, 2016). Logistic regression is popular in medical research because it functions well to the common question, whether a patient was cured or not (Belhekar, 2016). Since the target variable in this study was also dichotomous, logistic regression was appropriate (Belhekar, 2016). For the study, the target variable, retention, was represented by a one if the student was retained, or a zero if the student was not retained (Bingham & Solverson, 2016).

To determine if the model was successful in measuring what was intended, the model was tested for fit (Levine et al., 2016). One method for determining fit and assessing overall performance of a model is to apply the use of R^2 statistics ((Jia & Maloney, 2015)). The R^2, or the coefficient of multiple determinations, is a measure of the proportion of variability in a data set which is accounted for by a statistical model (Levine, Stephan, Krehbiel, & Berenson, 2016). The R^2 can assume values between zero, no variation explained, to one, 100% of variation explained (Levine et al., 2016). The closer R^2 is to one, the better the model fits the data (Levine et al., 2016). The larger the R^2, the greater the amount the dependent variable's variance is accounted for (Punch, 2014). The R^2 is sometimes inflated based on the number of dependent variables included in a model (Levine et al., 2016). For this reason, adjusted R^2 was used (Levine et al., 2016). Adjusted R^2 considers the number of dependent variables and enables comparison of models with changing numbers of variables included (Levine et al., 2016).

To conduct logistic regression, many variables were included in analysis. The rationale for inclusion of the variables was provided in Chapter Two and are revisited in the conclusions section of Chapter Five. The following sections include discussion of variables which were included in the binary logistic regression analysis.

To simplify the conversation regarding variables, they have been grouped by type within this text. In all, 24 variables were included within this analysis. Of the 24 independent variables, 14 were scale variables (Institutional Data, 2018). Scale variables included fall credits, composite ACT score, high school grade point average, accepted student loan amount, accepted scholarship amount, student income, parent income, as well as touchpoint data for the aquatic center, student-activity-fee-funded events, fitness programs, intramural sports participation, the recreation facility, sporting events, and career services walk-ins (Institutional Data, 2018).

Ten variables were categorical. Eight of the categorical variables were binary. For gender, 1 represented male, while 0 represented female (Institutional Data, 2018). For the other binary variables, 1 represented membership within the subpopulation and 0 represented non-membership (Institutional Data, 2018). Binary categorical variables included gender, student senate membership, whether the student was a tutee, campus resident status, whether the student was traditional or non-traditional, whether the student was a student athlete, and whether the student was a member of a fraternity or sorority (Institutional Data, 2018). Additionally, ethnicity was represented by eight category options, 0 for decline to answer, 1 for Native American or Alaskan Native, 2 for Asian, 3 for Black non-Hispanic, 4 for Hispanic or Latino, 5 for multiple races, non-Hispanic, 6 for Hawaiian or Pacific Islander, and 7 for White non-Hispanic (Institutional Data, 2018). Residency was the final categorical variable. Residency was designated as 0 for foreign, 1 for in-state resident, 2 for out of state resident but in-state tuition recipient, 3 for member of the Midwest Student Exchange, and 4 for non-resident (Institutional Data, 2018). When inputting categorical variables into the binary logistic regression analysis, the contrast for each was set as an indicator.

To determine which independent variables collaborate to form the best model, the stepwise selection approach to model building was utilized (Bingham & Solverson, 2016). Prior to beginning the stepwise selection process, all independent variables in the model were inputted as predictors (Bingham & Solverson, 2016). A stepwise selection approach utilizes either a forward-based method or a backward method (Field, 2017). The forward method starts with one constant then adds variables to the model, while the backward method begins the process with all predictors included (Field, 2017). For this study, the backward method was conducted. The backward method approach was recognized as appropriate in previous research on retention modeling (Bingham & Solverson, 2016). The backward stepwise approach to model building utilizes three removal criteria (Field, 2017). Beginning with all of the model’s variables, variables were then added and subtracted according to the operation which results in the lowest measure of the Akaike Information Criterion (Bingham & Solverson, 2016). The stepwise selection procedure is a common method in model construction (Punch, 2014).

The Akaike Information Criterion is a goodness-of-fit measure, which accounts for the number of parameters estimated (Field, 2017). Whichever variables are left in the model can be organized into an equation with the variable's coefficients and model's intercept (Bingham & Solverson, 2016). In addition to regression coefficients, data analysis tools in SPSS were used to determine standard errors and *p*-values for each of the model's predictors (Bingham & Solverson, 2016).

Following development of the model, goodness-of-fit was assessed (Bingham & Solverson, 2016). One goodness-of-fit tool employed was the Hosmer-Lemeshow test (Bingham & Solverson, 2016). The Hosmer-Lemeshow test produces the *C*-Statistic (Bingham & Solverson, 2016). The *C*-Statistic is the measurement of the area which falls under the receiver operator characteristic curve (Bingham & Solverson, 2016). One way to graphically depict the interplay between sensitivity and specificity is the receiver operator characteristic curve (Jia & Maloney, 2015). Sensitivity is how likely the non-retention outcomes are correctly determined, while specificity is the likelihood a student's retention is correctly identified (Jia & Maloney, 2015). Sensitivity and specificity information can help to understand the predictive power a model has outside of the samples, also known as its reliability (Jia & Maloney, 2015). Typically, a value between 0.6 and 0.7 for the *C*-Statistic is considered to have a predictive power, while values 0.7 to 0.8 are considered fair, and any values above 0.8 are good (Bingham & Solverson, 2016).

Following the Hosmer-Lemeshow test, further analysis was conducted to better investigate specificity and sensitivity in the model, comparable to the analysis of Bingham and Solverson (2016). Sensitivity represented the ability of the model to accurately project retention (Bingham & Solverson, 2016). These instances are also referred to as true positives (Bingham & Solverson, 2016). Specificity accounted for the number of correct non-retention predictions, also known as true negatives (Bingham & Solverson, 2016). Bingham and Solverson (2016) established the process of comparing three cutoffs for determining whether to predict a retained or non-retained outcome. Bingham and Solverson (2016) utilized cutoffs at the 0.65, 0.7, and 0.75 levels. Due to a lower rate of retention in the sample of this study compared to the one utilized by Bingham and Solverson (2016), cutoffs at the 0.5, 0.6, and 0.7 marks were tested.

Another method in which validity of the developed instrument was tested was to compare predicted and actual outcomes (Jia & Maloney, 2015). Comparing predicted results to actual results was an essential piece in the evaluation of a model produced by regression analysis (Punch, 2014). To accomplish the test, predicted probabilities were ranked and sorted into deciles (Jia & Maloney, 2015). Sorting into deciles was a way to subgroup a dataset into 10 groups of equal size (Bluman, 2017). If the model had no fit, the top decile would not feature a higher retention rate than other deciles (Jia & Maloney, 2015). Additionally, the lowest decile would not demonstrate a retention rate lower than other deciles (Jia & Maloney, 2015). Any percentage higher than 10% demonstrates varying degrees of validity within the model (Jia & Maloney, 2015). By using the actual retention percentage of students in the sample, the researcher then determined how many students were projected to have a probability higher than the mark of those who were not retained, and how many lower than the mark were not retained (Jia & Maloney, 2015). These percentages were then compared to the actual percentage to assess validity of the model (Jia & Maloney, 2015).

Ethical Considerations

There are serious ethical considerations in all studies which involve big data (De Freitas et al., 2015). When analyzing big data, it is essential to do so with caution and without bias (De Freitas et al., 2015). To assure confidentiality of the population of this study, all identifying information was removed from the institutional data (Institutional Data, 2018). Deidentification of the data were completed prior to any other step in the analysis process of the study. It is important when predictive models are developed predictions are based on more than race, ethnicity, or socioeconomic status (Ekowo & Palmer, 2017). By including other variables, the production of discriminatory results was avoided (Ekowo & Palmer, 2017).

Summary

In Chapter Three, the research methodology for the study was presented in depth. Presenting research questions and hypotheses for the study was the first step. Information regarding the quantitative nature of the study and why this approach was deemed most appropriate was provided in Chapter Three. A population and sample for the study were identified. In the study, the researcher attempted to develop a statistically significant model, which offered both validity and reliability to the user (Creswell & Creswell, 2018). Data were analyzed using logistic regression, while procedures were conducted to test the model for goodness-of-fit. In Chapter Four, results of this analysis are presented.

Chapter Four: Analysis of Data

Over the past decade, as tuition rates increased across the country, there has been a growing interest in accountability of higher education institutions (Valbrun, 2018). With added interest came additional pressure for higher education administrators to improve retention and graduation rates (Bingham & Solverson, 2016). Amplifying the challenge to improve retention and graduation rates was a recent crunch in higher education appropriations (Seltzer, 2018). From 2017 to 2018, the average increase in state appropriations for higher education was lower than the rise in the consumer price index (Seltzer, 2018). Despite budget woes, state and federal lawmakers are not just demanding improved performance in retention and graduation outcomes, but also an improved collegiate experience for students (Valbrun, 2018).

In the following sections, the data analysis process is discussed in depth. Chapter Four is organized by research question. The first six research questions have been grouped together because each required the same statistical test. In Chapter Four e an overview of research question seven is featured. Within the research question seven subsection, there is an explanation regarding results of the logistic regression, the developed model, and the goodness-of-fit assessment of the model.

Data Analysis

A quantitative research design was selected as the most appropriate method to complete the research of this study. Data used in the study were deidentified student data from an entire cohort of a Midwestern four-year public institution. Beginning this chapter is an overview of the seven hypotheses tested in the study. The first six of these hypotheses required a *z*-test for the difference between two proportions (Gurnsey, 2017).

Test results determined if, when considered in isolation, the social integration variable had a statistically significant impact on student retention from first- to second-year at a Midwestern four-year public institution.

The seventh hypothesis required the use of SPSS to generate a binary logistic regression model for the data, which could accurately predict whether a student would be retained from the first year of college into the second year (Belhekar, 2016; Bingham & Solverson, 2016). Twenty-seven variables were tested as predictors for the target variable, retention (Institutional Data, 2018). These variables included previously established measurements for demographics, student attributes, and academic performance, as well as social integration data (Institutional Data, 2018). Following the development of a model, several goodness-of-fit measures and model assessments were conducted to evaluate significance of the model. These tests included measurements of R^2, a Hosmer-Lemeshow Test, the *C*-Statistic, further sensitivity and specificity testing, and a decile comparison of the model's predicted outcomes and actual outcomes (Bingham & Solverson, 2016; Levine et al., 2016).

Research Questions: Tests and Results

Research questions one through six in the study were posed to analyze the effects of social integration on a student's retention outcome from first- to second-year. To answer the first six questions, retention and social integration were tested for a statistically significant difference. Social integration measurements included campus fitness programs, fraternity or sorority membership, intramural sports participation, recreation facility usage, whether the student lived on campus, and attendance at student-activity-fee-funded events. When comparing proportions of one group of the population to another with a parameter applied for questions one through six, it was appropriate to use a *z*-test to determine if there was a statistically significant difference between the proportions of the two groups (Gurnsey, 2017; Salkind, 2016). The null hypothesis could also have been expressed as $\bar{p}1 = \bar{p}2$. Thus, the alternative hypothesis was also expressed as $\bar{p}1 \neq \bar{p}2$. Each test was conducted at a 95% confidence interval or $\alpha = 0.05$ (Salkind, 2016). A 95% confidence interval means, with 95% certainty there is a statistically significant difference between the two groups tested within the *z*-test (Salkind, 2016).

Summary of *z*-test for research question one. Research question one, *Does student participation, minimum one class attended, in campus fitness programs have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution*, was analyzed using inferential statistics. The purpose of the analysis was to attempt to reject the null hypothesis. For research question one, the null hypothesis was expressed as having no statistically significant difference in first-year to second-year retention between campus fitness program participants and non-participants.

Table 1 contains the results of the *z*-test for null hypothesis $H1_0$. In the sample, 58 of the 939 student sample participated in campus fitness programs, while 881 did not. Overall, the retention rate of campus fitness program participants was 81%, while the retention of non-participants was 63.5%. The *z*-test value was 2.71, which translates to a *p*-value of 0.007. Since the value is greater than the two-tail critical value *Z* of 1.96, the null hypothesis, $H1_0$, was rejected. At the 0.05 level of significance, there was sufficient evidence to conclude the first- to second-year retention of campus fitness program participants was different than the first- to second-year retention of campus fitness program non-participants.

Table 1

Z-test Results for Fitness Program Participation

Groups	Participants		Non-Participants
Retained in Group	47		559
Sample Size of Group	58		881
Proportion of Group	0.810		0.635
Z-Test Value		2.711	
Two-tail(+/-) Critical Value *Z*		1.960	

Note. Retained in Group is the number of students within the population who were retained. Proportion of Group represents the percentage of the population who were retained.

Summary of *z*-test for research question two. Research question two, *Does student membership in a fraternity or sorority have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution*, was analyzed using inferential statistics. The purpose of analysis was to attempt to reject the null hypothesis. For research question two, the null hypothesis was expressed as having no statistically significant difference in first-year to second-year retention between fraternity or sorority members and non-members.

Table 2 contains the results of the *z*-test for null hypothesis $H2_0$. In the sample, 42 of 939 students sampled were members of fraternity or sorority programs, while 897 were not. Overall, the retention rate of fraternity or sorority members was 85.7%, while the retention of non-participants was 63.5%. The *z*-test value was 2.94, which translates to a *p*-value of 0.003. Since the value was greater than the two-tail critical value *Z* of 1.96, the null hypothesis, $H2_0$, was rejected. At the 0.05 level of significance, there was sufficient evidence to conclude the first- to second-year retention of fraternity and sorority participants was different than the first- to second-year retention of fraternity and sorority non-participants.

Table 2

Z-test Results for Fraternity or Sorority Participation

Groups	Participants		Non-Participants
Retained in Group	36		570
Sample Size of Group	42		897
Proportion of Group	0.857		0.635
Z-Test Value		2.935	
Two-tail(+/-) Critical Value *Z*		1.960	

Note. Retained in Group is the number of students within the population who were retained. Proportion of Group represents the percentage of the population who were retained.

Summary of *z*-test for research question three. Research question three, *Does student participation, minimum one intramural event attended, in intramural sports programs have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution,* was analyzed using inferential statistics. The purpose of analysis was to attempt to reject the null hypothesis. For research question three, the null hypothesis was expressed as having no statistically significant difference in first-year to second-year retention between intramural sport program participants and non-participants.

Table 3 contains the results of the *z*-test for null hypothesis $H3_0$. In the sample, 178 of the 939 student sample were participants in the intramural sport program, while

761 were not. Overall, the retention rate of intramural sport participants was 67.4%, while the retention of non-participants was 63.9%. The *z*-test value was 0.89, which translates to a *p*-value of 0.37. Since the value was less than the two-tail critical value *Z* of 1.96, the null hypothesis, $H3_0$, failed to be rejected. At the 0.05 level of significance, there was not sufficient evidence to conclude the first- to second-year retention of campus intramural sport participants was different than the first- to second-year retention of intramural sport non-participants.

Table 3

Z-test Results for Intramural Sports Program Participation

Groups	Participants	Non-Participants
Retained in Group	120	486
Sample Size of Group	178	761
Proportion of Group	0.674	0.639
Z-Test Value	0.892	
Two-tail(+/-) Critical Value *Z*	1.960	

Note. Retained in Group is the number of students within the population who were retained. Proportion of Group represents the percentage of the population who were retained.

Summary of *z*-test for research question four. Research question four, *Does student participation, minimum one check-in, at a university recreational facility have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution*, was analyzed using inferential statistics. The purpose of analysis was to attempt to reject the null hypothesis. For research question four, the null hypothesis was expressed as having no statistically significant difference in first-year to second-year retention between university recreational facility participants and non-participants.

Table 4 contains results of the *z*-test for null hypothesis $H4_0$. In the sample, 792 of the 939 student sample were participants of the campus recreation facilities, while 147 were not. Overall, the retention rate of intramural sport participants was 66.5%, while the retention of non-participants was 53.7%. The *z*-test value was 2.98, which translates to a *p*-value of 0.003. Since the value was greater than the two-tail critical value *Z* of 1.96, the null hypothesis, $H4_0$ was rejected. At the 0.05 level of significance, there was sufficient evidence to conclude the first- to second-year retention of campus recreation facility participants was different than the first- to second-year retention of campus recreation facility non-participants. The results of the analysis are presented in Table 4.

Table 4

Z-test Results for University Recreational Facility Participation

Groups	Participants	Non-Participants
Retained in Group	527	79
Sample Size of Group	792	147
Proportion of Group	0.665	0.537
Z-Test Value	2.979	
Two-tail(+/-) Critical Value *Z*	1.960	

Note. Retained in Group is the number of students within the population who were retained. Proportion of Group represents the percentage of the population who were retained.

Summary of *z*-test for research question five. Research question five, *Does student housing status, living on-campus or not, have a statistically significant impact on* *first-year to second-year retention at a Midwestern four-year public institution*, was analyzed using inferential statistics. The purpose of analysis was to attempt to reject the null hypothesis. For research question five, the null hypothesis was expressed as having no statistically significant difference in first-year to second-year retention between on-campus living participants and non-participants.

Table 5 contains the results of the *z*-test for null hypothesis $H5_0$. In the sample, 432 of the 939 student sample were on-campus living participants, while 507 were not. Overall, the retention rate of on-campus living participants was 65.5%, while the retention of non-participants was 63.7%. The *z*-test value was 0.58, which translates to a *p*-value of 0.57. Since the value was less than the two-tail critical value *Z* of 1.96, the null hypothesis, $H5_0$, failed to be rejected. At the 0.05 level of significance, there was not sufficient evidence to conclude the first- to second-year retention of on-campus living participants was different than first- to second-year retention of on-campus campus living non-participants.

Table 5

Z-test Results for On-campus Living Participation

Groups	Participants		Non-Participants
Retained in Group	283		323
Sample Size of Group	432		507
Proportion of Group	0.655		0.637
Z-Test Value		0.575	
Two-tail(+/-) Critical Value *Z*		1.960	

Note. Retained in Group is the number of students within the population who were retained. Proportion of Group represents the percentage of the population who were retained.

Summary of *z*-test for research question six. Research question six, *Does student participation, minimum one event attended, in student-activity-fee-funded events have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution,* was analyzed using inferential statistics. The purpose of analysis was to attempt to reject the null hypothesis. For research question six, the null hypothesis was expressed as having no statistically significant difference in first-year to second-year retention between student-activity-fee-funded event participants and non-participants.

Table 6 contains results of the *z*-test for null hypothesis $H6_0$. In the sample, 748 of the 939 student sample were student-activity-fee-funded event participants, while 191 were not. Overall, the retention rate of student-activity-fee-funded event participants was 67.4%, while the retention of non-participants was 53.4%. The *z*-test value was 3.60, which translates to a *p*-value of 0.00. Since the value was greater than the two-tail critical value *Z* of 1.96, the null hypothesis, $H6_0$, was rejected. At the 0.05 level of significance, there was sufficient evidence to conclude the first- to second-year retention of student-activity-fee-funded event participants was different than the first- to second-year retention of student-activity-fee-funded event non-participants.

Table 6

Z-test Results for Student Activity-Fee-Funded Event Participation

Groups	Participants	Non-Participants
Retained in Group	504	102
Sample Size of Group	748	191
Proportion of Group	0.674	0.534
Z-Test Value	3.604	
Two-tail(+/-) Critical Value Z	1.960	

Note. Retained in Group is the number of students within the population who were retained. Proportion of Group represents the percentage of the population who were retained.

Research question seven. Research question seven, *Does the inclusion of social integration variables, in association with already established variables, which account for demographics, student attributes, and academic performance, produce a statistically significant model which can be used as an instrument for projecting a student's likelihood to be retained,* was analyzed using logistic regression and goodness-of-fit techniques. The discussion regarding research question seven begins with the binary logistic regression analysis. Following discussion regarding the binary logistic regression analysis, the model, which was produced by the analysis, was provided. Once the model was provided, further dialogue regarding assessment of the model through various goodness-of-fit measures follows.

For the final hypothesis, a binary logistic regression analysis was conducted to determine if a statistically significant model could be produced with the dataset (Belhekar, 2016; Bingham & Solverson, 2016). Retention served as the target, or dependent variable. Retention was presented as a binary variable with a one indicating the student was retained, while a zero represented the student was not retained. All independent variables were entered into the model, then, as conducted in previous efforts to build models projecting student retention, the stepwise selection process was implemented (Bingham & Solverson, 2016).

When utilizing the stepwise selection process for model building, the backward based approach was conducted. The backward approach is preferable in regression analysis (Field, 2017). A backward stepwise selection process was better than a forward-based approach because of the possibility some variables have suppressor effects (Field, 2017). A suppressor effect occurs when a variable has a significant effect on a model, but the effect is contingent on another variable as a constant (Field, 2017). Due to the inability to accurately account for the suppressor effect, the forward-based method was more likely to exclude necessary predictors for the model (Field, 2017).

Through computing tools of SPSS, the stepwise approach utilizes one of three methods; likelihood ratio, conditional, or Wald (Field, 2017). In this circumstance, the likelihood ratio method was implemented because it was the most reliable of the three methods (Field, 2017). When the likelihood ratio was utilized, each stage the model was compared to what the model would be like if a variable were removed (Field, 2017). If removal of a variable had a statistically significant impact on the fit of the model for the data, the variable was kept in the model (Field, 2017). If there was not a statistically significant change in the model's fit for the data, the variable was rejected (Bluman, 2017).

With the stepwise selection process, in order of their removal, the following variables were taken out of the final model. Intramural sport touchpoints was the first variable removed. Intramural sport touchpoints were followed by fitness program touchpoints, parent income, accepted loan offer, age, athlete status, residency, student senate status, gender, ethnicity, recreation facility touchpoints, accepted scholarship offer, ACT score, traditional student status, student income, fall credits, campus resident status, aquatic center touchpoints, international touchpoints, and career services touchpoints. Variables which were retained in the final model are displayed in Table 7.

Table 7

Independent Variables Used in the Logistic Regression Retention Model

Independent Variable	Description
Tutee	1 if student received tutoring, 0 otherwise
StudActTP	Total number of student-activity-fee-funded events attended by the student
Greek	1 if student is a member of fraternity or sorority, 0 otherwise
StudEmp	1 if student employee, 0 otherwise
SportTP	Total number of sporting events attended by the student
HSGPA	High school grade point average
CarSerWalkIns	Total number of times the student visited the career services office for career assistance

Note. Greek is used as shorthand for membership within a fraternity or sorority.

Additional information regarding independent variables kept in the model related to logistic regression coefficients, standard errors, and *p*-values for each of the independent variables are listed in Table 8.

Table 8

Logistic Regression, Coefficients, Standard Errors, and p-values

Variable	Coefficient	Exp (Coefficient)	Standard Error	*p*-value
Intercept	-0.619	0.539	0.850	0.466
Tutee	-0.565	0.569	0.245	0.021
StudActTP	0.048	1.049	0.015	0.002
Greek	-1.074	0.342	0.467	0.021
StudEmp	-1.653	0.191	0.502	0.001
SportTP	0.092	1.097	0.049	0.058
HSGPA	1.187	3.277	0.148	0.000
CarSerWalkIns	3.496	32.974	1.013	0.001

Note. Greek represents students who were members of a fraternity or sorority.

Results of the binary logistic regression can be demonstrated in equation form. The coefficients in Table 8 are the same coefficients used to form Equation 1

$$ln\left(\frac{\text{p}}{\text{1-p}}\right) = -0.619 - (0.565 * \text{Tutee}) + (0.048 * \text{StudActTP}) - (1.074 * \text{Greek}) - (1.653 * \text{StudEmp}) + (0.092 * \text{SportTP}) + (1.187 * \text{HSGPA}) + (3.496 * \text{CarSerWalkIns})$$

The logistic regression equation shown in Table 8 is different from a traditional multiple regression formula, as the equation relies on log odds to produce a result between 0 and 100 % (Belhekar, 2016; Field, 2017). Log odds are represented in the portion of the equation on the left side of the equal sign, with *p* representing the probability the student is retained from first-year to second-year (Bingham & Solverson, 2016). To convert to log odds, students' specific inputs can be placed within the equation (Bingham & Solverson, 2016).

The effect of the log odds process is each coefficient is exponentiated (Bingham & Solverson, 2016). For example, for each student-activity-fee-funded event a student attends, a student's likelihood to be retained at the institution increases by a factor of 1.049, which represents a 4.9% increase in the student's retention probability (Bingham & Solverson, 2016). Sporting event touchpoints, high school grade point average, and career services walk-ins can be interpreted in the same manner with the application of the log odds. A positive increase demonstrates the predictor's positive increase on a student's likelihood to be retained, while the opposite is true for predictors with a negative coefficient (Bingham & Solverson, 2016). As is the case with variables in the model which indicate a student's status within a group; each returned a negative coefficient. Greek students would be expected to be retained 34.2% as often as non-Greek students, when the model's other variables are accounted for and all variables are held equal (Bingham & Solverson, 2016). Additionally, the probabilities of a student's retention can also be found with the use of the following function in Equation 2 (Bingham & Solverson, 2016):

$$f(a) = \frac{exp(a)}{1 + exp(a)}$$

Assessment of model. The R^2 statistics are probably the most common measure for assessing regression models (Jia & Maloney, 2015). Two forms of R^2 were included in the SPSS output of the binary logistic regression analysis. These forms included the Nagelkerke R^2, as well as the Cox and Snell R^2. The Cox and Snell R^2 utilized the log-likelihoods of both the new model and original model to produce the statistic (Bluman, 2017). Nagelkerke's adjustment to R^2 was intended to address the fact the Cox and Snell variation did not allow for the measurement to reach 1, the theoretical maximum

(Bluman, 2017). While the two versions of R^2 are different, conceptually, both are ways to measure the same thing (Bluman, 2017). The model produced in this study provided a Nagelkerke R^2 of .243 and a Cox and Snell R^2 of .176. These R^2 mean the model represented in Equation 1, can account for approximately 24.3% or 17.6% of the variation in student retention from first- to second-year depending on the R^2 statistic applied.

Another goodness-of-fit measure is the Hosmer-Lemeshow Test (Bluman, 2017). As a goodness-of-fit measure, the Hosmer-Lemeshow test measures the interaction between the log and predictor variables (Bluman, 2017). A *p*-value of 0.423 was produced by the Hosmer-Lemeshow test. The *p*-value indicated no evidence of lack of fit in the model (Bingham & Solverson, 2016). There is no evidence for lack of fit because the value was greater than 0.05, which would have indicated the model did not fit the data (Bluman, 2017).

The *C*-Statistic was also generated to make another goodness-of-fit assessment. The *C*-Statistic is also a way to generate the target effectiveness of retention probabilities generated by the model (Jia & Maloney, 2015). The *C*-Statistic represents the area under the ROC curve, which can be viewed in Figure 1 (Bingham & Solverson, 2016). A *C*-Statistic with a value of 0.8 and over is considered good, a value between 0.8 and 0.7 is fair, and between 0.7 and 0.6 is poor (Bingham & Solverson, 2016). The *C*-Statistic indicated an improvement on previous models at 0.752. Therefore, the probability a randomly selected non-retained student in the study will have a lower likelihood to be retained than a randomly selected retained student was 75.2% (Jia & Maloney, 2015).

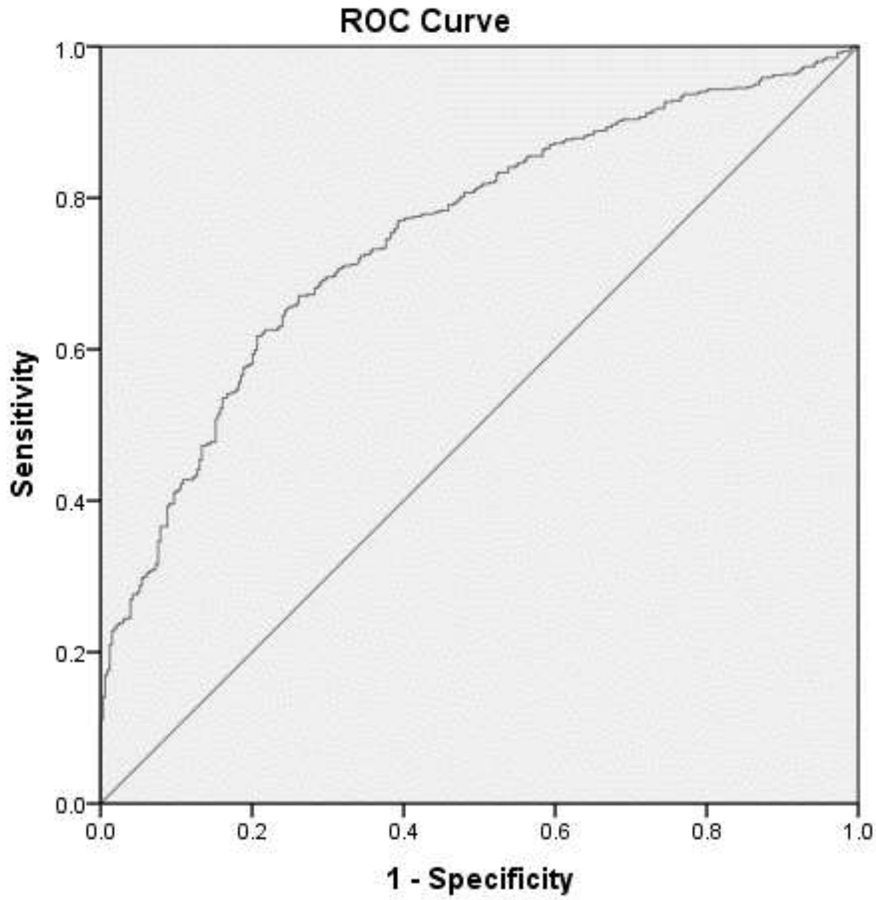

Figure 1. ROC Curve of first- to second-year retention model.

Following development of the ROC curve, additional analysis of the model's specificity and sensitivity was conducted. In Chapter Three, it was determined cutoff levels examined would be at the 0.5, 0.6, and 0.7 levels. A 0.5 cutoff level indicated if the model produced a value for a student above 0.5, or 50.0%, the student was predicted to be retained, while if the value was below the cutoff, the student was not predicted to be retained. Data included in Table 9 demonstrates the count of correct and incorrect predictions produced by the model at the three cutoff levels. Columns labeled as correct represent the number of correctly retained and not retained predictions at the cutoff level, while the opposite is true for those listed in the incorrect columns.

Table 9

Count of Correct and Incorrect Predictions

	Correct		Incorrect	
Cutoff Level	Retained	Not Retained	Retained	Not Retained
0.5	505	151	96	178
0.6	436	213	165	116
0.7	318	276	283	53

Note. Cutoff levels were determined and discussed at length in Chapter Three.

Utilizing information provided in Table 9, several other important pieces of information including sensitivity and specificity were computed. Table 10 includes results for overall effectiveness of cutoff, sensitivity, specificity, as well as false positive and false negative predictions. Sensitivity represents the percentage of student retentions correctly predicted (Bingham & Solverson, 2016). Specificity is the percentage of correctly predicted non-retentions (Bingham & Solverson, 2016). False positive accounts for the percentage of students predicted to be retained who were not (Bingham & Solverson, 2016). False negative is the percentage of students not predicted to be retained, but returned for the second-year of college (Bingham & Solverson, 2016).

As is demonstrated in Table 10, the cutoff level with the highest success rate is 0.5. At the 0.5 cutoff level, retention outcomes are predicted correctly 70.8% of the time. Utilizing the 0.5 cutoff, the model correctly forecasted a student to be retained 83.7% of the time and correctly predicted a non-retained outcome in 47.1% of cases. When the cutoff level is increased, the model becomes less accurate in predicting retention outcomes, which results in an overall less effective forecast.

However, as the cutoff level is increased, the number of correct non-retention outcomes increases, which results in a higher specificity. A tradeoff between sensitivity and specificity is to be expected as cutoff levels increase, which is the balance to be weighed by the researcher (Bingham & Solverson, 2016). Overall, cutoff levels at 0.5 and 0.6 perform reasonably well.

Table 10

Correct Predictions, Sensitivity, Specificity, False Positives, and False Negatives

Cutoff Level	Correct	Sensitivity	Specificity	False Positive	False Negative
0.5	70.8%	83.7%	47.1%	16.3%	52.9%
0.6	69.9%	71.4%	67.2%	28.6%	32.8%
0.7	63.0%	50.9%	85.1%	49.1%	14.9%

Note. Cutoff levels were determined and discussed at length in Chapter Three.

The final method to evaluate the model for effectiveness was to compare actual outcomes with predicted outcomes (Jia & Maloney, 2015). By applying the model in Equation 1, all students from the sample were ranked based on predicted probabilities to be retained (Jia & Maloney, 2015). The sample was then divided into 10 groups, or deciles (Jia & Maloney, 2015). Each decile included 93 students, except for decile 10, which had 92 students. The sum of the probabilities was divided by the total number of the students to create an expected and actual retention rate for each decile.

If the model is effective, a progression of retention rate is seen, which mirrors the predicted retention rate for each decile (Jia & Maloney, 2015). If the model is ineffective, actual retention rates of the deciles appear random or equal (Jia & Maloney, 2015). In Table 11, a comparison of predicted and actual outcomes is organized by decile.

Table 11

Comparison of Predicted and Actual Outcomes by Decile

	Predicted Outcomes			Actual Outcomes		
Decile	Retained	Not Retained	Retention Rate	Retained	Not Retained	Retention Rate
Decile 1	29.2	63.8	31.4%	34	59	36.6%
Decile 2	39.5	54.5	42.0%	36	58	38.3%
Decile 3	45.7	47.3	49.2%	44	49	47.3%
Decile 4	52.4	40.6	56.4%	48	45	51.6%
Decile 5	58.3	34.7	62.7%	54	39	58.1%
Decile 6	63.0	30.0	67.8%	67	26	72.0%
Decile 7	67.6	25.5	72.6%	72	21	77.4%
Decile 8	73.7	19.4	79.2%	78	15	83.9%
Decile 9	81.8	11.2	87.9%	78	15	83.9%
Decile 10	89.9	2.1	97.7%	90	2	97.8%

Note. Each decile was comprised by 93 students, except for Decile 10, which had 92 students.

The decile process is also beneficial to determine which groups of the sample are prime targets for intervention (Jia & Maloney, 2015). For example, given the results found in Table 11, 35.6% of not retained students would be captured if the bottom 20% of predicted outcomes were targeted. If the bottom 50.0% of predicted outcomes were targeted, the number rises to 76.0% of students within the study who were not retained into the second-year of college.

Summary

Chapter Four began with an introduction into data analysis. Seven null hypotheses were tested in the analysis. For hypotheses one through six, a *z*-test was conducted (Gurnsey, 2017; Salkind, 2016). For hypotheses 1, 2, 4, and 6, the null hypothesis was rejected, and a statistically significant difference in retention rates of the groups was found. For hypotheses 3 and 5, a failure to reject the null hypothesis occurred, and a statistically significant difference between groups was not found.

The analysis of hypothesis seven required development of a statistically significant model which could predict student retention outcomes from the first- to second-year of college. Overall, twenty-seven variables were tested using a binary logistic regression analysis with a backwards stepwise selection approach to model building (Institutional Data, 2018). Eight variables were included in the final model (Institutional Data, 2018). Following the model's development, tests were conducted to assess the model. These tests included measurements of R^2, a Hosmer-Lemeshow Test, the *C*-Statistic, further sensitivity and specificity testing, and a decile comparison of the model's predicted outcomes and actual outcomes. Higher R^2 scores than the 11.5% R^2 established in 2014 by Jia and Maloney were generated from the model. The Hosmer-

Lemeshow test returned a *p*-value of .423. Finally, the *C*-Statistic, or the area under the ROC curve, yielded a 0.752, an improvement over previous attempts, which yielded 0.718 (Jia & Maloney, 2015) and 0.651 (Bingham & Solverson, 2016). An improved *C*-Statistic is significant because it indicates a stronger model can be produced when social integration variables are introduced to the model building process.

In Chapter Five, findings of the research are presented. Each research question is considered and tied back to the student retention theory and literature review found in Chapter Two. Conclusions regarding the research were drawn based on data analysis and how information connected to theory. Additionally, implications of the findings within the study are discussed. Then, recommendations for future research are presented.

Chapter Five: Summary and Conclusions

This study was designed to determine the impact of social integration on student retention. In Chapter Two, prior research was discussed to determine previous studies from which a methodology could be expounded upon, as well as variables which had been previously linked to student retention. In Chapter Three, the study's methodology was presented. Chapter Three also included an explanation on how research questions one through six were answered with a two proportion *z*-test. Additionally, research question seven was answered with a binary logistic regression and assessed with several goodness-of-fit measures. In Chapter Four, the researcher reported on data analysis of statistical tests for each research question. Chapter Five begins with findings and conclusions from the research and continues with implications of research findings, as well as recommendations for future research.

Findings

Findings of the research have been broken down into seven subsections. Each research question is presented as a subsection in which the findings are provided. Further discussion of the findings, as well as the implications of the researcher's findings, are presented later in the chapter.

Research question one. A *z*-test with a confidence level of 95% or $\alpha = 0.05$ was conducted to determine statistical significance between retention rates of campus fitness program participants and non-participants. Since the *p*-value for research question one was 0.007, results were deemed statistically significant. The null hypothesis was rejected.

Research question two. A *z*-test with a confidence level of 95% or $\alpha = 0.05$ was conducted to determine statistical significance between the retention rates of fraternity or sorority members and non-members. Since the *p*-value for research question two was 0.003, results were deemed statistically significant. The null hypothesis was rejected.

Research question three. A *z*-test with a confidence level of 95% or $\alpha = 0.05$ was conducted to determine statistical significance between the retention rates of intramural sport participants and non-participants. Since the *p*-value for research question three was 0.372, results were not deemed statistically significant. There was a failure to reject the null hypothesis.

Research question four. A *z*-test with a confidence level of 95% or $\alpha = 0.05$ was conducted to determine statistical significance between the retention rates of recreation facility users and non-users. Since the *p*-value for research question four was 0.003, results were deemed statistically significant. The null hypothesis was rejected.

Research question five. A *z*-test with a confidence level of 95% or $\alpha = 0.05$ was conducted to determine statistical significance between the retention rates of on-campus residents and off-campus residents. Since the *p*-value for research question five was 0.565, results were not deemed statistically significant. There was a failure to reject the null hypothesis.

Research question six. A *z*-test with a confidence level of 95% or $\alpha = 0.05$ was conducted to determine statistical significance between the retention rates of student-activity-fee-funded event participants and non-participants. Since the *p*-value for research question six was 0.000, results were deemed statistically significant. The null hypothesis was rejected.

Research question seven. In research question seven, a model was developed with binary logistic regression techniques (Belhekar, 2016). Several goodness-of-fit measures were tested and compared with previous modeling efforts to determine if social integration was also a functional way to predict student retention. The model produced a Nagelkerke R^2 of 0.243 and a Cox and Snell R^2 of 0.176. The Hosmer-Lemeshow test returned a p-value of 0.423. The area under the ROC curve, otherwise known as the C-statistic, was 0.752 meaning there was a 75.2% chance a randomly selected non-retained student in the study would have a lower likelihood to be retained than a randomly selected retained student (Jia & Maloney, 2015).

Conclusions

The design of this study was built around the theoretical framework outlined in Chapter One and Chapter Two. Theories to support this research included Gennep's (1960) rites of passage, Tinto's (1993) theory of student departure, and Astin's (1999) theory of student involvement. The theories served as the basis for the establishment of social integration as critical in the student retention process (Astin, 1975, 1999; Gennep, 1960; Tinto, 1993). The review of current and seminal research led to a quantitative research design to test statistical significance of various social integration variables on a student's likelihood to be retained at a Midwestern four-year public institution. Data analysis required z-tests to determine whether proportions of the two groups featured a statistically significant difference (Salkind, 2016). Additional statistical analysis required use of a binary logistic regression technique to develop a model to predict the likelihood of a student's retention given a variety of inputs (Punch, 2014). The use of binary logistic regression as a strategy for retention analysis was established through prior research (Bingham & Solverson, 2016; Jia & Maloney, 2014). The product of this research was designed to bridge the gap between the theorized importance of social integration on retention and recent attempts to model student retention.

Research question one. The first research question was: *Does student participation, minimum one class attended, in campus fitness programs have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution*? The null hypothesis was rejected for this question, and results were deemed to be statistically significant. A finding of statistical significance was consistent with previous research. Tinto's (1993) theory of student departure stated there was a relationship between student success outcomes and involvement in campus recreation programs. In 1999, Astin also established relationship between exercise and student fitness with student satisfaction and degree attainment. More recently, fitness participation was tested as a link to student success outcomes through data analysis (Belch et al., 2001). Research of Belch et al. (2001) found a relationship between fitness and positive student success outcomes. Findings of this research provide additional statistical support of a positive link between student retention and participation in campus fitness programs.

Research question two. The second research question was: D*oes student membership in a fraternity or sorority have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution*? The null hypothesis was rejected, and results were deemed statistically significant. Findings of this research question validated previous student retention theory and statistical research. Fraternities and sororities provide students the opportunity to connect with other students

(Tinto, 1988). Connections with other students foster and increase sense of belonging, which is theorized to result in a greater likelihood of retention (Astin, 1999). Relationships a student develops through Greek life results in stronger integration within the institution's community (Tinto, 1988). Prior statistical analysis found involvement within student groups, such as fraternities or sororities, during a student's second-year of college increased the likelihood a student would be retained and eventually graduate (Branand et al., 2015). In addition to a predictability of retention outcomes for second-year students involved in student groups, results of research question two indicated a difference in retention outcomes for first-year students involved in fraternities and sororities. Findings of research question two corroborated the 2015 Branand et al. findings.

Research question three. The third research question was: D*oes student participation, minimum one intramural event attended, in intramural sports programs have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?* The null hypothesis was not rejected for this question as the results were not deemed statistically significant. Results of research question three were not consistent with research discussed in Chapter Two. Both Astin (1993, 1999) and Tinto (1988, 1993) established a theoretical relationship between student success outcomes and participation in intramural sports. In fact, some research has even indicated intramural sport participation is among the most positive factors which influence student persistence (Astin, 1999; Belch et al., 2001). Intramural sports provide students with an enormous opportunity to interact with fellow students and integrate into a social community (Belch et al., 2001). For some students, a considerable portion of the university experience is made up of these interactions (Blumenthal, 2009). For first-year students, intramural sports can provide opportunities to build study groups, seek advice, find study partners, and make friends (Belch et al., 2001). Despite findings of other research, analysis of this student population did not yield statistically significant results to strengthen those perspectives.

Research question four. The fourth research question was: D*oes student participation, minimum one check-in, at a university recreational facility have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?* The null hypothesis was rejected for this question, and results were deemed to be statistically significant. These findings corroborated student retention theory and prior research on the effect of recreation facilities on student retention. In 1982, Tinto theorized the rise in construction of recreation facilities on college campuses could serve as a mean to centralize interactions between students, as well as between students and faculty. Research findings, as discussed in Chapter Four validate Tinto's (1993) claims, which stated it is equally important for students to be engaged in areas of college life such as recreational activities, as it was to be academically involved. In 1993, Astin also found links between exercise and student success outcomes such as degree attainment. Prior analysis of recreation center users versus nonusers found students were 7.0% more likely to be retained from the first- to second-year if they were recreation facility users (Belch et al., 2001). In the sample analyzed for research question four, recreation facility users were retained at a rate 12.8% higher than non-users. As theorized in previous research, recreation facilities serve as community hubs for the college campus (Belch et al., 2001; Tinto, 1982). Facilities provide students the opportunity to develop a sense of belonging to the institution through connections with other students, faculty, and staff (Belch et al., 2001; Tinto, 1982). Results indicate interaction opportunities first-year students obtain through recreation facility usage could lead to greater satisfaction with the institution, and as a result a higher likelihood of retention (Belch et al., 2001).

Research question five. The fifth research question was: D*oes student housing status, living on-campus or not, have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?* The null hypothesis was not rejected for this question, as results were not deemed statistically significant. Results of research question five were not consistent with the research discussed in Chapter Two. Findings of previous researchers indicated living on campus had a positive effect on student retention (Astin, 1999). These findings were consistent across various institution types and were relevant among all types of students (Astin, 1999). However, findings demonstrated in Chapter Four did not establish an increased level of student retention for resident students (Institutional Data, 2018). The finding runs contrary to prior work in student retention theory, which would lead one to believe the increased time on campus would result in stronger connections to the university (Astin, 1999; Vlanden & Barlow, 2014). After all, a student who lives on campus should have a stronger attachment to the institution (Astin, 1999; Bronkema & Bowman, 2017). Results of research question five were also inconsistent with more recent research findings (Branand et al., 2015; Bronkema & Bowman, 2017). Findings of research question five were not statistically significant to conclude the student population living on campus analyzed effected first- to second-year retention.

Research question six. The sixth research question was: D*oes student participation, minimum one event attended, in student-activity-fee-funded events have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?* The null hypothesis was rejected for this question and results were deemed to be statistically significant. Findings serve to validate Tinto's (1993) theory, which included student activity events and campus organizations were the most critical ways for students to develop social integration. Through social integration, students are, in theory more likely to be retained (Astin, 1993, 1999; Tinto, 1988, 1993). Results of research question six is further evidence student activity events are ways to integrate within the campus community (Belch et al., 2001; Branand et al., 2015).

Research question seven. The seventh research question was: D*o social integration variables, in association with already established variables, which account for demographics, student attributes, and academic performance, produce a statistically significant model, which can be used as an instrument for projecting a student's likelihood to be retained?* Additionally, could the model be used as an instrument for projecting a student's likelihood to be retained? To determine the answer to these questions, a logistic regression analysis was conducted and followed by several goodness-of-fit measures for the model produced by the analysis. In Chapter Two, a section was devoted to providing a historical context for why certain variables should be included in the model building efforts of this project. Broadly, these variables were divided into subgroups, which included demographics, student attributes, academic performance, and social integration.

Demographics. In the demographics subsection of the inputs and variables section of Chapter Two, four variables were discussed. These variables included age, gender, race, and ethnicity (Institutional Data 2018). In prior research, a student's age was found to impact the likelihood a student was retained (Jia & Maloney, 2015). For this reason, age was included in the binary logistic regression analysis. However, when the Akaike Information Criterion was the lowest remaining in step six of the stepwise selection, the variable was removed. Gender was also included in the binary logistic regression analysis, but it was removed in step 10. While the removal of gender corroborated the findings of Jia and Maloney (2014), it did not validate previous findings, which indicated gender was a statistically significant predictor of student loyalty and retention (Astin, 1975, 1999; Vlanden & Barlow, 2014). Additionally, the removal of gender as a variable within the model did not match Bingham and Solverson's (2016) finding which showed gender to be predictive when paired with race within a model. Only ethnicity was included in deidentified data provided by the Midwestern four-year public institution, and so race was not included as a separate variable. Ethnicity was removed from the model in step 11. Ethnicity's removal from the model ran counter to previous findings (Jia & Maloney, 2014; Márquez-Vera et al., 2016).

Student attributes. In the student attributes subsection of the inputs and variables section in Chapter Two, five variables were included for research. These variables included enrollment status, finances, domestic status, program of study, and traditional or non-traditional student status (Institutional Data, 2018). Enrollment status was the first of these variables. Findings of previous studies demonstrated students enrolled full-time were more likely to be retained than those who were not (Calvert, 2014; Fain, 2016; Jia &

Maloney, 2014). In the model construction phase, the researcher determined to exclude enrollment status as each student in the Integrated Postsecondary Education Data System (IPEDS) defined cohort was a full-time student (IPEDS 2016-17 Glossary, 2017). As a result, no additional information was yielded to indicate whether enrollment status was predictive of student retention.

The second variable discussed in the student attributes subsection was finances. Four variables were included in the binary logistic regression analysis to capture students' financial situation within the sample (Institutional Data, 2018). Variables included total loan amount accepted by the student, total scholarship amount accepted by the student, as well as the student's and parent's income indicated by the FAFSA filed for the academic year (Institutional Data, 2018). None of these variables were included in the final model produced in the analysis due to low Akaike Information Criterions (Field, 2017). Parent income was removed in step four. Loan amount accepted was removed in step five. Scholarship amount accepted was removed in step 13. Student income was removed in step 16. In 1982, Tinto theorized finances to be a major factor in student retention. However, Tinto (1982) also believed not enough effort had been given to include these financial variables in retention modeling efforts. An effort was placed on including financial variables in the modeling attempts of this study, however, none were included in the final product.

Two student attribute variables, which were not included within the model but discussed in Chapter Two, included domestic status and program of study (Institutional Data, 2018). Unfortunately, only a very small number of students in the sample were non-domestic students, while several programs of study also included small student numbers. To account for this, the variable was combined with a residency variable to provide an accounting for domestic status within the analysis. Whether the student was a traditional or non-traditional student was also included within the data analysis. The traditional or non-traditional student status variable was included primarily due to Pike and Graunke's (2014) research regarding the impact of non-traditional students on retention of the overall student population. The variable was removed in step 15 of the binary logistic regression analysis.

Academic performance. In the academic performance subsection of the inputs and variables section in Chapter Two, five variables were included for research. These variables included credit hours, first-year grade point average, high school grade point average, standardized tests, and tutoring (Institutional Data, 2018). Credit hours had previously been found to be predictive of student success outcomes (Branand et al., 2015). In the binary logistic regression analysis, credit hours were removed in step 17 of the backward stepwise process.

A student's grade point average in the first year is also strongly linked with student retention outcomes (De Freitas et al., 2015; Harvey & Luckman, 2014). First-year grade point average was to be included in the binary logistic regression analysis. Due to the dataset missing 531 of 939 values, it was determined to remove the variable prior to analysis. Student high school grade point average was included in analysis and was still present in the final model. Presence of grade point average in the final model backs prior research findings, which indicated high school grade point average was a statistically significant indicator of a student's first-year persistence (Belch et al., 2001; Bingham & Solverson, 2016). While prior findings also indicated standardized test scores were predictive of student retention, composite ACT scores were removed from the stepwise analysis in step 14 (Jia & Maloney, 2014; Pike & Graunke, 2014).

The final variable discussed in Chapter Two was tutoring. It is worth considering the tutoring variable could also reflect the value in the social interaction between the tutor and tutee, as Branand et al. (2015) postulated, any strong relationship could engrain a student into the college community. In the binary logistic regression analysis, a student's status as a tutee was included as a binary, categorical variable. Following the binary logistic regression analysis, the variable remained present in the final step.

Social integration. In the social integration subsection of the inputs and variables section in Chapter Two, 10 variables were included for research. These variables included campus activities, student organization membership, college athletics, fraternity or sorority membership, Honors program membership, student government, recreation facilities, recreation programs, and student residency (Institutional Data, 2018). Several of the variables listed in Chapter Two were demonstrated through multiple variables in the deidentified student data (Institutional Data, 2018).

In 1993, Tinto theorized participation in campus activities was one of the two most critical ways in which a student developed a sense of belonging to an institution. Further analysis of Tinto's theories has yielded similar results, involvement in campus activities positively affects student retention outcomes (Belch et al, 2001; Branand et al., 2015). Several variables were used to measure participation in campus activities events, which included career services walk-in appointments, international event touchpoints, and student-activity-fee-funded event touchpoints (Institutional Data, 2018). Each of these represented a type of campus activity program available to students at the

Midwestern four-year public institution (Institutional Data, 2018). Walk-in appointments at the studied institution's career services office was one variable which withstood the backward stepwise selection process and remained in the final model. The number of times a student participated in a student-activity-fee-funded event also was included in the final output of the binary logistic regression. The number of international themed events a student attended was provided as an input for the model but was removed in step 20. These results would seem to indicate participation in campus activities can be positively related to student retention outcomes, depending on the activity.

The other most critical way students can engage in campus, according to Tinto (1993), was through campus organization membership. Unfortunately, this information could not be provided by deidentified student data received from the Midwestern four-year public institution for all student organizations. Membership information was provided for student government and the institution's fraternity and sorority programs (Institutional Data, 2018). As with student organizations, participation in a university's student government programs has been shown to have a positive influence on student satisfaction and social integration (Astin, 1999; Branand et al., 2015). Clarity was not provided through regression analysis as to whether student government played a role on retention outcomes. Removed at step nine, the variable likely did not have a large enough sample to yield any findings.

Sample size was not a challenge in testing fraternity or sorority membership, as tested in research question two. Fraternity or sorority membership remained in the model through the final step. These findings provided additional support of Tinto's (1988), which stated membership in an institution's Greek system provided more social integration opportunities to those students than non-members. Astin (1999) found fraternity and sorority members to have an increased sense of belonging, in addition to being more likely to participate in other institutional activities.

There are other student organizations and groups on campus which have been theorized to impact student success outcomes (Astin, 1999; Branand et al., 2015). Other campus groups, which have not yet been discussed include honors and athletic programs (Astin, 1999; Branand et al., 2015). Unfortunately, honors program membership was not included within the dataset. However, information regarding membership within student athletics indicated 72 of the 939 were student athletes (Institutional Data, 2018). Athlete status was included in the binary logistic regression analysis but was removed in step seven. Another way in which the impact of college athletics on retention was measured was the number of times a student attended an athletic event as a spectator. Athletic event touchpoints were input within the binary logistic regression analysis and remained present through the final step of the process.

Utilization of recreation facilities and programs have also been favorably linked to positive student success outcomes (Belch et al., 2001; Carter & Yeo, 2015). These findings built upon the prior theoretical relationship posited by Tinto (1993) and Astin (1999). Tinto (1993) and Astin (1999) each theorized student involvement in recreation facilities and recreation programs would boost the likelihood a student develops a sense of belonging and is retained at the institution. When testing prior research, four variables were included within the binary logistic regression analysis. Variables capturing recreation facility usage and recreation program participation were aquatics center and recreation center usage, as well as fitness program and intramural sports participation

(Institutional Data, 2018). These data were accounted for via student identification check-ins. None of the variables were present in the final model produced by the analysis. Aquatics center usage was removed in step 19, recreation center usage was removed in step 12, fitness program participation was removed in step three, and intramural sports participation was removed in step two. These findings failed to build upon prior research, which indicated these factors were significant in student retention (Astin, 1999; Belch et al., 2001; Carter & Yeo, 2015; Tinto, 1993).

In 1999, in the theory of student involvement, Astin postulated students who held on-campus jobs would be retained at higher rates than students who were not employed on campus. In the sample, student employment was implemented as a binary, categorical variable. The variable remained in place through each step and was present in the final model of the binary logistic regression.

Another way in which students could be socially integrated within an institution is through on campus housing (Branand et al., 2015). In previous research, it has been found living on campus has a statistically significant impact on the likelihood a student is retained at a university (Belch et al., 2001; Bronkema & Bowman, 2017). More recent findings concur with Astin's (1999) theory of student involvement, which hypothesized living on campus could be the most significant factor in student retention. As with the *z*-test in research question five, the model also removed two forms of residency status in the logistic regression analysis of research question seven. Residency status was removed in step 18 of the backward stepwise process. Additionally, another variable for residency type was also included in the analysis (Institutional Data, 2018). Residency type was a categorical variable and grouped student by residency type at the institution

(Institutional Data, 2018). Residency type included groupings for foreign, in-state resident, in-state tuition recipient but out of state resident, and a non-resident (Institutional Data, 2018). As with residency status, residency type did not make it to the final model and was removed in step 8.

Retention model. Equation 1, which can be found in Chapter Four, represents the model produced by the logistic regression analysis. The model represents a parsimonious model, meaning as few predictor variables were used as possible (Levine et al., 2016). The model featured a Nagelkerke R^2 of 0.243 and a Cox and Snell R^2 of 0.176. The R^2 is used to demonstrate the amount of variability in the output variable, which can be attributed to the model (Levine et al., 2016). In this case, depending on which of the two versions of R^2 is utilized, either 24.3% or 17.6% of the variability is explained (Levine et al., 2016). The R^2 of the model developed in this study represents and improvement on Jia and Maloney's (2014) modeling efforts, which yielded a R^2 of 0.115. Seven variables featured in the final model, including binary categorical variables, to represent whether the student was a tutee, a member of a fraternity or sorority, or a student employee. Additionally, the model included scale variables including the number of student-activity-fee-funded events, sporting events, and career services walk-in appointments the student engaged in. Finally, high school grade point average was left as the sole academic performance variable. Each of the variables in the final model along with the respective coefficients, standard errors, and *p*-values are included in Table 8, which can be found in Chapter Four.

Results of research question seven indicated the addition of social integration variables could expand upon the statistical significance of the model produced through this study. Consider the following goodness-of-fit measures, Bingham and Solverson's (2016) model produced a *p*-value of 0.289, while the model produced through this research returned a *p*-value of 0.423. Another measure, the *C*-Statistic, or the measure of the area under the receiver operator curve, yielded a result of 0.752 (Bingham & Solverson, 2016). A *C*-Statistic of 0.752 means there is a 75.2% chance a non-retained student chosen at random will have a lower likelihood to be retained than a retained student chosen at random (Jia & Maloney, 2015). A *C*-Statistic of 0.752 represented a higher mark than Jia and Maloney's (2014) figure of 0.718 and, more recently, the 0.651 mark produced by Bingham and Solverson in 2016. An improvement to previous modeling efforts due to the introduction of social integration variables are indicated by the strength of the goodness-of-fit measures (Bingham & Solverson, 2016; Jia & Maloney, 2014).

Implications for Practice

Within this research, higher education administrators are provided additional data, which corroborates theoretical concepts student retention theory is based. These findings do not represent the first-time student behavioral inputs which have been used in student retention modeling efforts (Bingham & Solverson, 2016; Bronkema & Bowman, 2017; Jia & Maloney, 2014). Findings provide administrators with a broader understanding of how they can track student engagement behavior on campus within retention modeling efforts. Additionally, inclusion of this information only furthers the ability of an institution's faculty and staff to better serve students in more meaningful and impactful ways (De Freitas et al., 2015).

Validation of theoretical framework. One of the primary objectives for this research was to bridge the gap between existing studies on student retention statistical modeling and student retention theory. Modeling techniques used previously relied heavily on student demographics, attributes, and in some cases academic performance (Bingham & Solverson, 2016; Márquez-Vera et al., 2016; Pike & Graunke, 2014). In this research, card swipe data was introduced as a to capture sense of belonging and social integration information. Many of the variables included in prior research were pre-college characteristics, academic, or in some cases qualitative, in how sense of belonging was included (Bingham & Solverson, 2016; Jia & Maloney, 2014; Márquez-Vera et al., 2016; Pike & Graunke, 2014). Through *z*-testing and logistic regression analysis, which included student integration data, Astin's (1975, 1993, 1999) and Tinto's (1982, 1988, 1993, 2001, 2007, 2017) theories on student retention have been validated.

There is a perception which holds institutions responsible for retaining students (Tinto, 2006). Since the origin of student retention theory, it has been postulated a student's pre-college characteristics including demographics and student attributes combined with a student's interactions with the college environment determine which students are retained or drop out (Tinto, 1982). Those students who connect with campus develop the strongest institutional bonds (Astin, 1999). Involved students connect with peers, faculty, and staff, and participate in extracurricular activities (Astin, 1999). It is the students who are involved who develop a sense of belonging, while uninvolved students do not (Tinto, 2017). It is sense of belonging, which ties students to the institution and binds the student to the institution's communities (Tinto, 2017). Results of this research provide a statistical validation of those theoretical ideas. Finally, through analytics presented in research findings, higher education professionals can also measure the sense of belonging they are instilling on the student level.

Combating student attrition. To be financially viable, higher education institutions must maintain enrollment levels (Harvey & Luckman, 2014). A key component of enrollment is to combat student attrition (Harvey & Luckman, 2014). The findings of this research could help to accomplish this challenge in several critical ways. This model leveraged with several strategies could prove to be extremely effective in influencing institutional retention rates. Several key pieces to this puzzle would need to be implemented to ensure effectiveness of the overall analytic strategy. Implementation would require a central repository of data, continuous evaluation and identification of at-risk students, and outreach strategies. All of these would need to coincide with a constant effort to enhance and improve modeling efforts.

The first step in the process of applying findings of the research would be to develop a central data repository. At many institutions, data are compartmentalized throughout the organization (De Freitas et al., 2015). An essential first step in any process to apply predictive modeling solutions is to get the data flowing into one system (De Freitas et al., 2015). Until infrastructure is developed, it is impossible to truly integrate data into the decision-making process of university personnel (De Freitas et al., 2015). Once centralized, data should be available to all stakeholders (Blumenstyk, 2016). Stakeholders must also be diligent in treating data predictions ethically, appropriately, and with confidentiality (De Freitas et al., 2015). By developing appropriate data infrastructure, stronger conclusions regarding students is possible. With an increased level of information regarding each student, any predictive analytic strategies would be more accurate and effective.

To appropriately combat student attrition, institutional leaders must develop methods to identify students who are least likely to be retained (Harvey & Luckman, 2014). Identification of at-risk students early in the student's college experience is essential to any strategy aimed at improving retention outcomes (Márquez-Vera et al., 2016). Identification of at-risk students is not a new technique, as many schools have already began using academic and demographic variables to identify low- and high-risk students (Ekowo & Palmer, 2017). The quantification of variables aimed at measuring social integration and sense of belonging should only improve the identification process. A quality analytics system should be capable of identifying at risk students for continuous evaluation throughout the semester as variables and inputs change. Many institutions have already developed early warning systems to identify at-risk students (Márquez-Vera et al., 2016). Early warning systems should be developed so when a student's likelihood to be retained falls below a certain threshold, relevant university faculty and staff should be notified (Marquez -Vera et al., 2016). Relevant faculty and staff would serve as a circle of care of the student, and would include the student's faculty and advisor, as well as other key figures. Examples of other individuals who may be a part of the circle of care include the student's financial aid counselor, if the student works on campus his or her supervisor, residence life staff, student organization advisor, disability services personnel, counseling staff, or any other faculty or staff member who has developed a relationship with the student. It is the relationships a student builds with university faculty and staff, which will result in retention of more students (Vlanden & Barlow, 2014).

Findings of this research could also serve to improve outreach strategies to high-risk students. Institutions on the cutting edge of predictive analytics have utilized data in such a way as to make student outreach feel more personal to students (Straumsheim, 2017). Any additional information on student behavior creates more opportunities to personalize the outreach a student receives. By taking into consideration social integration data, university leadership can better identify personnel who may have connected with a student. A faculty or staff member who has had interactions with the student is more likely to have successful outreach than a random faculty or staff member (Straumsheim, 2017). When individuals who are conducting outreach have more information, they are more likely to build an impactful relationship (Astin, 1999; Branand et al., 2015; Straumsheim, 2017; Supiano, 2018; Tinto, 2006).

Forecasting. Student retention forecasting is extremely helpful for higher education administrators (Calvert, 2014). With an enhanced ability to forecast future enrollments via an improved method of projecting student retention, higher education administrators could be more strategic in long-term planning. One of the key components and benefits of predictive analytics is the increased ability to forecast for future growth (Calvert, 2014). An ability to model the likelihood of student retention could also be applied to model the number of students who will be retained from one semester to the next. Simply put, predictive analytics provide higher education leaders with a means to forecast the behavior of a student population (Calvert, 2014).

As more states across the country continue to re-appropriate resources away from higher education, the ability to better forecast the size of the institution's student population, as well as the needs of those students, is extremely advantageous (Kelderman, 2018). By improving the institution's ability to project which students are more likely to return than not, administrators are able to better deploy the university's resources. The difference in the cost of recruiting a new student, versus the costs of retaining a current student is enormous (Vlanden & Barlow, 2014). A proper pairing of student and outreach strategies allows the university to boost revenues and alleviate pressure on enrollment management personnel to fill in the persistence gap with additional student recruits (Vlanden & Barlow, 2014). With the rising pressure for leaders in higher education to run a cost-efficient university, it is easy to understand why student retention should be a top priority (Page & Gehlbach, 2018; Vlanden & Barlow, 2014). Not to mention, graduation of students should be at the core of higher education, and an institution cannot graduate a student without first retaining said student. Through predictive analytics, higher education has a path to achieve objectives in the most economical way possible (De Freitas et al., 2015). With an economical approach, leaders can change the lives of more students than less fiscally responsible methods (De Freitas et al., 2015).

The choice between a data driven approach and a relationship based approach to student retention is not necessary. Predictive analytics create the opportunity for institutions to be more refined in relationship based approaches through the support of data (Straumsheim, 2017). Predictive analytics and big data are not intended to replace the people behind the relationships, as they are instrumental in a student's sense of belonging. Instead, predictive analytics should be thought of as an integral piece within a loop. Every student is unique. Rather than a blanket outreach approach, complexities should be tracked within an analytic system. A quality analytic system provides university faculty and staff with information which allows them to connect with a student on a more effective level. When a student responds to outreach, the student's profile changes within the system. The cycle then repeats itself. Relationships will always be a core piece to student retention (Astin 1993; Tinto, 2006). However, analytics provide the tools needed to build these relationships in more impactful and meaningful ways on a larger scale (Straumsheim, 2017).

Recommendations for Future Research

There are several ways in which future research on the topic of predictive modeling and student retention could be explored. First, there should be considerations as to how to improve upon the quantitative methodology presented in Chapter Three. These strategies could include carrying out research longitudinally to explore various transitions in the college experience. Other opportunities could be collecting additional variables to include in the model, utilizing a different selection criteria technique, or changing the sample through multiple years of data, or examining another institutional type.

Another adjustment which could be made to the quantitative process would be to add descriptive quantitative data collection strategies. One example would be surveying students in the sample to determine sense of belonging or social integration. Another opportunity would be to survey at specific engagement opportunities to see if students' felt as though events made them more a part of the community. Yet another opportunity would be to use a similar research method to determine if a student's state sense of belonging correlated with social engagement indicators collected by card swipe data. Unlimited opportunities are available to expand upon the research presented, but the suggestions seem to be the most logical next steps.

According to Astin (1999) and Tinto (1982), it is important retention is analyzed in a longitudinal manner. In this research, students were analyzed from first- to second-year in college. The decision to analyze retention from first to second-year was made for two primary reasons. First, institutional official retention rates are based on first to second-year data (IPEDS 2016-17 Glossary, 2017). Second, analysis of first to second-year allows the research to be conducted in a longitudinal manner as Astin (1999) and Tinto (1982) advocated for. Future research could piece these findings together with other studies which examine second- to third-year retention, third- to fourth-year retention, or other time frames.

Another way to build upon the study's research would be to add additional variables into the model building process. Twenty-four variables were included within the data collection process for this study. While this may seem like a large amount, it is a small number when other things a student does on or before arriving to campus are considered. More data could have been collected regarding the student's financial situation through FAFSA. Additionally, characteristics of a student's hometown or high school could also play a major role in the student's preparedness or comfort level with an institution. Finally, as more classes are set up within an online platform, there are enormous opportunities for a student's utilization of these mediums to be captured. How frequently is the student logging in? Is the student watching lecture videos? Has the student participated in discussion board opportunities? These are just a few questions which could open an enormous wealth of information as higher education professionals seek to better understand and predict actions of students.

Future researchers could also alter the methodology of the study in other manners. For the reasons discussed in Chapter Three, the backward stepwise selection process was utilized within data analysis software (Field, 2017). However, there are other options which could yield different results. Another way in which the methodology could be adjusted is by changing the sample. Each university comes with its own unique set of characteristic and demographic profiles. The sample of this research came from a commuter-based, liberal arts, four-year Midwestern public institution. This research could be further examined through any number of ways due to the sample coming from an institution with a different profile. Another opportunity to adjust the research would be to collect data for more than one cohort. Additional cohorts could provide a larger sample, which could result in more statistically significant data (Bluman, 2017).

The most impactful, and possibly most interesting way, future research could build upon this study would be to introduce a mixed methods approach. Data analysis for the model building process described in Chapter Three was built upon Astin's (1975, 1992, 1999) belief which indicated a more involved student would have a stronger connection to the university's community. Further research could introduce qualitative elements or other quantitative elements to capture a student's sense of belonging and attachment to collegiate communities (Tinto, 2017). A change in quantitative approach could be accomplished via post-event surveys, or surveys to the entire sample at various points in the student's college experience. Post-event or engagement surveys could even provide researchers the chance to build a stronger understanding of which types of engagement opportunities are perceived by students to be the most impactful, versus which are statistically the most impactful. It is possible future research could find terms, which better resonate with the student's integration process than the ones currently used. A qualitative method to add to the study would be to gather student sense of belonging sentiments via interviews or focus groups (Creswell & Creswell, 2018). Information obtained in those sessions could then be included with other quantitative data to obtain a more holistic portrait of the sense of belonging process.

There are a multitude of directions future research could take in building upon results of this study. Whether it be through methodology, sample modification, or the introduction of qualitative methods, findings presented in this study represent just a step toward better understanding. It will be for future researchers to determine which direction they believe could offer higher education the most to make retention practices and outreach strategies most effective.

Summary

This quantitative study was pursued to expand upon knowledge concerning the role of student involvement, social integration, and sense of belonging in the student retention process. The basis of the study was built upon the theoretical framework of Gennep's (1960) rites of passage theory, Tinto's (1982, 1993) model of college dropout and theory of student departure, as well as Astin's (1975, 1999) work on college dropouts and his theory of student involvement. Through the study, the researcher attempted to bridge the gap between the theoretical framework of student retention and recent attempts to develop a statistical model, which could forecast a student's likelihood to persist given a set of demographic, student attribute, and academic performance inputs. Seven research questions were developed to determine the role of social integration variables in the student retention process. Research question seven was intended to provide further clarity as to how social integration inputs interplayed with other student inputs to project the likelihood a student was retained at an institution.

In Chapter Two, the researcher presented a theoretical framework built upon Astin's (1975, 1993, 1999) and Tinto's (1982, 1988, 1993, 2001, 2007, 2017) theories on student involvement and retention. Through the literature review, a deep dive was then provided into concepts of social integration and sense of belonging. Predictive analytics' current application in higher education was included with the intent to enhance the reader's knowledge of its use in the higher education realm. Additionally, in the predictive analytics section of Chapter Two, discussion was included on current examples of successful predictive analytic programs in higher education, as well special considerations institutions must take when implementing such strategies. Chapter Two concluded with literature backing the variables, which were requested for the data analysis of the study.

Findings of the study mirrored much of what was discussed in Chapter Two. Campus fitness programs, fraternity or sorority membership, recreation facility usage, and student-activity-fee-funded event participation were deemed as having a statistically significant relationship to student retention, which validated theory the questions were based on. Intramural sports participation and on-campus living did not yield statistically significant results, and thus findings failed to corroborate the literature review presented in Chapter Two. Built for research question seven, the statistical model provided stronger goodness-of-fit scores than previous modeling efforts, which did not account for social integration via card scan data. Social integration data's critical nature in the model building exercise served to confirm Astin's (1975, 1993, 1999) and Tinto's (1982, 1988, 1993, 2001, 2007, 2017) theories on student retention, which postulated student involvement was integral to the retention process.

Findings of this research should provide university leadership with several key takeaways. Big data and predictive analytics have a place in higher education. If an institution's faculty and staff can develop an ethical, organized, and thoughtful approach to collecting, storing, and implementing student data in meaningful ways, they can be better equipped to serve students. Student outreach can be more strategic and specialized to students if they are targeted for all, while being better stewards of the institution's resources. These strategies should be implemented in a way which allows university personnel to forge stronger and better relationships with students.

Most importantly, findings of this research provide statistical backing to the emphasis placed a student's relationship to institution stakeholders (Branand et al., 2015). One of the most impactful components of a student's retention is the relationship a student builds with fellow students, faculty, and staff (Astin, 1993). Results of the analysis presented in this study should serve to highlight the importance of relationship building. Retention should be viewed as a cycle, in which a relationship between students and the college community must always be evaluated and improved upon. The greater sense of community higher education administrators can instill in students, the more likely the student will persist, graduate, and ultimately live a more successful life.

Appendix A

Approval from University Institutional Review Board

[redacted]

DATE:	March 20, 2018
TO:	Landon Adams
FROM:	[redacted]
PROJECT TITLE:	[1200510-1] Understanding the Relationship Between Student Demographic, Attribute, Academic, and Social Engagement Factors with Retention
REFERENCE #:	1200510-1
SUBMISSION TYPE:	New Project
ACTION:	APPROVED
APPROVAL DATE:	March 20, 2018
EXPIRATION DATE:	March 20, 2019
REVIEW TYPE:	Facilitated Review

Thank you for your submission of New Project materials for this project. The [redacted] IRB has APPROVED your submission. This approval is based on an appropriate risk/ benefit ratio and a project design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Facilitated Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Federal regulations require that each participant receives a copy of the consent document.

Please note that any revision to previously approved materials must be approved by this committee prior to initiation. Please use the appropriate revision forms for this procedure.

All UNANTICIPATED PROBLEMS involving risks to subjects or others (UPIRSOs) and SERIOUS and UNEXPECTED adverse events must be reported promptly to this office. Please use the appropriate reporting forms for this procedure. All FDA and sponsor reporting requirements should also be followed.

All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to this office.

This project has been determined to be a Minimal Risk project. Based on the risks, this project requires continuing review by this committee on an annual basis. Please use the appropriate forms for this procedure. Your documentation for continuing review must be received with sufficient time for review and continued approval before the expiration date of March 20, 2019.

Generated on IRBNet

Please note that all research records must be retained for a minimum of three years after the completion of the project.

If you have any questions, please contact [redacted]. Please include your project title and reference number in all correspondence with this committee.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within [redacted] IRB's records.

Generated on IRBNet

Appendix B

Support Letter from University Administration

[redacted]

Student Affairs Office

Lindenwood University Institutional Review Board
Office for Human Research Protections
209 South Kingshighway Street
St. Charles, MO 63301

November 30, 2017

Dear Lindenwood University IRB:

On behalf of [redacted] I am writing to grant permission for Landon Adams, an Ed. D. candidate at Lindenwood University, to conduct his research titled, "Understanding the Relationship Between Student Demographic, Attribute, Academic, and Social Engagement Factors with Retention".

This permission is contingent upon Institutional Review Board approval from Lindenwood University and [redacted] if deemed a procedural step by the administration. I understand that Landon will be utilizing institutional student data from the fall 2016 and spring 2017 semesters to develop a model, which could lead to a better understanding of student retention at our institution.

We at [redacted] are in full support of Landon's research and are thrilled to contribute to this important research.

Sincerely,

Darren Fullerton
Vice President for Student Affairs
[redacted]

Appendix C

Determination of Exempt Status from University Institutional Review Board

LINDENWOOD

LINDENWOOD UNIVERSITY ST. CHARLES, MISSOURI

DATE:	February 13, 2018
TO:	Landon Adams
FROM:	Lindenwood University Institutional Review Board
STUDY TITLE:	[1137183-1] Understanding the Relationship Between Student Demographic, Attribute, Academic, and Social Engagement Factors with Retention
IRB REFERENCE #:	
SUBMISSION TYPE:	New Project
ACTION:	DETERMINATION OF EXEMPT STATUS
DECISION DATE:	February 13, 2018
REVIEW CATEGORY:	Exemption category # *4*

Thank you for your submission of New Project materials for this research study. Lindenwood University Institutional Review Board has determined this project is EXEMPT FROM IRB REVIEW according to federal regulations.

We will put a copy of this correspondence on file in our office.

If you have any questions, please send them to IRB@lindenwood.edu. Please include your project title and reference number in all correspondence with this committee.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within Lindenwood University Institutional Review Board's records.

Appendix D

Data Request to University Institutional Effectiveness

Recipient Data:

Time Finished: 2018-03-22 09:42:37 CDT

IP: [redacted]

ResponseID: R_3TRgTxScuRFTGZX

Link to View Results:

URL to View

Results: [redacted]

Response Summary:

NAME

Landon Adams

CITY, STATE, ZIP CODE

Joplin, MO, 64801

DEPARTMENT, COMMITTEE, BUSINESS, INSTITUTION, or AFFILIATION

Lindenwood Ed.D program for Higher Education Administration

DESCRIBE THE REASON FOR YOUR REQUEST (check all that apply)

Student Research

DESCRIBE THE RESEARCH QUESTION YOU ARE TRYING TO ANSWER

1. Does student participation, minimum one class attended, in campus fitness programs have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

2. Does student membership in a fraternity or sorority have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

3. Does student participation, minimum one intramural event attended, in intramural sports programs have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

4. Does student participation, minimum one check-in, at a university recreational facility have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

5. Does student housing status, living on-campus or not, have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

6. Does student participation, minimum one event attended, in student-activity-fee-funded events have a statistically significant impact on first-year to second-year retention at a Midwestern four-year public institution?

7. Do student engagement variables, in association with already established variables, which account for demographics, student attributes, and academic performance, produce a statistically significant model, which can be used as an instrument for projecting a student's likelihood to be retained?

WHAT DATA ARE YOU NEEDING TO HELP YOU ADDRESS THAT RESEARCH QUESTION?

I need three total reports. Fall 2016, Spring 2017, and Fall 2017 census with the following variables for each student.

Age, Gender, Race, Ethnicity, Marital status, Enrollment Status (full-time vs part-time), Domestic Status (domestic vs international student), program of study, traditional vs non-traditional, athlete vs non-athlete, high school grade point average, credit hours enrolled, credit hours completed, cumulative grade point average, previous semester grade point average, ACT score, composite ACT scores for (reading, science, English, math), tutoring hours/sessions attended (could be gathered from Tutor Trax software), campus activities board event touchpoints (card scans), student organization membership, international event touchpoints (card scans), athletics event touchpoints (card scans), theater performance touchpoints, Greek Life membership, Honors Program membership, Student Senate membership, Show-Me-Gold membership, Recreation Center touchpoints (card scans), aquatic center touchpoints (card scans), intramural sport touchpoints (card scans), fitness class touchpoints (card scans), RHA event touchpoints (card scans), on-campus employment hours worked, living status, residence hall, financial aid gap, percent of need met, total gift funds, total gift funds, total financial aid package, financial aid gap, total income, parent income, department or program area, college of study (school of business, school of education, etc.), distance from campus, sport, number of days as admit, number of days FAFSA received, number of days packaged, .

For the touchpoint data, they won't be able to be ran as a part of a census report, because they aren't stored in Banner. They'd have to be gathered from the responsible areas. So it is no problem, if I get A) a semester report of how many times a certain student attended an event or B) a ZIP file of all the semesters banner participation reports of the individual events.

WHICH OF THE FOLLOWING ARE YOU NEEDING (check all that apply)?

Raw Data

Student Names, IDs, or other Identifying Information

Have you sent your study through the Institutional Review Board? If so, what is the Reference Number in IRBnet? (All research using identifiers must be approved by the IRB: [redacted]

Yes -- 1200510-1

IE staff time spent gathering data must be justified according to the extent to which the IE office will be able to use the results from the requested project to benefit the students of [redacted]. Please explain how [redacted] IE will receive results from your project and how [redacted] students will benefit from the project, directly or indirectly.

Results from dissertation will be provided back to [redacted]. Specifically, results will be shared with Dean of Student Success for retention purposes. Results will also be used to better understand the university specific dynamics and how they interact with one another to impact retention.

PROVIDE DETAILED INFORMATION CONCERNING DATA DEFINITIONS FOR THE DATA YOU ARE REQUESTING; BELOW ARE SOME QUESTIONS TO CONSIDER:

By "professor," do you mean all full-time instructors? Include adjuncts? By "international students," do you mean students with permanent residencies outside the US? By "graduated in this major" do you mean 1st or 2nd major? declared as of when? graduated any year?

I would be happy to further elaborate on any particular variable. Many of the variables included have been pulled because of their present on other reports I have seen or worked with prior to beginning my dissertation. For example, most of the variables are available within Banner and have been including in data dumps sent to Noel-Levitiz for retention study. The touchpoint data will likely need to be collected on a departmental level. The resident Life, recreation, and student life touchpoints are all collected by those areas and should be easily provided. I am not entirely sure what the system is for Theater, Athletics, or International events of tracking those engagement. If we can get reports or touchpoint data from these areas it is a bonus for me, I am primarily focused on adding student life touchpoint data to the already existing retention modeling efforts, which focus on the other variables I have requested.

Institutional Effectiveness seeks to respond to all requests within two weeks of posting. I would like to receive this information by -- Email or Phone Call

This request is urgent (please explain) -- I know you all are swamped with HLC. I am basically at a point in my dissertation process where I am at a standstill until I get the data. I'd appreciate any help you can give me in getting this to many as quick as possible so I can keep the ball rolling. I'm happy to help in anyway.

References

Astin, A. W. (1975). *Preventing students from dropping out*. San Francisco, CA: Jossey-Bass.

Astin, A. W. (1993). *What matters in college: Four critical years revisited.* San Francisco, CA: Jossey-Bass.

Astin, A. W. (1999). Student involvement: A developmental theory for higher education. *Journal of College Student Development, 40*(5), 518. Retrieved from https://search.proquest.com/docview/195180247/fulltextPDF/3BDBFA7B82A54F4EPQ/1?accountid=12570

Belch, H. A., Gebel, M., & Maas, G. M. (2001). Relationship between student recreation complex use, academic performance, and persistence of first-time freshmen. *NASPA Journal, 38*(2), 254. Retrieved from https://search.proquest.com/docview/223541151?accountid=12570

Belhekar, V. M. (2016). *Statistics for psychology using r* (1st ed.). Thousand Oaks, CA: SAGE Publications Inc.

Biemiller, L. (2017). *Big data for student success still limited to early adopters.* Retrieved from http://www.chronicle.com/article/Big-Data-for-Student-Success/239713

Bingham, M. A., & Solverson, N. W. (2016). Using enrollment data to predict retention rate. *Journal of Student Affairs Research and Practice, 53*(1), 51-64. doi:10.1080/19496591.2016.1110035

Bluman, A. G. (2017). *Elementary statistics: A step by step approach* (10th ed.). New York, NY: McGraw-Hill

Blumenstyk, G. (2016, February 23). *As big-data companies come to teaching, a pioneer issues a warning*. Retrieved from http://www.chronicle.com/article/As-Big-Data-Companies-Come-to/235400

Blumenthal, K. J. (2009). Collegiate recreational sports: Pivotal players in student success. *Planning for Higher Education, 37*(2), 52-62. Retrieved from https://search.proquest.com/docview/212568746?accountid=12570

Branand, B., Mashek, D., Wray-Lake, L., & Coffey, J. K. (2015). Inclusion of college community in the self: A longitudinal study of the role of self-expansion in students' satisfaction. *Journal of College Student Development*, *56*(8), 829-844. doi:10.1353/csd.2015.0080

Bronkema, R., & Bowman, N. (2017). A residential paradox? Residence hall attributes and college student outcomes. *Journal of College Student Development*, *58*(4), 624-629. doi:10.1353/csd.2017.0047

Calvert, C. E. (2014). Developing a model and applications for probabilities of student success: A case study of predictive analytics. *Open Learning*, *29*(2), 160-173. doi:10.1080/02680513.2014.931805

Carter, S., & Yeo, A. C. (2016). Students-as-customers' satisfaction, predictive retention with marketing implications. *The International Journal of Educational Management*, *30*(5), 635-652. Retrieved from https://search.proquest.com/docview/1793895327?accountid=12570

Copus, C., & McKinney, B. (2016). Early integration of tutorial support in beginning algebra. *Journal of Developmental Education*, *40*(1), 32-34. Retrieved from ezproxy.lindenwood.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=121361901&site=ehost-live.

Creswell, J. W. & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: SAGE Publications Inc.

De Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., & ... Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*, *46*(6), 1175-1188. doi:10.1111/bjet.12212

Denson, N., & Bowman, N. (2015). The development and validation of an instrument assessing student-institution fit. *Assessment & Evaluation in Higher Education*, *40*(8), 1104-1122. doi:10.1080/02602938.2014.970513

Dimeo, J. (2017). *Data dive.* Retrieved from https://www.insidehighered.com/digital-learning/article/2017/07/19/georgia-state-improves-student-outcomes-data

Ekowo, M., & Palmer, I. (2017). *Predictive analytics in higher education: Five guiding practices for ethical use.* Retrieved from https://www.newamerica.org/education-policy/policy-papers/predictive-analytics-higher-education/#

Fain, P. (2016). *Data on student engagement with an LMS is a key to predicting retention.* Retrieved from https://www.insidehighered.com/news/2016/06/13/data-student-engagement-lms-key-predicting-retentionlearning/article/2017/07/19/georgia-state-improves-student-outcomes-data

Fain, P. (2017). *Negative findings on performance based funding*. Retrieved from https://www.insidehighered.com/quicktakes/2017/12/18/negative-findings-performance-based-funding

Fernandes, A., Ford, A., Rayner, G., & Pretorius, L. (2017). Building a sense of belonging among tertiary commuter students: The Monash non-residential colleges program. *Student Success*, *8*(2), 31-42. doi:10.5204/ssj.v8i2.380

Field, A. (2017). *Discovering statistics using SPSS* (5th ed.). Thousand Oaks, CA: SAGE Publications Inc.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. (2014). *How to design and evaluate research in education* (9th ed.). New York, NY: McGraw-Hill Education.

Gennep, A. V. (1960). *The rites of passage.* Chicago, IL: University of Chicago Press.

Gill, J. (2017). *How big data is helping to close the student retention gap.* Retrieved from https://www.timeshighereducation.com/opinion/how-big-data-helping-close-student-retention-gap

Gurnsey, R. (2017). *Statistics for research in psychology: A modern approach using estimation* (1st ed.). Thousand Oaks, CA: SAGE Publications Inc.

Harvey, A., & Luckman, M. (2014). Beyond demographics: Predicting student attrition within the bachelor of arts degree. *The International Journal of the First-year in Higher Education, 5*(1), 19-29. doi:10.5204/intjfyhe.v5i1.187

IPEDS 2016-17 Glossary. (2017). *IPEDS 2016-17 data collections system*. Retrieved from https://surveys.nces.ed.gov/ipeds/VisGlossaryAll.aspx

Jacoby, B. (2015). Enhancing commuter student success: What's theory got to do with it? *New Directions for Student Services, 150*(1), 3-12. doi:12.1002/ss.20122

Jia, P., & Maloney, T. (2015). Using predictive modeling to identify students at risk of poor university outcomes. *Higher Education*, *70*(1), 127-149. doi:10.1007/s10734-014-9829-7

Kamenetz, A. (2016). *How one university used big data to boost graduation rates.* Retrieved from http://www.npr.org/sections/ed/2016/10/30/499200614/how-one-university-used-big-data-to-boost-graduation-rates

Kelderman, E. (2018). *State spending on higher education has inched upward. But most public colleges can't celebrate.* Retrieved from https://www.chronicle.com/article/State-Spending-on-Higher/242298

Leavy, P. (2017). *Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches* (1st ed.). New York, NY: The Guilford Press.

Levine, D. M., Stephan, D. F., Krehbiel, T. C., Berenson, M. L. (2016). *Statistics for managers* (8th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Manata, B., DeAngelis, B. N., Paik, J. E., Miller, V. D. (2017). Measuring critical aspects of the resident assistant role. *Journal of College Student Development, 58*(4), 618-623. doi:10.1353/csd.2017.0046

Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert Systems*, *33*(1), 107-124. doi:10.1111/exsy.12135

Masika, R., & Jones, J. (2016). Building student belonging and engagement: Insights into higher education students' experiences of participating and learning together. *Teaching in Higher Education*, *21*(2), 138-150. doi:10.1080/13562517.2015.1122585

Missouri Department of Higher Education. (2017). *Performance funding in Missouri*. Retrieved from http://dhe.mo.gov/PerformanceFunding.php

O'Flaherty, B., & Heavin, C. (2015). Positioning predictive analytics for customer retention. *Journal of Decision Systems*, *24*(1), 3-18. doi:10.1080/12460125.2015.994353

Page, L., & Gehlbach, H. (2018). *How Georgia State University used an algorithm to help students navigate the road to college.* Retrieved from https://hbr.org/2018/01/how-georgia-state-university-used-an-algorithm-to-help-students-navigate-the-road-to-college

Papandrea, D. (2017). *Analyze student data for real progress in student success; 5 signs that data analytics mindset has infiltrated campus culture.* Retrieved from https://www.universitybusiness.com/article/analyze-student-data-real-progress-student-success

Pike, G., & Graunke, S. (2014). Examining the effects of institutional and cohort characteristics on retention rates. *Research in Higher Education*, *56*(2), 146-165. doi:10.1007/s11162-014-9360-9

Punch, K. (2014). *Introduction to social research: Quantitative and qualitative approaches.* Thousand Oaks, CA: SAGE Publications Inc.

Reed, M. (2017). *Ethics, design, and control.* Retrieved from https://www.insidehighered.com/blogs/confessions-community-college-dean/ethics-design-and-control#.WL7JloyCaXE.mailto

Salkind, N. J. (2016). *Statistics for people who think they hate statistics* (6th ed.). Thousand Oaks, CA: SAGE Publications Inc.

Seltzer, R. (2018). *Anemic state funding growth.* Retrieved from https://www.insidehighered.com/news/2018/01/22/state-support-higher-ed-grows-16-percent-2018

Soria, K. M., & Taylor J. R. L. (2016). Strengths-based approaches in college and university student housing: Implications for first-year students' retention and engagement. *Journal of College & University Student Housing*, *42*(2), 60-75. doi:10.1111/hea/12090_36

Straumsheim, L. (2016). *A health care strategy for higher ed.* Retrieved from https://www.insidehighered.com/news/2016/04/19/colleges-embrace-strategy-health-care-industry-improve-student-retention

Straumsheim, L. (2017). *A new system for student success planning.* Retrieved from https://www.insidehighered.com/news/2017/07/05/colleges-need-enterprise-level-software-tackle-student-success-issues-company-says

Supiano, B. (2018). *Relationships are central to the student experience. Can colleges engineer them?* Retrieved from https://www.chronicle.com/article/Relationships-Are-Central-to/242230

Tinto, V. (1975). Dropouts from higher education: A theoretical synthesis of recent literature. *A Review of Education Research*, *45*(1), 89-125. doi:10.3102/00346543045001089

Tinto, V. (1982). Limits of theory and practice in student attrition. *The Journal of Higher Education, 53*(6), 687-700. doi:10.2307/1981525

Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. *The Journal of Higher Education, 59*(4), 438-455. doi:10.2307/1981920

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

Tinto, V. (2006). Research and practice of student retention: What next? *Journal of College Student Retention, 8*(1), 1-19. doi:10.2190/4YNU-4TMB-22DJAN4W

Tinto, V. (2017). Reflections on student persistence. *Student Success, 8*(2), 1-8. doi:10.5204/ssj.v8i2.376

Valbrun, M. (2018). *Study questions whether 'high impact' practices yield higher graduation rates.* Retrieved from https://www.insidehighered.com/news/2018/04/25/study-questions-whether-high-impact-practices-yield-higher-graduation-rates

Vlanden, J., & Barlow, P. J. (2014). Showing the love: Predictors of student loyalty to undergraduate institutions. *Journal of Student Affairs Research and Practice, 51*(1), 16-29. doi:10.1515/jsarp-2014-0002

Williams, J. (2017). *Four year graduation rate rises from 51 to 66 percent in five years.* Retrieved from https://news.utexas.edu/2017/09/20/four-year-graduation-rate-rises-from-51-to-66-percent

Vita

Landon Adams was born in 1990 in Joplin, Missouri. In 2008, he graduated from Carl Junction High School. Four years later Adams became a graduate of Missouri Southern State University, a four-year public liberal arts institution. There Adams received a bachelor of science in business administration in marketing, with a minor in sociology. Adams completed his master's in business administration in leadership through Adams State University in 2015.

Adams began his career in higher education at Missouri Southern University in May of 2013, as the Director of Student Activities. He remained in this position through June 2016, after which he began a new role as the Director of Student Life. In June 2017, Adams again took on additional responsibility and became the Director of Student Life and Conduct.

Throughout his career in higher education, Adams attended many conferences and trainings, most recently the Gehring Academy through the Association for Student Conduct Administration, and annual conferences for the Higher Learning Commission. Finally, Adams has served as a member, vice chair, and chair on a multitude of committees at Missouri Southern State University.

Adams lives in Carl Junction, Missouri, with his wife, Katie. Together they have one child, Annabelle Ruth Adams, born November 20, 2016.